Taxing Top Earners: A Human Capital Perspective*

Alejandro Badel
Georgetown University
Bureau of Labor Statistics
ale.badel@gmail.com

Mark Huggett
Georgetown University
mh5@georgetown.edu

Wenlan Luo
Tsinghua University
luowenlan@gmail.com

THIS DRAFT: NOVEMBER 13, 2017

Abstract
An established view is that the revenue maximizing top tax rate for the US is approximately 73 percent. We argue that theory and evidence suggest a lower value. First, theory provides a robust formula where three elasticities determine this top rate rather than the one elasticity highlighted in the established view. Second, theory and measurement suggest that the two new elasticities are positive and reduce the revenue-maximizing rate. Third, a human capital model is provided that follows the logic of the formula, is consistent with US regression evidence on the response of income after a tax reform and with US evidence for the coefficients entering the tax rate formula and yet features a revenue maximizing top rate well below 73 percent.

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

JEL Classification: D91, E21, H2, J24

*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.

1 Introduction

How should the tax rate on top earners be set? A well-established answer 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]

This answer is based on first determining the revenue maximizing top tax rate via a widely-used formula. The coefficient a in the formula is a simple statistic of the earnings (or income) distribution beyond a threshold and ϵ is the elasticity of aggregate earnings (or income) beyond a 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.

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

There is much to admire about the combination of theory, data and policy relevance behind this answer. However, we highlight two main issues with this analysis and argue that both lead to a lower top tax rate. First, there is a theoretical problem: the widely-used formula is not valid in dynamic models. Fortunately, there is a formula (see Badel and Huggett (2017)), featuring three elasticities, that applies very widely to static models and to steady states of dynamic models. The Badel-Huggett formula is a generalization of the widely-used formula and is stated below. Clearly, if the two new forces $a_2\epsilon_2$ and $a_3\epsilon_3$ are positive then they reduce the value of the revenue maximizing top tax rate. We present US regression evidence showing that point estimates for each of these new elasticities (ϵ_2, ϵ_3) are positive.

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

Second, there is the issue of measuring the formula coefficients. We calculate that $(a_1, a_2, a_3) = (1.70, 3.64, 0.16)$ using 2010 US data. The Pareto statistic $a_1 = \bar{y}/(\bar{y} - \underline{y})$ equals $a_1 = 1.70$ when the threshold $\underline{y}$ is the 99th percentile of our income measure and $\bar{y}$ is mean income above this threshold. Diamond and Saez (2011) 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. However, AGI includes capital gains and qualified dividends that are taxed at preferential rates.[2] Our income definition excludes capital gains

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

[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 rate on ordinary income was 35 percent.

and qualified dividends. If one wants to determine the top tax rate on ordinary income that is revenue maximizing, holding all other tax rates fixed, then a_1 needs to be calculated excluding income sources taxed at preferential rates.

To provide perspective on the consequences of increasing the marginal tax rate on top earners, we analyze a human capital model that is calibrated to properties of the US age-earnings distribution, the structure of US marginal tax rates and the 2010 US value for the coefficient a_1 that enters the tax rate formula. The model tax reform increases the tax rate on the top 1 percent, holding other tax rates and government expenditures unchanged, to fund a lump-sum transfer to all agents. The steady state model Laffer curve peaks at a top tax rate τ of 49 percent.

We apply the tax rate formula to the model. As indicated below, the two new forces $a_2\epsilon_2$ and $a_3\epsilon_3$ are positive and depress the revenue maximizing tax rate. The elasticity ϵ_2 is positive when an increase in the net-of-tax rate $1-\tau$, applying to income beyond a threshold, increases the total tax revenue from agents below the income threshold. The elasticity ϵ_3 is positive when an increase in $1-\tau$ increases the tax revenue from agents above the threshold based on all other tax bases. In the model and the US economy, consumption expenditures and capital income sources with preferential tax rates form these other tax bases. Positive values for both of the new elasticities imply that when τ increases then tax revenues decrease from both of these revenue sources. Thus, total tax revenue is maximized when the revenue from the top income group, based on the tax base that τ applies to, is still increasing.

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

The formula allows for a simple decomposition. Substituting only the model values for the two new forces $a_2\epsilon_2$ and $a_3\epsilon_3$ into the formula, while fixing $a_1\epsilon_1$ at Diamond-Saez values, implies that $\tau^* = \frac{1 - 4.508 \times .040 - .120 \times .739}{1 + 1.5 \times .25} = .532$. Thus, the two new forces account for the bulk of the movement from a 73 percent revenue maximizing top tax rate to the model's revenue maximizing rate of 49 percent.

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 in comparison to a permanent change in the log net-of-tax rate 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 the log net-of-tax rate applying to the top 1 percent of US tax units. We present regression evidence that the point estimates for $(\epsilon_1, \epsilon_2, \epsilon_3)$ are positive, consistent with the implications of the human capital model.

To what degree is endogenous skill accumulation important for the model's revenue maximizing top tax rate and lump-sum transfer? To answer this, an observationally-equivalent model is constructed where skills do not respond to a change in the top tax rate but evolve in the same way as in the human capital model. The revenue maximizing top tax rate $\tau^*_{exog} = 0.59$ is larger in the exogenous-skill model. This occurs because skills of top earners fall when the top tax rate increases in the human capital model but are unaffected in the exogenous-skill model. All three elasticities in the formula are smaller in the exogenous-skill model than those in the human capital model while the coefficients are exactly the same. The two new forces are again central for why the revenue maximizing rate in this model is well below the 73 percent value highlighted by Diamond and Saez (2011).

$$\tau^*_{exog} = \frac{1 - a_2\epsilon_{2,exog} - a_3\epsilon_{3,exog}}{1 + a_1\epsilon_{1,exog}} = \frac{1 - 4.508 \times .020 - .12 \times .633}{1 + 1.7 \times .222} \doteq 0.59$$

What is the mechanism by which the skills of top earners fall? An increase in the top rate decreases the marginal benefits of skill investment that are recieved 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 steady-state skills. For this logic to hold, top earners must have upward sloping earnings profiles. Earnings growth rates over the working lifetime are strikingly large for top US lifetime earners. The model produces strikingly large earnings growth rates for top lifetime earners.

The paper is organized as follows. Section 2 presents the model framework. Section 3 documents properties of the US age-earnings distribution, marginal tax rates and the tax formula coefficients. Section 4 and 5 describe model properties. Section 6 analyzes the model Laffer curve and discusses regression evidence. Section 7 concludes.

Related Literature

The paper is related to the literature on sufficient statistic formulae. 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.

The paper is related to the elasticity literature. This literature, reviewed by Saez, Slemrod and Giertz (2012), focuses on the response of income measures to changes in a tax rate after a tax reform. Our paper examines how well standard regression frameworks from this literature estimate elasticities in dynamic models. This can be done because model elasticities are defined and calculated separately from any regression framework. Our paper is the first to estimate the two new elasticities entering the Badel-Huggett formula.

The human capital model we employ builds on the model in Huggett, Ventura and Yaron (2011) but adds valued leisure. Heckman, Lochner and Taber (1998), Erosa and Koreshkova (2007) and Guvenen, Kuruscu and Ozkan (2014) also analyze income tax progression using human capital models. Unlike our paper, none of these papers focus on tax reforms directed at the upper tail. Altig and Carlstrom (1999), Guner, Lopez-Daneri and Ventura (2016), Kindermann and Krueger (2014) and Brueggemann and Yoo (2015) analyze tax reforms that are directed at the upper tail. Unlike our paper, they do not allow an agent's labor productivity or skill to respond to a tax reform. Blandin (2016) analyzes the elimination of the cap on social security earnings taxation in the US and finds that this reduces steady-state skill accumulation. The economic mechanism underlying his results is the same as the mechanism that we highlight.

2 Framework

This section presents a model and a revenue maximizing top tax rate formula.

2.1 Model

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

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

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.

$$
\begin{aligned}
K &= \sum_{j=1}^{J} \mu_j \int_{\hat{X}} E[k_j(\hat{x}, z^j)|\hat{x}]d\psi \text{ and } L = \sum_{j=1}^{J} \mu_j \int_{\hat{X}} E[h_j(\hat{x}, z^j) l_j(\hat{x}, z^j)|\hat{x}]d\psi \\
T &= \sum_{j=1}^{J} \mu_j \int_{\hat{X}} E[T_j(wh_j(\hat{x}, z^j) l_j(\hat{x}, z^j), c_j(\hat{x}, z^j), rk_j(\hat{x}, z^j)|\hat{x}]d\psi
\end{aligned}
$$

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: $C + K(n + \delta) + G = F(K, L)$.*

2.2 Tax Rate Formula

Badel and Huggett (2017, Theorem 1) derive a formula that states the revenue maximizing top tax rate τ^* in terms of three elasticities. Their formula applies widely to static models and to steady states of dynamic models, whereas the widely-used formula applies only to some static models.

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 Badel-Huggett formula is stated below. The aggregate 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.[3] 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.

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

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

2.3 Applying the Formula

The Badel-Huggett formula is not stated in terms of the primitives of a specific economic model. This gives it wide application. To apply the formula to the human capital model, map equilibrium model features into the three elements used in the formula. We do so in three steps. First, an agent type in the model is $x = (h_1, a, j, z^j)$ - a quadruple of initial skill h_1, learning ability a, age j and (partial) shock history z^j.[4] Second, define y_1 as labor income, y_2

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

[4]The probability measure P over agent types in the formula is constructed from the distribution of initial conditions, the exogenous shock process and the fractions μ_j of each age group in the population.

as consumption, y_3 as capital income and y_4 as social security transfers.[5] Third, the function T is determined by the model tax system T_j in Problem P1. T_j will later feature a progressive tax T^{prog} on labor income with top tax rate τ, proportional tax rates (τ_c, τ_k) on consumption and capital income and a lump-sum social security transfer. Section 6.3 calculates all model coefficients and elasticities in the formula.

$$y_1(x,\tau) \equiv w(\tau)h_j(h_1,a,z^j;\tau)l_j(h_1,a,z^j;\tau) \text{ for } j < Retire \text{ and 0 otherwise}$$

$$y_2(x,\tau) \equiv c_j(h_1,a,z^j;\tau)$$

$$y_3(x,\tau) \equiv r(\tau)k_j(h_1,a,z^j;\tau)$$

$$y_4(x,\tau) \equiv transfer \text{ for } j \geq Retire \text{ and 0 otherwise}$$

$$T(y_1,y_2,y_3,y_4;\tau) \equiv T^{prog}(y_1;\tau) + \tau_c y_2 + \tau_k y_3 - y_4$$

3 Empirics

We address four issues. 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? Is there US evidence for the elasticities (ϵ_2, ϵ_3)? Answers to the first three issues 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 of earnings and (iv) the mean fraction of time spent working.[6] 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.

[5]The notation employed in defining (y_1, y_2, y_3, y_4) emphasizes that equilibrium factor prices and decisions depend on the top tax rate τ.

[6]The Pareto statistic is $\frac{\bar{y}}{\bar{y}-\underline{y}}$, where $\underline{y}$ is a threshold and $\bar{y}$ is mean earnings for all observations above this threshold. The Pareto statistic is analyzed because it enters the tax rate formula and, thus, disciplines the model in a theoretically relevant way.

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.[7] 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.[8]

3.2 Tax Function

TAXSIM is used to characterize marginal tax rates in 2010.[9] 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.[10] The marginal tax rate schedule tends to increase with income. It jumps at thresholds where federal income tax brackets increase and 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 $T^{prog}(e;\tau)$ in section 2.3, is constructed by integrating the model marginal tax rate function.

[7] 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 tme dummy variables rather than time dummies and an age polynomial.

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

[9] 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).

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

3.3 Tax Formula Coefficients

The coefficients (a_1, a_2, a_3) are calculated using US data and an accounting framework. The framework divides the number of tax units N into those N_1 with income above the 99th percentile threshold and the remainder N_2. The framework divides an aggregate tax measure Tax into components: N_2Tax_2 is the part paid by tax units below the threshold and $N_1Tax_1 + N_1Tax_3$ is the part paid by units above the threshold. The latter component is subdivided into N_1Tax_1 that is in theory determined by income source "y_1" and into N_1Tax_3 that is based on taxes on all other incomes or expenditures.

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

The coefficients, defined in section 2.2, are easy to restate using the accounting framework. Let $\underline{y}$ denote the threshold and $\bar{y}$ denote mean income per tax unit beyond the threshold. The coefficient a_2 is the ratio of total taxes paid by units below the threshold to a measure of income above the threshold. Similarly, a_3 is the ratio of all the other taxes paid by units above the threshold to a measure of income above the threshold.

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

Table 1 Panel (d) calculates that $(a_1, a_2, a_3) = (1.70, 3.64, 0.16)$. To understand the logic behind these calculations, start with Panel (a). Tax is an aggregate measure of US taxes that is the sum of personal taxes, social insurance taxes and taxes on production.[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. All aggregate measures are based on 2010 Bureau of Economic Analysis data.

Panel (b) calculates the part of each of the three aggregate tax measures that is paid by the top 1 percent based on Statistics of Income (SOI) data. 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 aggregate 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. Capital income taxes are subtracted from personal taxes because, for tax units with high income, qualified dividends and long-term

[11] Our economic model and our empirical analysis abstracts from 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 methods and all the data sources employed.

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.

Capital income taxes paid by the top 1 percent are calculated in Panel (c) in three steps. First, specify the income measure based on SOI income categories.[12] Second, calculate $(\underline{y}, \bar{y})$, the 99th percentile of this income measure and the mean beyond this percentile. Panel (d) states the result. Third, *capital income tax* equals the sum of all the qualified dividends and capital gains received by tax units with income beyond the 99th percentile multiplied by the tax rate $\tau_k = 0.20$ calculated in section 4 based on federal and state tax rates in 2010.

3.4 Elasticities: US Evidence

The elasticities $(\epsilon_1, \epsilon_2, \epsilon_3)$ in the tax rate formula involve how aggregate income and tax revenue measures 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

[12]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.

permanent change in $\log(1-\tau)$ within the human capital model. This section estimates the two new elasticities (ϵ_2, ϵ_3) using US data and a statistical model that allows for a permanent shift in the net-of-tax rate. The statistical model captures, in a simple way, 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 measure used in Table 1. Aggregate taxes are decomposed following the logic of the formula: $Tax_t = Rev_{1t} + Rev_{2t} + Rev_{3t}$.[13] 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. The income definition and measurement are described in Appendix B.

Figure 3 plots the revenue series and the average marginal tax rate for the top 1 percent of tax units calculated by Mertens and Montiel-Olea (2017).[14] Much of the tax rate variation in Figure 3 appears to be transitory; however, the tax rate declined in a dramatic and roughly permanent fashion in the mid 1980s.

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

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.[15] We set $T = 1987$ 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.

---

[13] 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.

[14] 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.

[15] 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.

Table 2 - 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)
$\log Rev_3$	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)
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	$\epsilon_3 = \eta_2/\alpha_2$	-0.380	0.081	0.223	-0.449	0.001	0.218
		(0.673)	(0.609)	(0.865)	(0.673)	(0.633)	(0.908)
	Time Dummies	None	1986-87	1986-90	None	1986-87	1986-90
	N. Obs.	41	41	41	41	41	41

Note: The dependent variable is log revenue in columns (1)-(3) and log revenue per tax unit in columns (4)-(6). 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.

Table 2 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 in equation (1). Observations for 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). Columns (4) to (6) measure the dependent variable as (log) tax revenue per tax unit rather than (log) total tax revenues but otherwise repeat the estimation of columns (1) to (3).

The estimate for α_2, governing the shift in the net-of-tax rate, is positive. 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). The elasticity estimates in columns (4) to (6) mimic the signs and magnitudes of the those from columns (1) to (3).

What motivates adding time dummies? The literature documents that capital gains are especially responsive to anticipated changes in the applicable tax rate. Burman, Clausing and O'Hare (1994) document that capital gains realizations dramatically surged late in 1986 after key steps in the signing of the Tax Reform Act of 1986 but before the higher capital gains tax rate took effect in 1987. They argue that the spike in capital gains realizations in 1986 and the collapse in 1987 is due to the change in the capital gains rate, specifically the higher future rate, and not the change in the top 1 percent tax rate documented in Figure 3.

A separate argument for adding time dummies, or for modifying the benchmark statistical model, is that it is too parsimonious in that transitional dynamics are not allowed. All effects of the hypothesized permanent tax change are instantaneous even though transitional dynamics occur in the human capital model. Adding time dummies for years after a reform is a means for addressing, in a simple way, the possible importance of transitional dynamics.

Our view is that the elasticity estimates in columns (2)-(3) and (5)-(6) of Table 2 are the most relevant. Some means of accounting for the dramatic variation in capital gains in 1986 and 1987 is needed as the leading explanation is that it is due to variation in the capital gains tax rate. Point estimates for ϵ_2 and ϵ_3 are positive based on these columns.

Piketty and Saez (2003) and Saez and Zucman (2016) provide a narrative argument for a positive value for ϵ_3.[16] They document a strong U-shaped pattern for the top 1 percent income and wealth shares in the US over the last hundred years. An important 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 is a force bolstering tax revenues, measured in the variable Rev_{3t}, and a force behind a positive value for ϵ_3.[17]

The elasticity ϵ_3 can in theory be restated as the revenue-weighted average of the elasticities of its two components: taxes on production and capital income taxes paid by the top 1 percent. We estimate the elasticity of these two components $(\epsilon_{3,1}, \epsilon_{3,2})$ using system (2). System (2) modifies system (1) 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.

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

Figure 4 graphs the two new time series. It 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.

Table 3 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

[16] 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."

[17] Kaymak and Poschke (2016) and Hubmer, Krusell and Smith (2016) employ quantitative-theoretical models 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. These models provide quantitative-theoretical support for this narrative that is consistent with the implications of the human capital model.

Table 3 - 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 2.

negative point estimate for ϵ_3 in Table 2, for the case without time dummies, is driven entirely by a large, negative elasticity of capital income taxes paid by the top 1 percent.

4 Model Parameters

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.[18]

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 $\epsilon \sim LN(0, \sigma_\epsilon^2)$

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

Table 4 - 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	$(A, \gamma, \delta) = (0.411, 0.352, 0.044)$
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.

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.

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.[19]

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 (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.[20] 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 $\alpha_1 = 0.125$ 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 α_1 is 0.125 based on MaCurdy (1981) who uses data for white males age 25-55.[21]

[19] 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

[20] Appendix B5 specifies the objective that is minimized and the implemention of the minimization.

[21] This is the average from MaCurdy (1981, Table 1 row 5-6). Altonji (1986) finds similar results. Domeij and Floden (2006, Table 5) estimate $\alpha_1 = 0.16$ for male household heads age 25-60 using 1984-94 PSID data. Keane (2011) and Keane and Rogerson (2012) review this literature. MaCurdy (1981) argues that within exogenous-wage models the regression coefficient α_1 is an estimate of the preference parameter ν under appropriate conditions.

$$\Delta \log hours = \alpha_0 + \alpha_1 \Delta \log wage + \epsilon$$

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 and the instrumental variable methods employed. It also discusses reasons why the model regression coefficient is below the model value of ν. Disciplining the model to match the empirical regression coefficient α_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 Saez, Slemrod and Giertz (2012 p. 3-4)) that a very high top tax rate may be revenue maximizing.

Figure 5 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 $\alpha_1 = 0.125$ from US data while it is $\alpha_1 = 0.124$ in the model. The model parameters that produce these results are stated in Table 4.

5 Properties of the Model Economy

5.1 Age-Earnings Distribution

Figure 5 highlights model properties. The model produces a hump-shaped median earnings profile by a standard human capital mechanism. Agents concentrate learning time and human capital production early in the working lifetime. Towards the end of the working lifetime, both the median and the mean human capital levels fall. This occurs because time allocated to learning goes to zero towards the end of the working lifetime and because the mean of the multiplicative shock to human capital is below one (i.e. $E[exp(z)] = exp(\mu_z + \frac{\sigma_z^2}{2}) < 1$). Thus, on average skills depreciate.

Earnings dispersion measures increase with age in U.S. data. Specifically, Figure 5 shows that the 99-50 earnings ratio increases with age and that the Pareto statistic at the 99th percentile decreases with age. The model economy has two forces leading to increasing earnings dispersion: differences in learning ability and idiosyncratic shocks. 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 and hours facts. Higher learning ability, other things equal, rotates an agent's mean earnings profile counter clockwise because higher learning ability increases the marginal benefits derived from time investment early in life.

5.2 Distribution of Initial Conditions

We highlight two properties of the distribution of initial conditions. The first property is that there is dispersion in learning ability. In fact, the distribution which best fits the data has learning ability approximately following a Pareto distribution. Thus, the model requires a source for increasing earnings dispersion with age, beyond that coming from idiosyncratic risk, to produce the increase in the 99-50 earnings ratio and the decrease in the Pareto statistic with age observed in U.S. data.

The second property is that learning ability and human capital are positively correlated at age 23. This is implied by the fact that the model parameter β_1 is positive in Table 4. The positive correlation is consistent with Huggett, Ventura and Yaron (2006, 2011) who argue that a zero correlation would tend to produce a strong U-shaped earnings dispersion pattern with age not found in U.S. data. The positive correlation has two important implications: (i) agents with high lifetime earnings will tend to have high learning ability and (ii) agents with high lifetime earnings will have high average earnings growth rates over the working lifetime as learning ability is a key factor governing this growth rate. Section 6.5 of the paper will later examine earnings growth over the lifetime for top lifetime earners in US data and in the model.

6 Analyzing the Tax Reform

6.1 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 well 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 displays Laffer curves when the utility function parameter ν is set to $\nu = 0.5$ and to $\nu = 0.4$ rather than $\nu = 0.614$ in the benchmark model. For each of the values of ν

the estimation procedure is repeated, choosing the remaining model parameters to best match targets.[22] The revenue maximizing top tax rate increases as the parameter ν decreases and the revenue-maximizing transfer increases as the parameter ν decreases.

The steady state welfare consequences of increasing the top tax rate are measured as the 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 to the ex-ante expected utility achieved at alternative values of the top tax rate. This calculation is made before agents know their initial conditions. Figure 6 shows that within the benchmark model there is a very small welfare loss of 0.037 percent of consumption from increasing the top tax rate to the revenue-maximizing level. The welfare loss is driven by offsetting effects. While there is a redistributional benefit arising from the lump-sum transfer, there is also a small fall in the wage w. It turns out that factor prices change very little in percentage terms 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 in different directions for age 1 agents, conditional on an agent's initial conditions. Agents with high learning ability experience welfare losses. These agents will be directly impacted later in life when they cross the threshold at which the higher top tax rate applies. For the group of age 1 agents within the top 1 percent of the learning ability distribution, 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.

6.2 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.

In the exogenous human capital model, the time investment decision $s_j(\hat{x}, z^j)$ as a function of initial condition $\hat{x} = (h_1, a)$, age j and shock history z^j are fixed and do not vary as the tax system changes. These decisions are set equal to those in the human capital model under the benchmark tax system. Thus, individual human capital evolves in exactly the same way in the

[22] The targets and the objective function are the same as were used in choosing model parameters in Table 4.

exogenous human capital model regardless of the tax system.[23]

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. More tax revenue can be raised in the exogenous human capital model by raising the top tax rate as compared to the human capital model. 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.[24] The largest percentage fall in skills occurs at the end of the working lifetime from agents endowed with high learning ability.[25]

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

The mechanism behind the fall in skill is easily grasped from the Euler equation governing skill investment above. 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

[23]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.

[24]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.

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

top earners later in life.

6.3 Applying the Tax Rate Formula

The formula can be applied in several useful ways. It provides insight into the proximate reasons for why the Laffer curves have different revenue maximizing top tax rates in the human capital and exogenous human capital models. It also determines the proximate reasons for why the human capital model Laffer curve has a much smaller revenue maximizing top rate than the 73 percent rate suggested by Diamond and Saez (2011).

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. Define an agent type as $x = (h_1, a, j, z^j)$. Calculate the set X_1 of top earner types. 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 respective Laffer curves. Model social security transfers are not included in net taxes so that model calculations of (a_1, a_2, a_3) parallel the US data calculation in Table 1.

Table 5 calculates the three model coefficients and elasticities in both models. It shows that the coefficients are the same in both models. They are the same because the distribution of incomes, expenditures and taxes at the initial steady state is exactly the same in both models and because the coefficients are functions of this joint distribution. It also shows that exogenous human capital lowers all three elasticities. The economic mechanism behind the lower elasticity ϵ_1 in comparison to that in the human capital model was articulated in section 6.2. Lastly, Table 5 shows that differences across models in the elasticities (ϵ_2, ϵ_3) accounts for roughly half of the difference in the top tax rate across models.

Why does the Laffer curve for the human capital model peak at a lower tax rate than the 73 percent rate suggested by Diamond and Saez (2011)? Table 5 gives a concise answer. First, there are generally two new forces, captured by the terms $a_2\epsilon_2$ and $a_3\epsilon_3$, that need to be considered that are not accounted for in the formula $\tau^* = 1/(1 + a\epsilon)$ employed by Diamond and Saez (2011). These new forces are positive within the human capital model and depress the revenue maximizing top tax rate. Second, the model Pareto statistic is $a_1 = 1.70$, whereas Diamond and Saez employ a value of 1.5. Third, the model elasticity $\epsilon_1 = .317$ is slightly larger than the value 0.25 used by Diamond and Saez (2011).

How important are the two new forces in accounting for the difference with respect to the 73 percent rate implied by the Diamond-Saez analysis? This question is answered by plugging into the Badel-Huggett formula the model values for the two new forces $a_2\epsilon_2$ and $a_3\epsilon_3$ but setting $(a_1, \epsilon_1) = (1.5, .25)$ to the Diamond-Saez values. The formula implies that the revenue maximizing top rate goes from .727 to .531. Thus, the two new forces account for most of the movement from 73 percent to the 49 percent revenue maximizing top tax rate.

Section 3.4 of the paper provided US regression evidence for these new elasticities. When these elasticities are positive, increasing the top tax rate on the top 1 percent on ordinary income results in a decrease in the tax revenue coming from the bottom 99 percent and a decrease in tax revenue coming from the top 1 percent based on consumption-related taxes and on capital income sources with preferential tax rates. This implies that total tax revenue is maximized when the revenue from the top income group, based only on the tax base that τ applies to, is still rising.

Although we argue that the two new forces are quantitatively the most important, it is still useful to explain what accounts for why we set model parameters 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 lower, preferential rates. Based on 2010 Statistics of Income (SOI) data, we calculate in Table 1 that $a_1 = 1.70$ when income excludes qualified dividends and capital gains. We also calculate, using AGI as an income measure, that $a_1 = 1.50$ at the 99th percentile based on 2010 SOI data. 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 these years. Instead, it is that the upper tail is thinner after excluding capital income sources that are taxed at preferential rates.

6.4 Income Elasticities: Interpreting US Evidence

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 occuring 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 8 plots the transition path arising from the surprise reform. Aggregate capital and labor input per agent fall by a similar percentage.[26]

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 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 6 - 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
Model Elasticity ϵ_1	0.317	0.317	0.317	0.317	0.317	0.317

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 6 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 6 were employed in SSG. SSG (2012, Table 2, Panel B1) 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. Table 6 shows that, using model data for all seven time periods, the mean of the model point

[26] 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.

estimates range from a low of 0.37 to a high of 0.47.[27]

We view the results in Table 6 as a model validation exercise. Broadly, the model regression coefficients are consistent with the range of results documented by SSG. Thus, the economic model is validated as the short-term model responses are not restricted to match US short-term response evidence.[28]

6.5 Earnings Growth: US and International Evidence

Skill accumulation falls in the model after a permanent increase in the top tax rate. This fall in steady-state skills is largest late in the working lifetime for agents who enter the model with high learning ability. The logic behind this, see section 6.2, relies on upward sloping earnings profiles.

Figure 9 plots the relationship between lifetime earnings for individual males in SSA panel data and the growth rate of earnings based on the work of Guvenen, Karahan, Ozkan and Song (2015). Males are divided into 100 equal-sized bins based on percentiles of lifetime earnings. Figure 9 plots the ratio of mean real bin earnings at age 55 to mean real bin earnings at age 25 and at age 30. Figure 9 shows that earnings grow dramatically over the working lifetime for the top several bins of lifetime earnings. For example, earnings in the top bin increases by a factor of 15 between age 25 and 55 and by a factor of 7 between age 30 and 55.

Figure 9 shows that the human capital model displays the qualitative patterns and to a degree the quantitative patterns observed in US data for top lifetime earners when model agents are put into 100 bins following the procedure used on US data. These aspects of US data are not targeted in the model calibration. 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. Top lifetime earners are disproportionally those entering the model with high learning ability - see section 5.2. The dispersion in learning ability is disciplined to match the rise in the 99-50 earnings percentile ratio by age and the decrease in the Pareto statistic by age documented in Figure 1, taking as given the estimate of idiosyncratic risk from Huggett, Ventura and Yaron (2011).

Badel, Daly, Huggett and Nybom (2017) use administrative earnings data for males from Denmark and Sweden and apply the same data construction procedure used on US data. Figure 9 shows that earnings ratios over the working lifetime are strikingly large for the top several bins of lifetime earnings in all countries. Thus, theories of top earners will broadly need to account

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

[28] Appendix B examines US long-run response estimates and runs the same regressions on model data. The estimated short-term model responses reported in Table 6 vary somewhat with the details of the tax experiment.

for this empirical regularity. Badel et al. (2017) demonstrate that quantitative-theoretical models built on shocks with only a mean-reverting or a purely transitory impact on earnings fail to account for this striking empirical regularity.

7 Conclusion

How should the tax rate on top earners be set? The established answer to this question involves first determining the revenue maximizing top tax rate. The established answer is based on a static model - the Mirrlees model - and a simple formula that expresses this top rate in terms of one coefficient and one elasticity. The main quantitative conclusion from this work is that the revenue maximizing top tax rate in the US is roughly 73 percent and is well above the top rate of roughly 42 percent holding in the US in 2010.

We challenge this answer at a theoretical level and at a quantitative level. At a theoretical level, we provide a more general top tax rate formula that applies to static models, such as the Mirrlees model, and to steady states of dynamic models. The more general formula makes it clear that economists must analyze two new forces, in addition to the one standard force emphasized in previous theoretical and empirical work.

At a quantitative level, we provide a number of reasons for why the revenue maximizing top tax rate in the US is likely to be below 73 percent. First, we provide a quantitative-theoretical perspective and a reduced-form empirical perspective on the importance of the two new forces. The two new forces are reflected in positive elasticities for ϵ_2 and ϵ_3 within a quantitative human capital model. Thus, there are natural theoretical mechanisms behind positive values for each of these new elasticities. These two new forces are the main reason for why the model's revenue maximizing tax rate $\tau^* = 0.49$ is well below the value $\tau^* = 0.73$ from the established answer. At an empirical level, we present US regression evidence that the point estimates for both of the new elasticities are positive. Positive values imply, other things equal, a lower revenue maximizing top tax rate according to the Badel-Huggett formula.

A positive value for the elasticity ϵ_3 is consistent with a popular narrative. Piketty and Saez (2003 p. 37) state *"... 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."* Saez and Zucman (2016) document a U-shaped pattern for the top 1 percent wealth share in the last 100 years, with a strong increase after roughly 1980. This suggests that capital income taxes paid by top earners increase when the net-of-tax rate on ordinary income increases, other things equal.

Second, we measure that the US Pareto statistic at the 99th percentile of income in 2010 is

$a_1 = 1.7$ rather than the value $a_1 = 1.5$ used in previous work. This is a relevant empirical value when the goal is to determine the revenue maximizing top tax rate on ordinary income, holding other tax rates constant including the tax rate on capital gains and qualified dividends. A higher value of the Pareto statistic implies, other things equal, a lower revenue maximizing top tax rate.

The human capital model is consistent with a number of facts that theory says are relevant and that the literature has viewed as being relevant for determining the revenue maximizing top tax rate. For example, the model exactly matches the US Pareto statistic $a_1 = 1.70$ and approximates the US values for the coefficients (a_2, a_3). Theory dictates that these are key. The model approximates both the Frisch labor hours elasticity and the (short-term) income elasticity to the net-of-tax rate as estimated in the labor and public economics literatures. Both of these elasticities have been viewed as being highly relevant for the issue at hand - see Keane (2011) and Saez et al. (2012). Finally, the human capital model produces the strikingly large earnings growth of top US lifetime earners.

References

Altig, D. and C. Carlstrom (1999), Marginal Tax Rates and Income Inequality in a Life-Cycle Model, American Economic Review, 89, 1197-1215.

Altonji, J. (1986), Intertemporal Substitution in Labor Supply: Evidence from Micro Data, Journal of Political Economy, 94, 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, 87, 1-12.

Badel, A., Daly, M., Huggett, M. and M. Nybom (2017), Top Earners: Cross-Country Facts, HCEO Working Paper 2017- 061.

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.

Brueggemann, B. and J. Yoo (2015), Aggregate and distributional effects of increasing taxes on top income earners, IMFS Working Paper Series 94, Goethe University Frankfurt, Institute for Monetary and Financial Stability (IMFS).

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

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

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

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

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

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

Guner, N., Lopez-Daneri, M. and G. Ventura (2016), Heterogeneity and Government Revenues: Higher Taxes at the Top?, Journal of Monetary Economics, 80, 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, 81, 818-50.

Guvenen, F., Ozkan, S. and J. Song (2014), The Nature of Countercyclical Income Risk, Journal of Political Economy, 122, 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, 13, 15-51.

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

Heckman, J, Lochner, L. and C. Taber (1998), Tax Policy and Human Capital Formation, American Economic Review, Papers and Proceedings, May 1998.

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 (2006), Human Capital and Earnings Distribution Dynamics, Journal of Monetary Economics, 53, 265- 90.

Huggett, M., Ventura, G. and A. Yaron (2011), Sources of Lifetime Inequality, American Economic Review, 101, 2923-2954.

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, 77, 1-25.

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

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

Kindermann, F. and D. Krueger (2014), High Marginal Tax Rates on the Top 1% ? Lessons from a Life Cycle Model with Idiosyncratic Income Shocks, manuscript.

MaCurdy, T (1981), An Empirical Model of Labor Supply in a Life-Cycle Setting, Journal of Political Economy, 89, 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, 38, 175- 208.

Piketty, T. and E. Saez (2003), Income Inequality in the United States, 1913-1998, Quarterly Journal of Economics, 118, 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, 68, 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, 50:1, 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, 131, 519-78.

A Appendix

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.[29] 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.

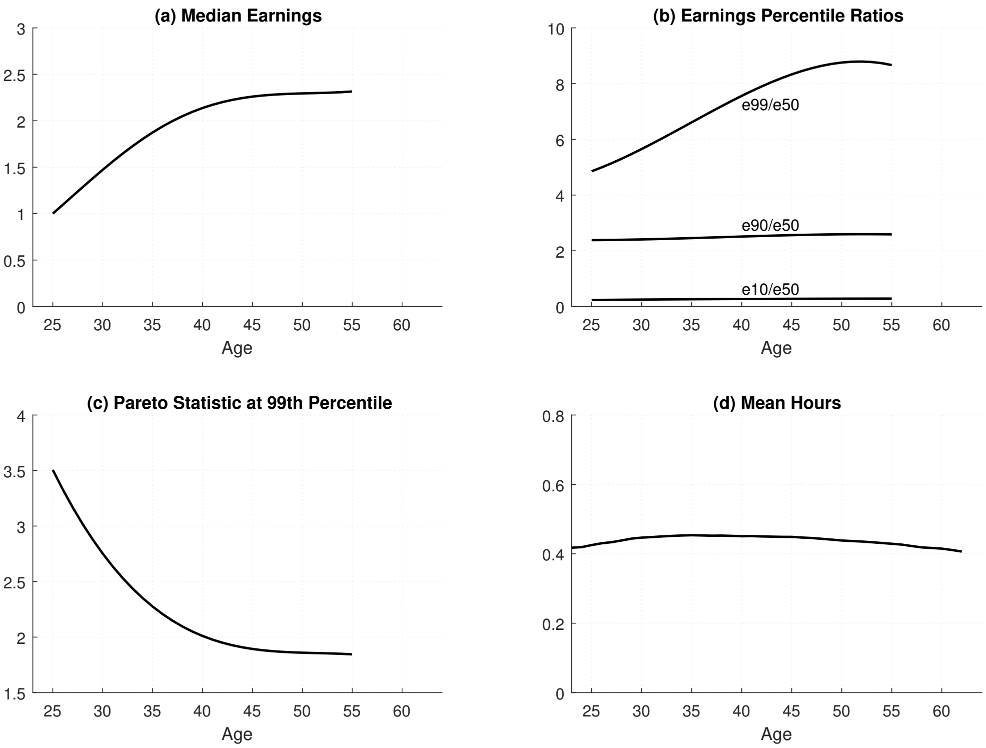
[29] SOI Tables: https://www.irs.gov/uac/soi-tax-stats-individual-income-tax-returns-publication-1304-complete-report

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.

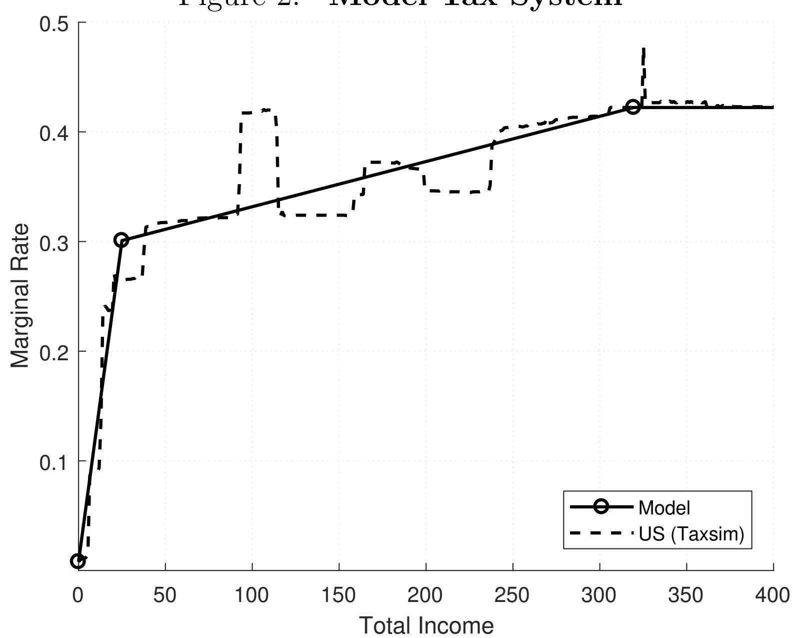
Step 4: Compute the entries in Panel (a) of Table 1. Personal Taxes are BEA Table 3.1 line 3. Social Insurance Taxes are BEA Table 3.6 line 4 plus line 22. Social insurance taxes are based only on the employer and employee parts of social security and medicare taxes. Taxes on Production are BEA Table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

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



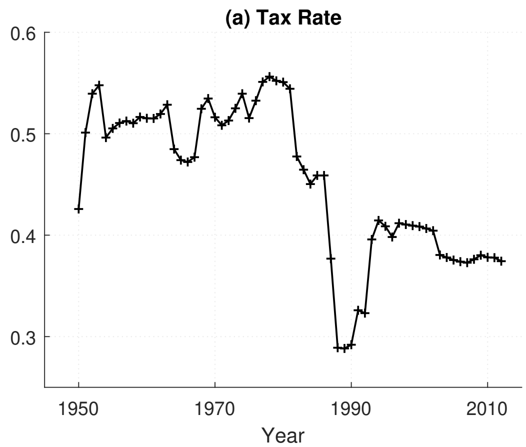
Note: Earnings profiles are based on SSA data. Hours profiles are based on PSID data.

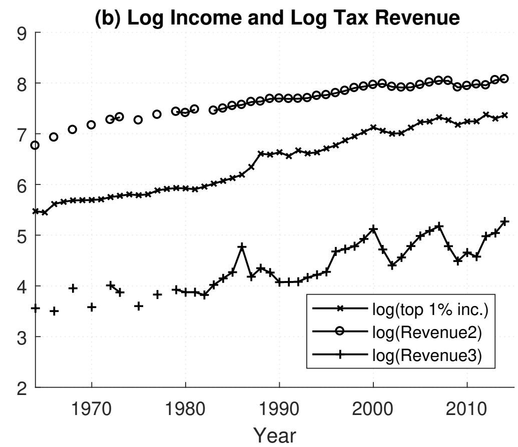
Figure 2: **Model Tax System**



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

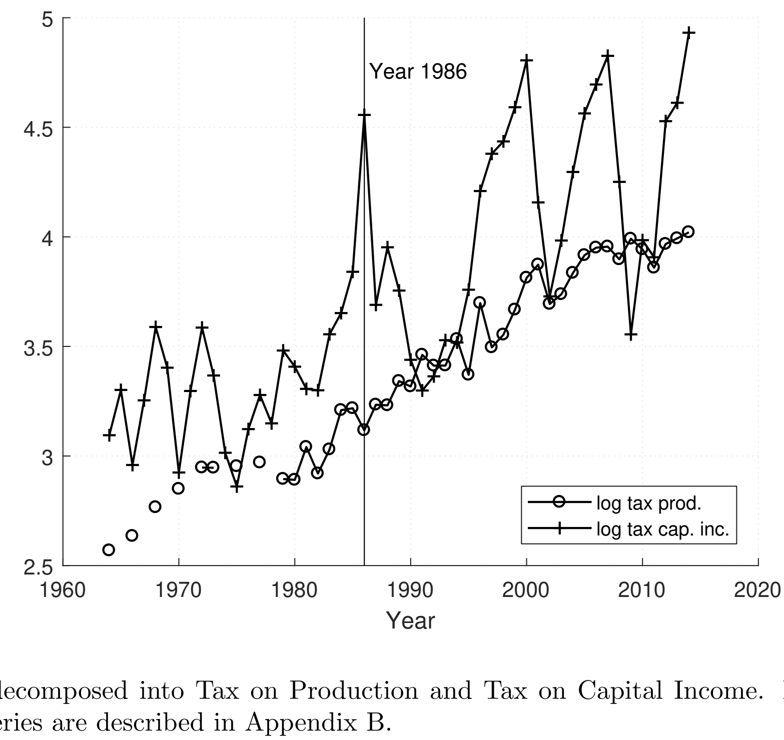
Figure 3: **Top 1 Percent Tax Rate, Log Income and Log Tax Revenue**





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 4: **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.

Figure 5: **Life-Cycle Profiles: Data and Model**

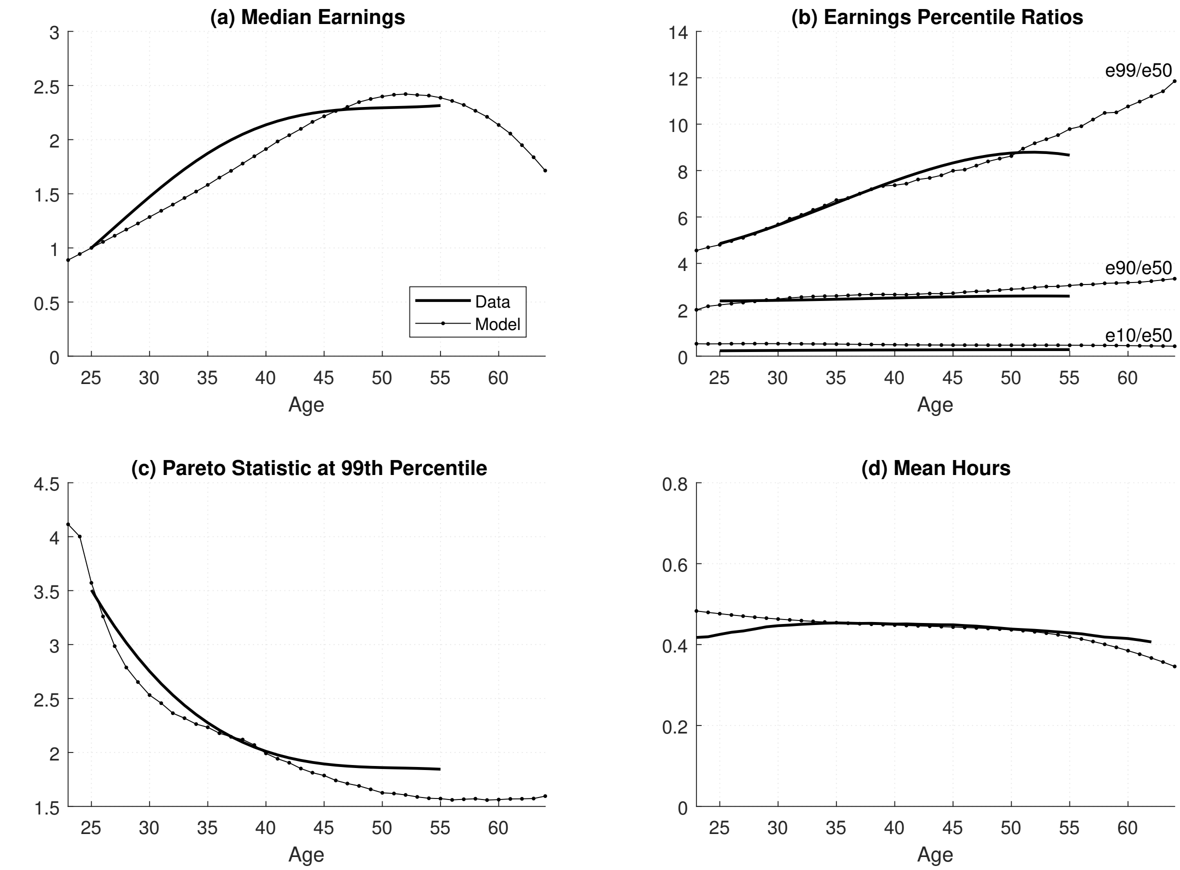
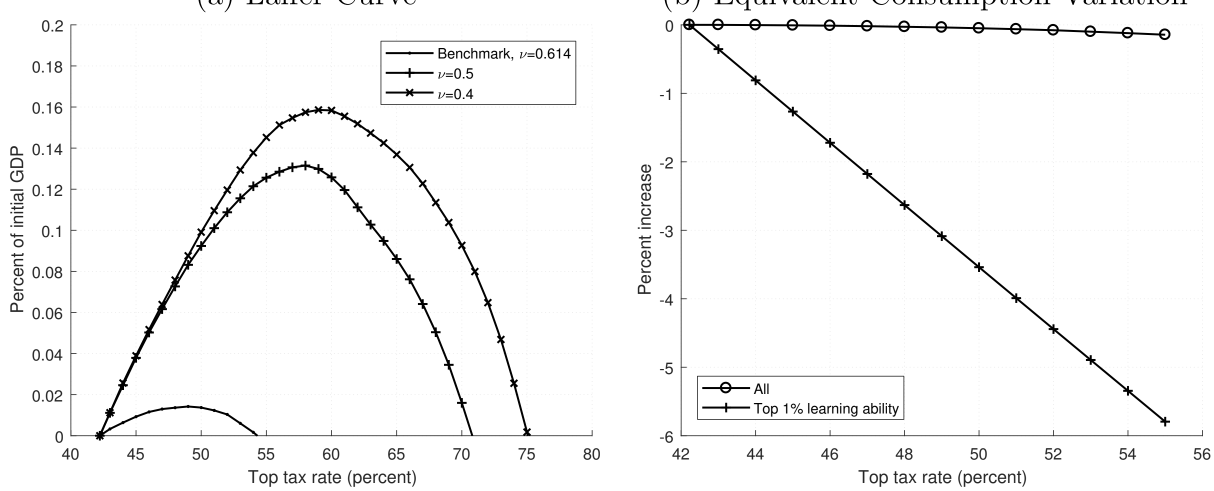


Figure 6: **Model Laffer Curves and Equivalent Consumption Variation**

(a) Laffer Curve (b) Equivalent Consumption Variation



Note: The Laffer curve plots the aggregate, lump-sum transfer per agent that results from the higher top tax rate as a percent of initial output per agent. The Laffer curve is plotted for the benchmark model with $\nu = .614$ and for models with $\nu = 0.5$ and 0.4. Consumption equivalents are calculated for the benchmark model.

Figure 7: **Laffer Curves**

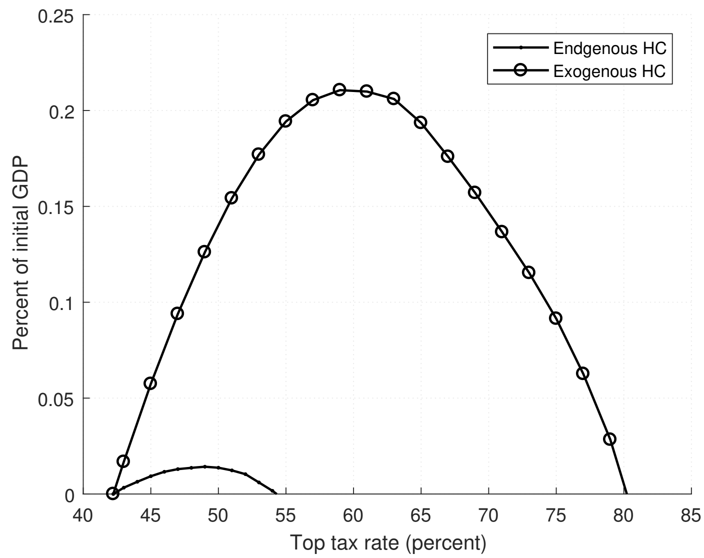
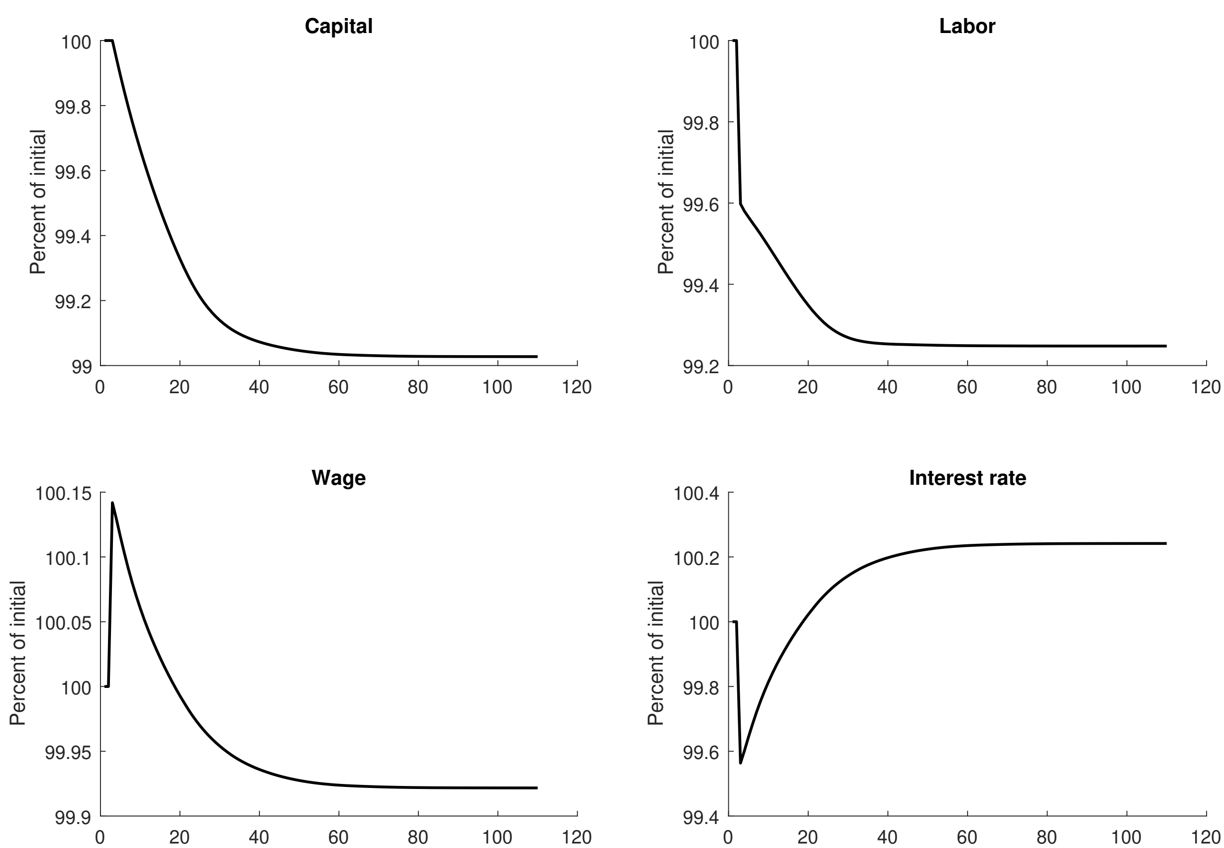
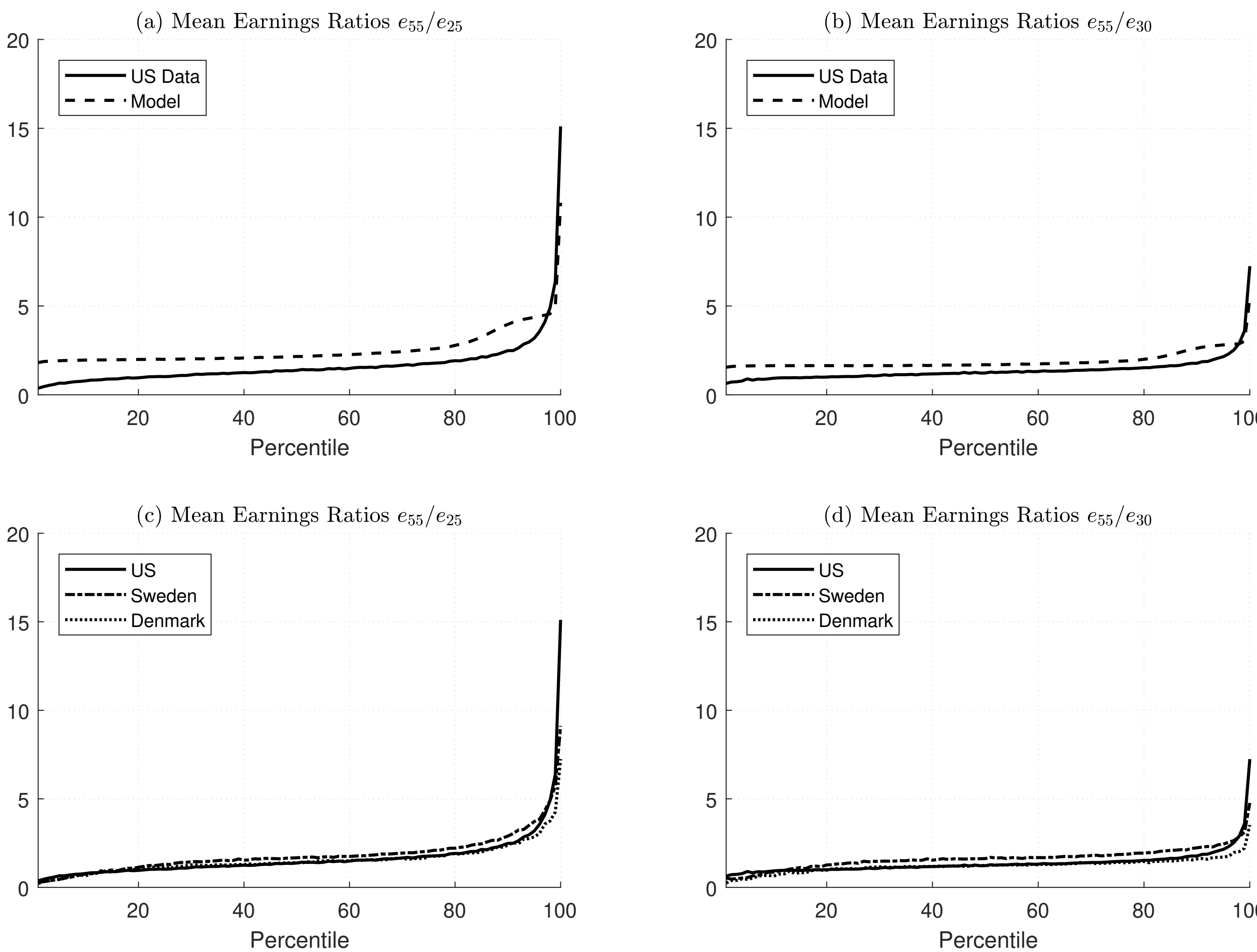


Figure 8: **Model Transition Path**



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 9: **Earnings Ratios By Lifetime Earnings Percentiles**



Note: US facts are from Guvenen et al. (2015). Facts for Denmark and Sweden are from Badel, Daly, Huggett and Nybom (2017).

B Appendix - NOT FOR PUBLICATION

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 $(\bar{y}_1, \bar{y}_2, p_1, p_2)$.

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 maximium 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 where there is no 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.[30]

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

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 set 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.[31]

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.[32] Third, some of the information needed to carry out steps 1.1 to 1.8 is not available is 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.

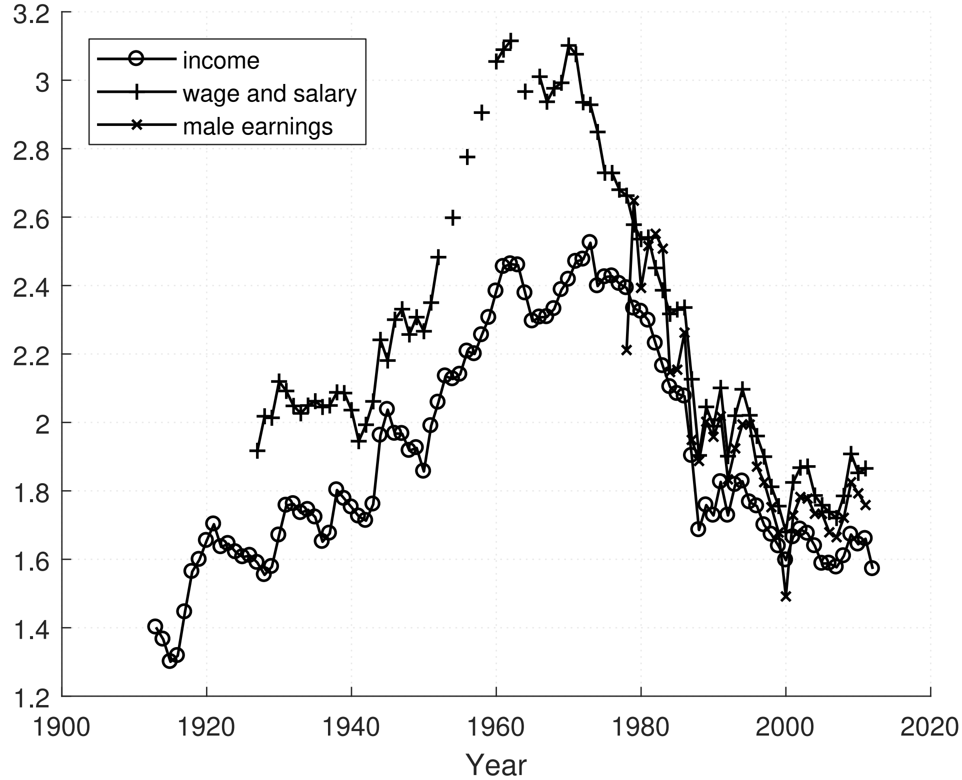
B.2 Pareto Statistic in US Data

Figure 10 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)

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

[32] 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 10: **Pareto Statistic at 99th Percentile: US Measures**

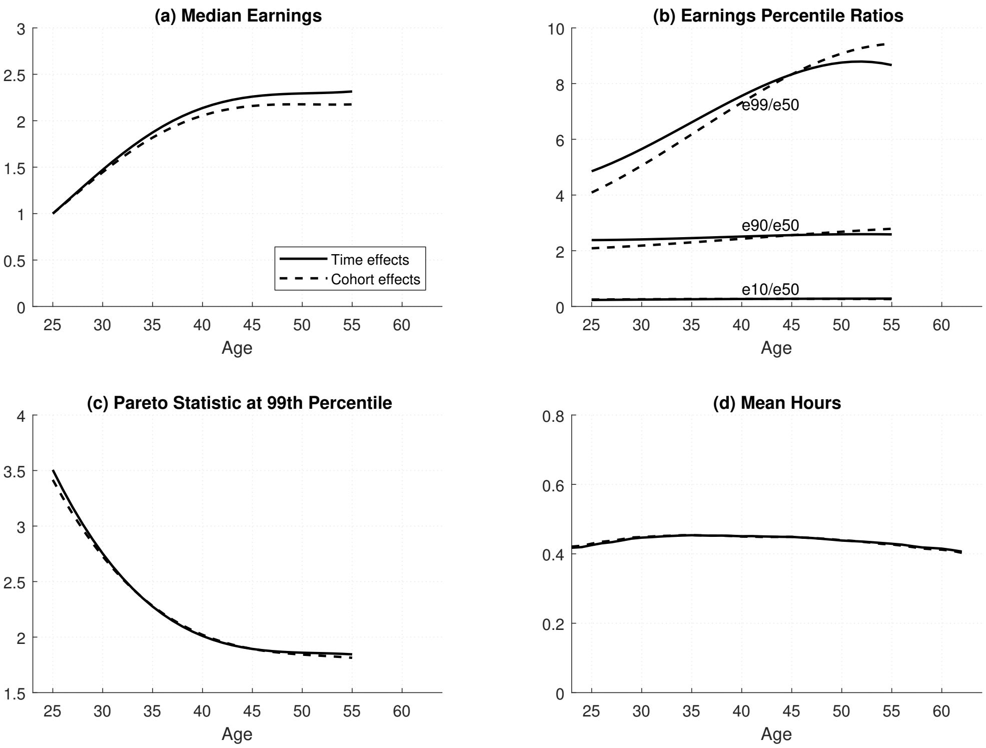


Note: Income measure is without capital gains and comes from the World Top Incomes Data Base. The wage and salary measure is for tax units using data from Piketty and Saez (2003, Update). The earnings measure is for males based on data from Guvenen, Ozkan and Song (2014).

is based on SSA data tabulated by Guvenen et al. (2014) and is constructed by assuming that earnings beyond the 99-th percentile follows a Pareto distribution and by using the reported 99-th and 99.999-th percentiles of earnings to infer the Pareto coefficient beyond the 99-th percentile.

Figure 10 shows that 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 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 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.

Figure 11: **Life-Cycle Profiles: Time and Cohort Effects**



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

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

Figure 11 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.

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 $\beta_1 > 0$.

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 $(0.0, 0.2, 0.4, 0.6, 0.8, 0.9, 0, 0.99, 0.995, 0.999, 1.0)$.

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) $\quad v_j(x, a) = \max_{(c,l,s,k')} u(c, l + s) + \beta E[v_{j+1}(k', h', a)]$ 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 $N = 2000$ 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 5, 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) $\quad v_j(k; \bar{k}, \bar{h}, a) = \max_{(c,l,k')} u(c, l + \bar{s}) + \beta E[v_{j+1}(k'; \bar{k}', \bar{h}', a)]$ 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.

$$\Delta \log hours_j = \alpha_0 + \alpha_1 \Delta \log wage_j + \epsilon_j$$

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 $N = 2000$ 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 $N = 2000$, 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 α_1

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

Note: 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.

$$u_{2,j}(c_j, l_j + s_j) + \lambda_j \left[wh_j(1 - \tau_j')\right] = 0 \; implies \; l_j + s_j = \left[\frac{\lambda_j \left(wh_j(1 - \tau_j')\right)}{\phi}\right]^{\nu}$$

$$\Delta \log(l_j + s_j) = \nu \left[\Delta \log \lambda_j\right] + \nu \Delta \log wh_j(1 - \tau_j')$$

$$\Delta \log(l_j + s_j) = \nu \left[-\log \beta(1 + r(1 - \tau_k))\right] + \nu \Delta \log wh_j(1 - \tau_j')$$

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 (2011) or 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.[33]

Table B1 shows a number of regularities. First, the regression coefficient for the 25-55 age group in the first row is positive but well below the value of ν from Table 4. 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 examined in Table B1 increase but still remain below the value of ν. Domeij and Floden (2006, Table 5) estimate in 1984-94 PSID data that $\alpha_1 = 0.16$ 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 total wealth.

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

Table B2 estimates all three elasticities, based on system (3) below, using the US data displayed in Figure 3 from section 3.

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

Table B2 - System Regression Based on 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 2. Appendix B describes data construction. Estimation follows methods in Table 2.

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

What is new in Table B2 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 B2 are the level and share regressions in Saez (2004) stated below. 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.[34] 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 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 B2 are larger than those in Saez (2004).[35] 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.

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

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

[34] 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).

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

Table B3 - 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)

We apply the regression methods used in Table B2 to model data in Table B3. 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 B3 shows that the true model elasticities $(\epsilon_1, \epsilon_2, \epsilon_3) = (.317, .040, .739)$ are all underestimated. One reason for this is that with a short time series the trend term captures part of the fall in income and in tax revenue due to the tax rate increase.

References Appendix B

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, 104, 122-26.

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

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

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

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

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

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

Saez, E. (2004), Reported Incomes and Marginal Tax Rates, 1960-2000: Evidence and Policy Implications, Tax Policy and the Economy, 18, 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, 50:1, 3-50.