Research Division
Federal Reserve Bank of St. Louis
Working Paper Series

Taxing Top Earners: A Human Capital Perspective

Alejandro Badel
and
Mark Huggett

Working Paper 2014-017B
http://research.stlouisfed.org/wp/2014/2014-017.pdf

July 2014
Revised October 2014

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

Taxing Top Earners: A Human Capital Perspective*

Alejandro Badel
Federal Reserve Bank of St. Louis
ale.badel@gmail.com

Mark Huggett
Georgetown University
mh5@georgetown.edu

THIS DRAFT: OCTOBER 29, 2014

Abstract
We assess the consequences of substantially increasing the marginal tax rate on U.S. top earners using a human capital model. We find that (1) the peak of the model Laffer curve occurs at a 52 percent top tax rate, (2) if human capital were exogenous, then the top of the Laffer curve would occur at a 66 percent top tax rate and (3) applying the theory and methods that Diamond and Saez (2011) use to provide quantitative guidance for setting the top tax rate to model data produces a tax rate that substantially exceeds 52 percent.

Keywords: Human Capital, Marginal Tax Rates, Inequality, 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.

1 Introduction

In a paper entitled *The Case for a Progressive Tax: From Basic Research to Policy Recommendations,* Diamond and Saez (2011) distill a few recommendations for tax reform from the vast literature on optimal income taxation. Their Recommendation 1 states *"Very high earnings should be subject to rising marginal rates and higher rates than current U.S. policy for top earners."* They argue that the marginal earnings tax rate on top earners should be 73 percent, using what they view as a mid-range estimate for a key elasticity. Their quantitative guidance comes from choosing the top tax rate to maximize the tax revenue obtained from top earners. They argue that the top rate that maximizes revenue will approximate the top rate that maximizes welfare for some welfare measures including the utilitarian measure.

The goal of this paper is to assess the consequences of increasing the marginal tax rate on top earners beyond the U.S. level in 2010. According to Diamond and Saez (2011) the marginal tax rate on top earners was 42.5 percent in 2010. This rate applies to roughly the top 1 percent of U.S. households.[1]

From a human capital perspective, one might question whether substantially increasing the marginal tax rate on top earners is misguided. First, while there is a simple formula for the revenue maximizing top tax rate in some static models, the formula is invalid in a dynamic human capital model. The quantitative guidance offered by Diamond and Saez (2011) is based on static theory and a short-run elasticity estimate. Second, the short-run response of labor input may be smaller than the long-run response as skills are largely fixed in the short run. Skill accumulation may be discouraged by the prospect of a substantially higher marginal tax rate applying later in life.

We use a human capital model to assess the consequences of increasing the marginal tax rate on top earners and returning any additional revenue in equal lump-sum transfers. Agents in the human capital model differ in terms of two initial conditions: initial human capital and learning ability. Initial human capital (i.e. skill) affects the intercept of an agent's mean earnings profile. Learning ability acts to rotate this age-earnings profile as good learners spend more time learning early in the working lifetime but reap the benefits later in life.

[1] The top federal tax rate was 35 percent in 2010. Diamond and Saez (2011) calculate that the top marginal rate is 42.5 percent based on federal and state income taxes, medicare taxes and sales taxes. The top federal rate starts at a taxable income level of $\$373,650$ for joint filers. This corresponds to a total income level of $\$392,350$ for joint filers with no dependents according to TAXSIM. The 99th percentile of the U.S. income distribution (including capital gains) in 2010 was $\$365,026$ in 2012 dollars according to the World Top Incomes Database.

The model has two forces that can produce an increase in earnings dispersion with age like that observed in U.S. data. First, agents differ in learning ability. Good learners have a mean earnings profile with a steeper slope other things equal. Second, agents differ ex-post due to differing realizations of human capital shocks over the lifetime. Shocks have a persistent effect on earnings as they have a persistent impact on skills. Thus, both luck and initial conditions are potentially important.

Our empirical strategy identifies the role of luck and initial conditions. Wage rates move over time for an agent late in life in the model only because of shocks. This implies that late in the working lifetime log wage rate differences identify the idiosyncratic shock variance following the line of argument in Huggett, Ventura and Yaron (2011). We use their estimate for the shock variance. Given this estimate, we then choose the distribution of learning ability and initial human capital, along with other model parameters, so a steady state of the model best matches features of the U.S. age-earnings distribution.

The model economies have a Laffer curve that relates the top tax rate to the resulting steady-state, lump-sum transfer. In the benchmark tax reform the top tax rate that applies to labor income is varied but the capital income tax rate is unchanged. We also analyze an alternative tax reform that increases the top tax rate and applies the progressive income tax formula to the sum of capital plus labor income. This tax reform can be viewed as increasing tax progression and eliminating the preferential tax treatment on types of capital income (e.g. capital gains and dividends). The peak of the Laffer curve occurs at a tax rate of 52 percent under the benchmark reform and 49 percent for the alternative tax reform.

We analyze the importance of human capital for the shape of the Laffer curve. We find that increasing the top tax rate leads to a fall in steady-state aggregate labor input. More than half of the fall in the aggregate labor input (i.e. aggregate skill-weighted labor hours) is due to the change in skill and the remaining part is due to the fall in work hours. Both components are driven by the behavior of agents with very high learning ability levels. We also find that the model Laffer curve is flatter with a smaller revenue maximizing top rate compared to the Laffer curve that would hold in an otherwise similar model that ignores the possibility of skill change in response to a tax reform. The alternative model could be viewed as an exogenous human capital model. The revenue maximizing top tax rate is more than 10 percentage points larger for either reform in the exogenous human capital model.

We use the ex-ante expected utility of young agents as a welfare measure. The steady-state welfare gain to young agents follows the same qualitative pattern as the

Laffer curve for transfers under the benchmark reform. However, the type-specific welfare results differ dramatically. Young agents with learning ability near or below the median level experience welfare gains that resemble the qualitative shape of the Laffer curve for transfers. In contrast, young agents with high learning ability experience welfare losses that increase as the magnitude of the top tax rate increases. Welfare results differ strongly by type for two main reasons. First, the majority of the variation in a number of measures of lifetime inequality is due to differences in initial conditions at age 23 rather than shocks over the remainder of the working lifetime. This is consistent with results from Huggett, Ventura and Yaron (2011). Second, agents with high learning ability tend to have high initial human capital at age 23 and face the highest probability of later becoming top earners.

There is a simple formula for the revenue maximizing top tax rate in static models (e.g. the model in Mirrlees (1971)) that depends on only two inputs: an earnings elasticity for top earners and a statistic of the upper tail of the earnings distribution. We argue later in the paper that the formula is invalid in dynamic models. Nevertheless, we use data from the human capital model and a variety of standard procedures from the literature that estimates elasticities in response to a tax reform to compute these two inputs. We find that the revenue maximizing top tax rate implied by the formula using these standard procedures is between 67 and 98 percent when the true top of the model Laffer curve is 52 percent. Thus, the formula, used in conjunction with standard empirical methods, does not accurately predict the top of the model Laffer curve. These are the tools that Diamond and Saez use to provide quantitative guidance for setting the marginal tax rate on top earners.

The paper is organized as follows. Section 2 presents the model framework. Section 3 documents properties of the U.S. age-earnings distribution. Sections 4 sets model parameters and section 5 describes model properties. Section 6 assesses the consequences of increasing the marginal tax rate on top earners. Section 7 discusses the main results of the paper.

2 Framework

The model we employ is closest to the human capital model developed by Huggett, Ventura and Yaron (2011). A key difference is that we add leisure.[2]

[2] The framework that we use is also related to the framework used in four other recent papers. Erosa and Koreshkova (2007) use a dynastic model with human capital accumulation to assess the

Decision Problem In Problem P1 an agent maximizes expected utility which is determined by consumption $c = (c_1, ..., c_J)$, work time decisions $l = (l_1, .., l_J)$ and learning time decisions $s = (s_1, ..., s_J)$. Consumption c_j, work time l_j and learning time s_j decisions at age j are functions of initial conditions $x = (h_1, a)$ 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. Idiosyncratic shocks z_{j+1} impact an agent's skill level. These shocks are independent and identically distributed over time.

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

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

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 the agent divides up his one unit of available time into distinct uses: work time l_j and learning time s_j. Leisure time is implicitly the difference between the one unit of available time and total labor time $l_j + s_j$. Earnings e_j equal the product of a rental rate w, skill h_j and work time l_j before a retirement age, denoted $Retire$, and is zero afterwards. Learning time s_j and learning ability a augment future skill through the law of motion for future human capital $h_{j+1} = H(h_j, s_j, z_{j+1}, a)$.

Equilibrium The model economy has an overlapping generations structure. The fraction μ_j of age j agents in the economy at a point in time obeys the recursion $\mu_{j+1} = \mu_j/(1+n)$, where n is the population growth rate. There is an aggregate production function $F(K, L)$ with constant returns which converts aggregate quantities of capital K and labor L into output. 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

impact of replacing a progessive tax system with a proportional tax system. Guvenen, Kuruscu and Ozkan (2014) use a human capital model to address whether differences in tax progression across countries is a key source of cross-country differences in earnings and income distributions. Finally, Guner, Lopez-Daneri and Ventura (2014) and Kindermann and Krueger (2014) also analyze tax reforms that are directed at the upper tail of the income distribution. A key difference is that they do not employ a human capital framework and thus do not allow labor productivity or skill to respond to a tax reform.

of agents, population fractions $(\mu_1, \mu_2, ..., \mu_J)$ and the distribution ψ of initial conditions. For example, the capital stock is the weighted sum of the mean capital holding within each age group.

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

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, given (w, r).*
2. *Prices: $w = F_2(K, L)$ and $r = F_1(K, L) - \delta$*
3. *Government Budget: $G = T$*
4. *Feasibility: $C + K(n + \delta) + G = F(K, L)$*

The model economies employ the functional forms stated below. The utility function u and the aggregate production function F are widely employed within a large applied literature. The parameter χ in the utility function affects the growth rate of total labor time $(l_j + s_j)$ over the working lifetime. This role for χ is discussed in section 5. The utility function parameter ϕ affects the mean of the total labor time. The human capital production function H takes the functional form used in Ben-Porath (1967). The idiosyncratic shocks z to an agent's stock of human capital are independent and identically distributed across periods and are normally distributed. Initial conditions $x = (h_1, a)$ follow a bivariate distribution with the property that the marginal distributions for learning ability and initial human capital are both right-tailed Pareto-Log-Normal (PLN) distributions - see Appendix A.4. This bivariate distribution is characterized by 6 parameters.[3] We delay the discussion of the functional form for the tax function T_j until section 4.

Benchmark Model Functional Forms:

Utility: $u_j(c, l + s) = \frac{c^{(1-\rho)}}{1-\rho} - \phi \exp(\chi(j-1)) \frac{(l+s)^{(1+\frac{1}{\nu})}}{1+\frac{1}{\nu}}$

[3] We use the Pareto-Log-Normal distribution as early work with bivariate lognormal distributions led to difficulties matching upper tail properties of the US age-earnings distribution.

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

The model structure implies that the idiosyncratic shock variance σ_z^2 can in principle be pinned down by very specific aspects of the data. Huggett, Ventura and Yaron (2011) argue that observations on log wage rate differences late in the working lifetime pin down σ_z^2 independently of other model parameters. Intuitively, this occurs because the change in the log wage rate for an agent across periods is entirely determined by shocks when an agent's skill investments go to zero. Moreover, skill investments decline late in life within the model as the number of periods where an agent can recoup the gains to these investments falls. We employ the estimate of the standard deviation σ_z from Huggett, Ventura and Yaron (2011).

One might also conjecture that the utility function parameter ν can be pinned down by very specific moments of the data. This parameter has been a focus of the literature that estimates the Frisch elasticity of labor hours using the structure implied by models with exogenous wage rates. However, the mapping from regression coefficients in existing empirical work to values of the model parameter ν within our human capital model is not as straightforward as many economists might conjecture. Thus, we do not set this parameter using off-the-shelf estimates based on exogenous-wage models. Instead, we set ν along with other model parameters so that a regression of the change in log work hours on the change in log wage rates, based on earnings and hours data produced by our human capital model, approximates the regression coefficient found in the work of MaCurdy (1981).

3 Empirics

This section characterizes how the distributions of earnings and work hours for male workers move with age. Our data come from the Social Security Administration (SSA) and the Panel Study of Income Dynamics (PSID). We use tabulated SSA male earnings data from Guvenen, Ozkan and Song (2013) and PSID male earnings and hours data from Heathcote, Perri and Violante (2010). These data sets are described in Appendix A.1. SSA data are notable because the sample size is very large and earnings are not top coded. Earnings in both datasets are deflated to be in real units.

We characterize age profiles for a number of earnings and hours statistics. When we use SSA data, we calculate an earnings statistic from the data for males age j in year t and then run an ordinary-least-squares regression of this statistic on a third-order polynomial in age plus a dummy variable for each year. We plot the age effects from the estimated age polynomial after vertically shifting the polynomial to run through the mean across years of the data statistic at age 45.[4] The earnings statistics of interest for each age and year are (i) median earnings, (ii) the 10-50, 90-50 and 99-50 earnings percentile ratios and (iii) the Pareto statistic at the 99th percentile of earnings.

The Pareto statistic at the 99th percentile for age group j in year t is the mean earnings $\bar{e}^{99}_{j,t}$ for observations above this percentile divided by $\bar{e}^{99}_{j,t}$ less the 99th percentile $e^{99}_{j,t}$. The Pareto statistic is an inverse measure of the thickness of the upper tail of the earnings distribution as the statistic takes on a lower value when the upper tail is thicker. We analyze the Pareto statistic because it plays a key role in the revenue maximizing tax rate formula used by Diamond and Saez (2011). The basic idea is that by setting model parameters to best match the Pareto statistic by age then the purely mechanical effect on tax revenue of increasing the tax rate on top earners is approximated.

$$Pareto_{j,t} = \frac{\bar{e}^{99}_{j,t}}{\bar{e}^{99}_{j,t} - e^{99}_{j,t}}$$

The methodology for characterizing age effects for hours statistics in PSID data is slightly different. We run a regression of the hours statistic on age and time dummy variables and use the age dummy variables to highlight age effects. The hours statistic for each age group j and time period t is mean hours stated as a fraction of total discretionary time.[5] The construction of age-year cells is described in Appendix A.1.

Figure 1 highlights how the earnings and hours statistics move with age. Median earnings are hump-shaped with a peak at around age 50. The 90-50 and the 99-50 earnings percentile ratio both increase over most of the working lifetime. The increase in the 99-50 earnings percentile ratio is particularly strong as it approximately doubles from a ratio of near 4 at age 25 to a ratio of roughly 8 at age 55. Thus, earnings dispersion increases with age in the upper half of the distribution.

[4] This procedure is applied to all earnings statistics with the exception of median earnings, which is normalized to equal 100 at age 55.

[5] We assume that total discretionary time is 14 hours per day times 365 days per year.

The Pareto statistic decreases with age and is below 2.0 after age 45. Figure 1 also shows that the mean work hours profile for males in PSID data is hump-shaped but fairly flat with age.

We have examined the sensitivity of the profiles in Figure 1 in two directions. First, we analyze profiles based on SSA data when we control for cohort effects rather than time effects. The main change is that the magnitude of the increase in earnings dispersion with age in the top half of the distribution is slightly greater than for the time effects case. Second, we analyze earnings using PSID data rather than SSA data. The methodology differs when using PSID data as we run a regression allowing both age and time dummy variables or allowing age and cohort dummy variables. The age effects based on PSID data display the same qualitative behavior as the age effects based on SSA.[6] However, we find that measures of earnings dispersion found in SSA data display greater dispersion at a given age than the measures documented in widely-used PSID data.

4 Model Parameters

We set parameter values following three main considerations. First, we set some parameters to fixed values without computing equilibria to the model economy. Parameters governing demographics, technology and the tax system are set in this way as is the coefficient of relative risk aversion. Second, the parameter governing the standard deviation of human capital shocks is set to an estimate from Huggett, Ventura and Yaron (2011). Third, the remaining model parameters are set so that equilibrium properties of the model best match empirical targets, including those displayed in Figure 1, by minimizing a sum of squared deviations criterion. The Appendix describes how we compute an equilibrium.

Demographics We use a model period of one year. An agent enters the model at a real-life age of 23, retires at age $Retire = 63$ and dies after age 85. These ages correspond to model ages 1 to 63. The population growth rate $n = 0.012$ is set to the geometric average growth rate of the U.S. population over the period 1940-2012. Population fractions μ_j sum to 1 and decline with age by the factor $(1 + n)$.

[6] PSID data have few observations beyond the 99th percentile for age groups 50 and older. Thus, we do not calculate the Pareto statistic for such age groups. The results for the Pareto statistic in PSID data rely on imputed values for top coded observations. See Heathcote, Perri and Violante (2010) for a description of the imputation procedure.

Table 1 - Parameter Values

Category	Functional Forms	Parameter Values
Demographics	$\mu_{j+1} = \mu_j/(1+n)$	$Retire = 41, n = 0.012$ $j = 1, ..., 63$ (ages 23-85)
Technology	$Y = F(K,L) = AK^{\gamma}L^{1-\gamma}$ and δ	$(A,\gamma) = (0.919, 0.322)$ $\delta = 0.0673$
Tax System	$T_j = T_j^{ss} + T_j^{inc}$	statutory rates - see text
Preferences	$u_j(c, l+s) =$ $\frac{c^{1-\rho}}{1-\rho} - \phi \exp(\chi(j-1)) \frac{(l+s)^{1+\frac{1}{\nu}}}{1+\frac{1}{\nu}}$	$\phi = 16.3, \beta = 0.975, \chi = -0.00514$ $\rho = 1.0$ (log utility), $\nu = 0.551$
Human Capital	$H(h,s,z,a) = \exp(z)\left[h + a(hs)^{\alpha}\right]$ and $z \sim N(\mu_z, \sigma_z^2)$	$\alpha = 0.632$ $(\mu_z, \sigma_z) = (-0.0133, 0.111)$ σ_z follows HVY (2011)
Initial Conditions	$a \sim PLN(\mu_a, \sigma_a^2, \lambda_a)$ and $\epsilon \sim LN(0, \sigma_\epsilon^2)$ $\log h_1 = \beta_0 + \beta_1 \log a + \log \epsilon$	$(\mu_a, \sigma_a^2, \lambda_a) = (-0.442, 0.00149, 3.99)$ $(\beta_0, \beta_1, \sigma_\epsilon^2) = (5.44, 1.18, 0.253)$

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

Technology We target empirical values for capital's share of output, the capital-output ratio K/Y, the real return to capital r together with the normalization $w = 1$. We set $\gamma = 0.322$ to produce the capital share. Then, given γ, we set (A, δ) so that $(r, w) = (0.42, 1.0)$ when $K/Y = 2.947$. Finally, when we set the remaining model parameters, we impose the restriction $K/Y = 2.947$. The empirical sources for these values are described in Huggett, Ventura and Yaron (2011).

Tax System Taxes in the model are the sum of a social security and an income tax: $T_j(e_j, k_j r) = T_j^{ss}(e_j) + T_j^{inc}(e_j, k_j r)$. The model social security tax function is $T_j^{ss}(e_j) = \tau^{ss} \min\left[e_j, e_{max}\right]$ for $j < Retire$ and $T_j^{ss}(e_j) = -b\bar{e}$ otherwise. Earnings are taxed at a rate τ^{ss} for earnings up to a maximum taxable earnings level e_{max}. After a retirement age, agents receive a common benefit set to b times the mean earnings $\bar{e}$ in the model. We set $\tau^{ss} = 0.106$, $e_{max} = 2.56\bar{e}$ and $b = 0.4$.[7]

The model income tax is based on statutory federal tax rates and a combination of other tax rates. Figure 2 plots marginal federal tax rates in 2010 for different tax brackets as a function of total income. We state total income in Figure 2 in multiples of the 99th percentile of the income distribution in 2010. The top federal tax rate of

[7]We set e_{max} to equal the ratio of the maximum taxable earnings level \$106,800 in 2010 to average earnings \$41,673 in 2010 from the Social Security Administration's Annual Statistical Supplement (2012, Table 2.A.8). The model tax rate $\tau = 0.106$ is the old-age and survivor's insurance tax rate in the U.S. social security system. We set $b = 0.4$ so that the benefit in the model is 40 percent of mean earnings. The benefit implied by the U.S. old-age benefit formula is approximately 40 percent of mean earnings for an individual who earns mean earnings in each year of the working lifetime - see Huggett and Parra (2010, Figure 1).

35 percent in 2010 starts at a total income level somewhat above the 99th percentile. Figure 2 also plots a combined marginal tax rate that equals the federal rate plus a constant. The constant is set to 7.5 percent so that the combined top income tax rate in the model equals the 42.5 percent top rate calculated by Diamond and Saez (2011, p.168).[8] We pass a smooth curve through the data points describing the combined marginal tax rate to construct the model income tax function. The marginal rate is fixed at 42.5 percent for income levels in the top income bracket. Appendix A.2 discusses the construction of the tax function.

The model income tax T_j^{inc} is the sum of two components. The first component approximates the combined marginal tax rates as displayed in Figure 2. This component applies to income from earnings and social security transfers. The second component taxes capital income $k_j r$ at a proportional capital income tax rate $\tau^{cap} = 0.209$ that equals the federal tax rate of 15 percent on dividends and capital gains in 2010 plus the average state top tax rate of 5.9 percent reported in Diamond and Saez (2011). Thus, the model income tax function features progressive taxation of earnings and a flat tax rate on capital income. The lower federal tax rate on some forms of capital income (e.g. dividends and capital gains) is one reason why average federal income tax rates for extremely high income groups in U.S. data are well below the top federal tax rate.[9] Diamond and Saez (2011, footnote 3) claim that the lower tax rate on capital gains is key for accounting for this fact. We view the flat tax on capital income within the model as a useful way to approximate the taxation of capital income for high income households.

$$T_j^{inc}(e_j, k_j r) = T\left(e_j + b\bar{e} \times 1_{j \geq Retire}\right) + \tau^{cap} k_j r$$

Preferences We set the coefficient of relative risk aversion to $\rho = 1$ which is the log utility case. 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 ... ."* This parameter controls the strength of the income effect of a tax reform. All remaining model parameters,

[8] Their calculation accounts for federal and state income taxes, un capped medicare taxes, average sales taxes and rules on the deductibility of various taxes.

[9] Guner, Kaygusuz and Ventura (2013) document this fact using Internal Revenue Service data.

including the remaining parameters governing the utility function, are set to best match empirical targets.

Remaining Model Parameters We set all remaining model parameters so that equilibrium properties of the model best match the earnings and hours properties documented in Figure 1, the average cross-sectional Pareto coefficient for earnings at the 99th percentile for earnings over the period 1978-2011 and a regression coefficient from MaCurdy (1981). The remaining parameters are those governing (i) initial conditions $(\mu_a, \sigma_a^2, \lambda_a)$ and $(\beta_0, \beta_1, \sigma_\epsilon^2)$, (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, χ).

The last target mentioned above is based on evidence from an empirical literature that regresses the change in log labor hours on the change in a log wage measure and a constant term. The literature on the Frisch elasticity of labor supply employs this approach. The regression equation used in the literature is stated below. The target value for α_1 is 0.125 based on MaCurdy (1981, Table 1 row 5-6) who uses earnings and hours data for white males age 25-55.

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

To connect to evidence on this regression coefficient, we produce data on earnings and hours from the model and calculate model wages as earnings divided by hours. Hours data within the model is taken to be total hours: the sum of work time and learning time. The sample within the model is based on agents age 25-55 following MaCurdy. We then estimate the coefficients in the linear regression.[10] Section 5 discusses the results of the estimation of the regression equation and provides theoretical perspective.

[10]Keane and Rogerson (2011) state: "Economists should be seeking to identify the underlying structural parameters of these choice problems and then use that information to infer elasticities, rather than trying to explicitly estimate something called a labor supply elasticity that is then applied across different situations." Our approach is consistent with such methodological advice as we choose model parameters to best match empirical targets. One useful empirical target is the regression coefficient that has often been estimated in the labor literature.

5 Properties of the Model Economy

5.1 Age-Earnings Distribution

Figure 3 highlights model and data properties for a number of earnings and hours statistics that were directly targeted in setting model parameters. The model economies produce a hump-shaped median and mean earnings profile by a standard human capital mechanism. Specifically, the mean human capital profile in the model is hump-shaped as agents concentrate learning time, and thus human capital production, early in the working lifetime. Towards the end of the working lifetime, both the median and the mean human capital profile fall. This occurs for two main reasons. First, time allocated to learning goes to zero as the number of future working periods over which the agent can recoup these investments fall. Second, the mean of the multiplicative shock to human capital is below one (i.e. $E[exp(z)] = exp(\mu_z + \frac{\sigma_z^2}{2}) < 1$) based on the parameter values in Table 1. Thus, on average skills depreciate. Skill depreciation is how the model squares the curvature of the median earnings profile late in life with the relatively flat shape of the mean hours profile late in the working lifetime.

Measures of earnings dispersion increase with age in U.S. data. The 99-50 earnings ratio doubles from age 25 to age 50 and the Pareto coefficient falls with age. The model economy has two forces leading to increasing earnings dispersion: differences in learning ability and human capital 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 uisng specific moments of log wage rate changes for older workers in panel data. Given this estimate, the parameters of learning ability and initial human capital are set to match the earnings and hours facts in Figure 3, including the increase in the 99-50 ratio and the fall in the Pareto statistic. These parameters are discussed in the next section.

5.2 Distribution of Initial Conditions

The distribution of initial conditions is determined by the functional form from section 2 and the parameter values in Table 1. For computational reasons we employ a discrete approximation to this distribution. The approximation has 9 different learning ability levels and for each learning ability level there are 20 human capital levels. We put the majority of the learning ability levels in the upper tail of the ability distribution as the focus of the paper is on the behavior of top earners. Our approximation methods are described in Appendix A.4.

Simple summary measures of this joint distribution are given in Table 2. There is substantial heterogeneity in initial human capital as the coefficient of variation is 0.73. Human capital follows a right-skewed distribution with a mean-median ratio of 1.24. Learning ability also displays substantial variation with a coefficient of variation of 0.34. The dispersion in learning ability reflects the fact that the model requires a source for increasing earnings dispersion with age beyond that due to idiosyncratic risk in order to produce the properties in Figure 3. Log learning ability

Table 2 - Distribution of Initial Conditions

$SD(h_1)/Mean(h_1)$	$Mean(h_1)/Median(h_1)$	$SD(a)/Mean(a)$	$Corr(h_1, a)$
0.73	1.24	0.34	0.60

and log human capital are positively correlated at age 23 because the parameter β_1 in Table 1 is positive. The correlation in levels, rather than log units, is 0.60. The positive correlation is consistent with Huggett, Ventura and Yaron (2006, 2011) who analyzed human capital models with some of the features in the current paper. They argued that a zero correlation would tend to produce a U-shaped earnings dispersion pattern with age not found in US data and that a positive correlation serves to eliminate such counter factual implications.

The positive correlation between human capital and learning ability has two important implications: (i) top earners will disproportionally be high learning ability agents at prime earnings ages and (ii) high learning ability agents will tend to have high lifetime earnings. A consequence of (i)-(ii) is that agents with high lifetime earnings will have high earnings growth. Figure 4 provides empirical support for this pattern. Guvenen, Karahan, Ozkan and Song (2014) document a strong positive relationship between a measure of lifetime earnings and a measure of earnings growth for U.S. male earners. Figure 4 also shows that a similar pattern holds in the model even though model parameters are not set to target this fact.

The fact that high lifetime earners have on average a very high earnings growth has an important implication within our model. The implication is that many top earners are top earners late in life but not earlier in life. Section 6 will argue that this is a key condition for investment in skills to fall as the top tax rate increases.

5.3 Mean Earnings, Wage and Human Capital Profiles

Figure 5 highlights the mean profiles for earnings, wage rates, human capital and hours. Figure 5a shows that the wage rate grows more over the lifetime in percentage terms than human capital. Individual human capital is proportional to the ratio of earnings e_j to work time l_j while wage equals the ratio of earnings e_j to total hours $l_j + s_j$. Thus, the mean wage profile is steeper than the mean human capital profile because the profile of total hours in Figure 5b is flatter than the work time profile. Clearly, in setting model parameters, we assume that what is measured in PSID data between ages 23 to 62 is total hours which comprises model work time and model learning time. This interpretation of measured hours data is also adopted by Wallenius (2011) among others.

The shape of the mean total hours profile is governed by several parameters. The parameter ϕ helps to control the level of total hours over the lifetime. A number of parameters (the parameter χ, the discount factor β and the interest rate r) control the average growth rate of the total hours profile over the lifetime. This point is argued in the next section.

5.4 Regressing the Change in Hours on the Change in Wages

The labor literature has estimated the coefficients in the linear regression equation below. The literature constructs a wage measure dividing reported labor earnings by reported labor hours. See Keane (2011) and Keane and Rogerson (2011) for recent reviews. For example, MaCurdy (1981) uses PSID data for white males age 25-55 and finds a regression coefficient of 0.125.[11] Altonji (1986) reexamines MaCurdy's framework and concludes that regression coefficients between 0 and 0.35 can be obtained using PSID data for prime-age males. This type of evidence is behind the view that labor hours are not very elastically supplied by prime-age males. Moreover, MaCurdy and Altonji argue that within exogenous-wage models the regression coefficient α_1 is an estimate of the preference parameter ν under appropriate conditions. Both authors calculate instrumental variables (IV) estimates of α_1.

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

We set the parameters of the human capital model to minimize the distance between data statistics and model statistics. The data statistics are those characterized in Figure 1 together with the average Pareto statistic at the 99th earnings percentile across years based on SSA data and the regression coefficient $\alpha_1 = 0.125$. The model counterpart to the empirical regression coefficient is based on the 25-55 age group when $wage_j = e_j/(l_j + s_j)$ and when IV methods are employed.

[11] This is the average of the point estimates from MaCurdy (1981, Table 1 row 5-6).

The first row of Table 3 shows the regression coefficient α_1 produced by the human capital model when the wage measure is earnings e_j divided by total labor hours $l_j + s_j$. The results for the 25-55 age group can be compared to those in MaCurdy (1981). This is because the wage rate is calculated as in MaCurdy (i.e. wage equals earnings divided by the hours measure used on the left-hand-side of the regression), the 25-55 age group is the same and IV methods are employed as in MaCurdy. The model regression coefficient of $\alpha_1 = 0.114$ is close to the average value $\alpha_1 = 0.125$ estimated by MaCurdy.[12]

Table 3 - Model Regression Coefficient α_1

Wage Measure	Age 25-55	Age 36-62	Age 50-60
$wage_j = e_j/(l_j + s_j)$	0.114	0.130	0.128
$wage_j = e_j/l_j$	0.134	0.137	0.190
$wage_j = e_j(1-\tau_j')/l_j$	0.152	0.153	0.221

Note: Model hours on the left-hand side of each regression are calculated as $hours_j = l_j + s_j$. The symbol τ_j' denotes the marginal earnings tax rate. The results are based on the parameters in Table 1, where $\nu = 0.551$.

To interpret the results in Table 3 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. This necessary condition is then restated using the functional form assumption on the period utility function from section 2. The second equation takes first differences of the log of the necessary condition. The third equation uses the Euler equation for asset holding to replace the change in the Lagrange multiplier with model variables and parameters. In this last step we assume 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. It is well understood that an extra Lagrange multiplier term enters the last equation when the agent is at a corner (see Domeij and Floden (2005)). In addition, when there is risk, the last equation is modified by an additive

[12] We create a data set of pairs $(\Delta \log hours_j, \Delta \log wage_j)$ in two steps. Step 1: For each initial condition $x = (a, h) \in X_1^{grid}$, draw $N = 2000$ lifetime shock histories. Appendix A.4 describes the construction of X_1^{grid} and associated probabilities $\psi(x), \forall x \in X_1^{grid}$. Step 2: For each $x \in X_1^{grid}$, shock history and age j in the age range in Table 3, 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 $\frac{1}{N}\mu_j\psi(x)$ on an observation, where $N = 2000$, μ_j are age shares defined in Table 1 and $\psi(x)$ are probabilities of initial conditions.

"forecast error" term (see Keane (2011) or Keane and Rogerson (2012)) where the additive term is based on a linear approximation.

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

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

$$\Delta \log(l_j + s_j) = \nu \left[-\chi - \log \beta(1 + r(1-\tau^{cap}))\right] + \nu \Delta \log wh_j(1-\tau'_j)$$

The last equation above suggests that the human capital model is in a sense 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.

Table 3 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 $\nu = 0.551$ in the human capital model from Table 1. Second, the regression coefficient for any age group increases as the wage measure better approximates the wage concept relevant in the human capital model. One reason for this is that the growth in the baseline wage measure (earnings divided by total labor hours) exceeds the growth of human capital over the lifetime. Figure 5 from the previous section highlighted this point.[13] Another reason for this is that changing marginal tax rates are taken into account. Third, 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,

[13] This logic for why the regression coefficient is lower than the utility function parameter ν within human capital models is not new but it may be under appreciated. Imai and Keane (2004) and Keane and Rogerson (2012) make this point using a human capital model with learning by doing. Wallenius (2011) makes this point using a human capital model that is closer to the framework that we use.

the regression coefficient is still less than half the value of ν. Domeij and Floden (2005) 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 assets), then all regression coefficients in Table 3 increase markedly but still remain below the value of ν.

5.5 Earnings, Income and Wealth Distributions

Table 4 compares some key statistics of the distribution of earnings, income and wealth in the model economy with those from the U.S. economy. The model economies target earnings distribution facts. Thus, model parameters are set without directly targeting cross-sectional properties of income or wealth.

Table 4 - Distribution of Earnings, Income and Wealth

Variable	Economy	Top 1 %	Top 5 %	Top 20 %	Pareto Coefficient at the 99th percentile
Earnings	Model	11.6	26.5	52.2	2.00
	US 1978-2011	10.8	24.1	49.9	2.00
Income	Model	11.5	26.5	52.2	2.06
	US 1978-2011	14.3	29.0	-	1.86
Wealth	Model	17.3	40.1	74.9	2.15
	US 2007	33.6	60.3	83.4	1.81

Note: (1) US earnings distribution facts are averages over the years 1978-2011 based on our calculations using male earnings data tabulations from the SSA data set constructed by Guvenen et al. (2013). (2) US income distribution facts are averages over the years 1978-2011 based on data from World Top Incomes Database that include capital gains in the income measure. (3) US wealth distribution facts are from Diaz-Gimenez, Glover and Rios-Rull (2011) based on the 2007 Survey of Consumer Finances.

The model economy produces an earnings distribution that is roughly consistent with the upper tail properties of the U.S. cross-sectional distribution for male earnings when these values are averaged over the period 1978-2011. For example, the top 1 percent of earners in the model receive 11.6 percent of earnings whereas the average value in SSA data is 10.8 percent. The Pareto coefficient of earnings at the 99th percentile in cross section data is 2.0 in the model and averages 2.0 in SSA data.

The model economies produce an income distribution that does not concentrate as much income in the upper tail compared to the U.S. distribution. The U.S. income data summarized in Table 4 use the tax unit (see Alvaredo, Atkinson, Piketty and Saez, The World Top Incomes Database) as the unit of observation, measure income including capital gains and average each statistic over the period 1978-2011. Model and data are much closer when income excludes capital gains as the top 1 percent, top 5 percent and Pareto statistic in the U.S. average $12.7, 27.4$ and 2.03 over the 1978-2011 time period when income excludes capital gains. Regardless of whether capital gains are included or excluded from the empirical income measure, it seems sensible to compare to income facts averaged over the 1978-2011 period given that the earnings process in the model targets earnings statistics calculated from data over the same period. Over the period 1978-2011 the top shares of both the male earnings distribution and the income distribution trend upwards. The share of income received by the top 1 percent of tax units increased particularly strongly starting at 9.0 percent in 1978 and ending at 19.6 percent in 2011 while the Pareto coefficient for income decreased from 2.19 to 1.53 when income includes capital gains. The model economy produces upper tail income properties more consistent with the middle rather than the end of this time period.

The model economies concentrate too little wealth in the top 1 percent of the distribution compared to the U.S. wealth distribution. The model produces more than half of the fraction of wealth held by the top 1 percent of U.S. households based on 2007 Survey of Consumer Finances (SCF) data analyzed by Diaz-Gimenez, Glover and Rios-Rull (2011). This finding is not surprising as it is well known, at least since Huggett (1996), that life-cycle models that are calibrated to match features of the U.S. earnings distribution have difficulty matching the concentration of wealth held by the top 1 percent of the U.S. distribution. While income inequality has grown substantially over time, the concentration of wealth in the top of the U.S. wealth distribution has not changed dramatically in SCF data over the period 1983-2007 according to Wolff (2010).

6 Assessing the Tax Reform

6.1 Laffer Curve

We analyze Laffer curves for two reforms. Reform 1 alters the top tax rate on earnings but leaves the tax rate on capital income unchanged. Thus, the top tax rate of 42.5 percent, graphed previously in Figure 2, is changed without changing the tax rate schedule below the top tax bracket. Lump-sum transfers are positive if more revenue is collected in equilibrium under the new tax system. Government spending is held constant across all steady-state equilibrium comparisons. Reform

2 alters the top tax rate and imposes that capital income is no longer taxed at a flat rate. Labor income, capital income and social security transfers are now summed and this total income is subject to the progressive income tax function previously summarized graphically in Figure 2. This means that marginal earnings and marginal capital income tax rates arising from the income tax system are equal for an agent and that high income agents face a higher marginal income tax rate than lower income agents.[14]

Figure 6 displays Laffer curves. The horizontal axis measures the top tax rate and the vertical axis measures the equilibrium lump-sum transfer as a percentage of pre-reform GDP. The Laffer curve in the benchmark model peaks at a tax rate of roughly 52 percent for Reform 1 and 49 percent for Reform 2. More revenue is raised under Reform 2 - the reform that increases both the top earnings tax rate and the capital income tax rate. The lump-sum transfer for both reforms is below 0.10 percent of the initial steady-state output. Thus, both Laffer curves in the benchmark model are somewhat flat in that they do not raise much additional revenue. In addition, the top of both Laffer curves occur at a tax rate that is well below the 73 percent top rate that Diamond and Saez (2011) highlight as revenue maximizing.

Figure 6 also displays Laffer curves in models where the utility function parameter ν is moved away from the benchmark value of $\nu = 0.551$ in Table 1. We consider two alternative values $\nu = 0.35$ and $\nu = 0.75$. For each of these values, we repeat the estimation procedure from section 4, choosing the remaining parameters to best match targets.[15] We find that the revenue maximizing top tax rate decreases as the parameter ν increases and that the tax revenue at the top of the Laffer curve decreases as the parameter ν increases.

Figure 7 plots a measure of welfare gains associated with the tax reforms in the benchmark model. We calculate the ex-ante expected utility of a newborn agent in the benchmark model as well as in a steady-state equilibrium corresponding to each value of the new top rate. This could be viewed as a calculation of ex-ante expected utility behind the veil of ignorance so that agents do not know their initial conditions. We then calculate the percentage increase in consumption at all ages and states that is equivalent in expected utility terms to the ex-ante expected utility obtained in

[14] We acknowledge that marginal earnings and marginal capital income taxes differ for agents who are below the maximum taxable social security earnings when one focuses on the entire model tax system.

[15] The targets are the same as those used in section 4 with the exception that we do not target the regression coefficient estimated by MaCurdy (1981).

steady state under the new tax system. We call this an equivalent consumption welfare gains measure. The results are presented in Figure 7 under the legend label ALL to indicate that agents with all levels of learning ability are included in the welfare calculation. This welfare gains measure has roughly the same qualitative shape as the Laffer curve for transfers for Reform 1. For Reform 2 we find that all the values of the top tax rate in Figure 7 lead to welfare losses relative to the benchmark steady state. The wage rate is lower in these steady states compared to the benchmark steady state because the capital-labor ratio falls as the top tax rate increases.

Figure 7 also documents the equivalent consumption welfare gains measure for newborn agents conditional on learning ability. The welfare measure falls for agents with the two highest learning ability levels (i.e. ability levels 8-9) as the top tax rate increases. These agent types have the highest probability of becoming top earners. This occurs because these agents have a high average level of initial human capital and because they have a steeply-sloped mean earnings profile.

For agents with learning ability at or below median ability, the equivalent consumption welfare measure has the same qualitative pattern as the Laffer curve for transfers in Reform 1. Intuitively, this follows because the probability that these agents will pass the threshold associated with the top rate is nearly zero. Thus, these agents are impacted mainly by the transfer and the change in factor prices associated with the new aggregate factor inputs. Moreover, factor prices change very little across the steady states that we analyze under Reform 1.

6.2 Understanding the Role of Human Capital Accumulation

What role does human capital accumulation play in governing the shape of the Laffer curve? We answer this question in two ways. First, we calculate that the steady-state aggregate labor input falls in response to the top tax rate increase. Then we decompose this fall into a skills component versus an hours of work component. Second, we contrast the Laffer curve in the human capital model with the Laffer curve that would obtain if human capital were exogenous.

Decomposing the Change in Labor Input

We find that an increase in the tax rate on top earners is associated with a fall in both aggregate output and aggregate labor input for both of the tax reforms that we analyze. Aggregate labor input equals average human-capital-weighted labor hours.

We decompose the fall in the aggregate labor input into two components. One component measures the fall in the aggregate input due to changes in work hours at fixed skill levels. The remaining component captures the fall in the aggregate input associated with changing skill levels. ΔL denotes the steady-state labor input after the tax reform less the labor input in the original steady state. Terms with a hat symbol denote human capital and work hours under the higher top tax rate:

$$\Delta L = \sum_{j=1}^{J} \mu_j \int_X E[\hat{h}_j \hat{l}_j - h_j \hat{l}_j | x] d\psi + \sum_{j=1}^{J} \mu_j \int_X E[h_j \hat{l}_j - h_j l_j | x] d\psi$$

We find that, typically, more than half of the fall in the aggregate input is attributed to changing skill levels with the remaining portion due to a fall in work hours. We also apply this decomposition to the labor input from distinct learning ability types. We find that the largest percentage fall in skill-weighted labor input comes exclusively from agents with very high learning ability levels.

We describe an economic mechanism behind the fall in skill. Consider a deterministic version of Problem P1 for the benchmark tax system.[16] The first equation below is a necessary condition for an interior solution to the learning time decision s_j. It states that the marginal utility loss at age j from an extra unit of time devoted to acquiring skill equals the marginal utility gain to the extra future consumption that results from the extra skill investment. The second equation reorganizes the first equation, where $\hat{r}$ is the after tax interest rate.

$$\begin{aligned} u_1(c_j, l_j + s_j) w h_j (1 - \tau_j') &= \sum_{k=j+1}^{Retire-1} \beta^{k-j} u_1(c_k, l_k + s_k) \frac{dh_k}{ds_j} w l_k (1 - \tau_k') \\ w h_j (1 - \tau_j') &= \sum_{k=j+1}^{Retire-1} \left(\frac{1}{1+\hat{r}} \right)^{k-j} \frac{dh_k}{ds_j} w l_k (1 - \tau_k') \end{aligned}$$

Consider an agent who faces a marginal earnings tax rate τ_j' below the top rate early in life but who will face the top marginal rate in some future period $k > j$ later in life. Now consider an increase in the top tax rate. The left-hand side of the second equation does not change but some of the marginal net-of-tax rate terms $(1 - \tau_k')$ on the right-hand side will decrease, other things equal. Thus, some adjustment is needed. One source of adjustment is that a fall in time devoted to skill investment s_j

[16] Allowing for idiosyncratic risk simply involves adding expectation terms into the necessary condition.

raises the future marginal product terms $\frac{dh_k}{ds_j}$. If future labor hours l_k also decrease in response to the tax change, consistent with the results for high learning ability agents, then this requires an even larger fall in skill investment at age j.[17]

Laffer Curve with Exogenous Human Capital

What role does skill change play in accounting for the shape of the Laffer curve? To answer this question, consider an alternative economy that has the same preferences, technology and initial conditions as the benchmark model. This alternative economy has exactly the same steady-state equilibrium properties as the benchmark model under the benchmark tax system. The evolution of human capital for each agent is exactly the same in both models for the benchmark tax system. The key difference is that when the tax system changes then human capital investments change in the benchmark model but they remain unchanged in the alternative economy. Thus, the alternative economy could be called an exogenous human capital model.

At a mechanical level, it is simple to describe the difference in skill accumulation. In the alternative model, the time investment decisions $s_j(x, z^j)$ as a function of initial condition $x = (h_1, a)$ and shock history z^j are fixed and do not vary as the tax system changes. These decisions are set to equal those in the benchmark model under the benchmark tax system. In the benchmark model, these decisions are allowed to adjust when the tax system changes. In the alternative model all decisions other than the time investment decision are allowed to be adjusted to maximize expected utility when the tax system changes. Appendix A.3 describes our methods for computing steady-state equilibria in the alternative model.

Figure 8 plots the Laffer curve in the benchmark model and in the exogenous human capital model. The top of the Laffer curve for the exogenous human capital model raises more than three times as much revenue compared to the benchmark human capital model. Thus, one role that endogenous skill change plays is that it flattens out the Laffer curve compared to models that ignore the possibility of skill change in response to changes in the tax system. Another important difference is that the top of the Laffer curve in the model with exogenous skills occurs at a much higher top tax rate than the benchmark model. This difference is quantitatively

[17] This intertemporal tax wedge mechanism is not operative for agents that are in all periods and shock histories above the threshold at which the top tax rate increases. Such agents experience an income effect from the tax reform but would not experience a change in the ratio of net-of-tax rates across model periods. Consistent with this reasoning, the agents with the highest learning ability level experience a smaller percentage fall in labor input as the top tax rate increases compared to agents with the second highest learning ability level.

large as the top tax rate increases from 52 percent to roughly 66 percent for Reform 1 and increases from 49 percent to roughly 62 percent for Reform 2. Intuitively, these differences occur because aggregate labor supply is more elastic with respect to a change in the top rate in the human capital model.

Figure 9 shows that the mean earnings profile for high learning ability agents rotates clockwise as the top tax rate increases. The percentage change in mean earnings is greatest at older ages as the fall in time investments in skill production over the lifetime have a cumulative effect on skill that is greatest late in the working lifetime. Our analysis assumes that the tax reform leaves the initial distribution of skill and learning ability at age 23 unchanged.

6.3 Revisting the Revenue Maximization Formula

We revisit the revenue maximization formula that underlies the quantitative guidance of Diamond and Saez (2011) for setting the top tax rate. Consider the problem of choosing the top tax rate to maximize tax revenue within a static model. Implicitly, agents face a tax system with a top tax rate τ that applies beyond an earnings threshold $\underline{e}$. These agents make best choices and produce earnings e. The (per person) tax revenue extracted from agents based only on their earnings above the threshold is denoted $\tau E[e-\underline{e}|e \geq \underline{e}]$. The revenue maximization problem allows for agent heterogeneity as earnings are a function of the top rate and, implicitly, other agent characteristics.

Revenue Maximization Problem: $\max \tau E[e-\underline{e}|e \geq \underline{e}]$

Solution: $\tau^* = \frac{1}{1+a\epsilon}$, where $a \equiv \frac{E[e|e \geq \underline{e}]}{E[e-\underline{e}|e \geq \underline{e}]}$ and $\epsilon \equiv \frac{dE[e|e \geq \underline{e}]}{d(1-\tau)} \frac{(1-\tau)}{E[e|e \geq \underline{e}]}$.

A solution to this problem implies that at a maximum the top tax rate τ^* satisfies the formula above. Thus, the top tax rate is pinned down by the Pareto statistic (denoted by a) characterizing the top of the earnings distribution in the model and by an earnings elasticity ϵ for top earners with respect to the net-of-tax rate $(1-\tau)$. A basic assumption used in the derivation of the formula is that agents below the threshold $\underline{e}$ do not respond to changes in the top rate at the maximum. This is consistent, for example, with the way earnings are determined in the Mirrlees (1971) model. For this reason, neither the nature of the tax system nor the earnings distribution for earnings below $\underline{e}$ enter into the formula. Thus, the lack of a response from agents below the threshold accounts for the extreme simplicity of the formula.

The quantitative guidance offered by Diamond and Saez (2011) for setting the top earnings tax rate comes from plugging empirical values for (a, ϵ) into this formula.

The revenue maximizing top rate formula is not valid in dynamic models. One reason for this is that some agents who are currently below the top earnings threshold will respond to a change in the top rate. This would hold even if factor prices and transfers were to be held fixed. This occurs as some agents forecast that with positive probability they will cross the threshold later in life. In our model this leads agents with high learning ability to change their time allocation decisions before they become top earners. Thus, a tax rate change that is directly focused only on top earners ends up impacting the earnings distribution both above and below the threshold.

The main question that we want to address in this section is whether an economist who is equipped with the revenue maximizing tax rate formula $\tau^* = \frac{1}{1+a\epsilon}$ and proxies for (a, ϵ) based on data drawn from the human capital model would accurately predict the actual top of the model Laffer curve. To answer this question, we calculate the Pareto coefficient a in the benchmark model. We then consider a tax reform in the model that corresponds to setting the top tax rate equal to the rate defining the top of the model Laffer curve. Using model data from this reform, we then apply standard regression procedures like those used in the literature to calculate a proxy for ϵ. Finally, we plug the proxies for (a, ϵ) into the formula and report the results.

Saez, Slemrod and Giertz (2012) review the literature that estimates the income elasticity with respect to the net-of-tax rate. They also estimate this elasticity using U.S. data from 1991-97. They note that the great majority of the empirical studies on this issue use panel data. Their estimation is based on data from a tax reform that occurred in 1993. They calculate that this reform increased the average marginal tax rate of the top 1 percent but did not significantly change the average marginal tax rate for the rest of the top 10 percent. They examine a panel of the top 10 percent of US taxpayers over 1991-97 and run a battery of regressions, utilizing different controls, instruments and subperiods, to highlight a range of methods employed in the literature for estimating the regression parameter ϵ.

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

Saez et al. (2012, Table 2, Panel B) estimate the regression equation above, where ϵ is the empirical proxy for the elasticity parameter in the formula, z_{it} is income of individual i at time t, $\tau_t(z_{it})$ is the marginal tax rate, $f(z_{it}) = \beta \log z_{it}$ is an income

Table 5 - Elasticity with Respect to the Net-of-Tax Rate: Model Data

	(1)	(2)	(3)	(4)	(5)	(6)
Mean Elasticity ϵ	0.0089	0.1744	0.1840	0.2189	0.2129	0.2451
	(.0220)	(.0231)	(.0237)	(.0240)	(.0228)	(.0230)
Income Control $f(z)$	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: $1_{\{i \in T_2 \ and \ t=2\}}$			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
Tax rate implied by $\tau^* = \frac{1}{1+a\epsilon}$	0.98	0.74	0.73	0.70	0.70	0.67

Note: (1) We draw 100 balanced panel data sets of 30,000 agents. The agents in each balanced panel have labor income above the 90 percentile of earnings at $t = 1$. We follow these agents from periods $t = 1$ to $t = 7$. The tax reform in the model occurs at $t = 3$. (2) We report means and in parethesis standard deviations of the point estimates of ϵ across 100 randomly drawn balanced panels. (3) The tax rate implied by the formula uses the mean estimate of the elasticity and the model Pareto coefficient of $a = 1.97$ that holds beyond the earnings cutoff defined by the top bracket in Figure 2.

control and α_t are time dummy variables. They estimate ϵ using three different choices of instruments for log net-of-tax-rate changes and a two-stage-least-squares estimator. Instrument 1 is the indicator function taking the value 1 if individual i is in the top 1 percent in 1992 (i.e. $1_{\{i \in T_{1992}\}}$). Thus, T_{1992} denotes the set of individuals in the top 1 percent in 1992 which is the pre-reform year. Instrument 2 is $1_{\{i \in T_{1992} \ and \ t=1992\}}$. Instrument 3 is $\log\left(\frac{1-\tau_{t+1}(z_{it})}{1-\tau_t(z_{it})}\right)$ and represents the log of the ratio of net-of-tax rates across years if income were not to change across years.[18]

We outline this empirical framework to highlight a range of assumptions and techniques that have been used in the literature to calculate an empirical proxy for the parameter ϵ in the tax rate formula. We now use the same regression equation and apply the same techniques (i.e. two-stage-least squares) to a tax reform in the human capital model. The benchmark tax system applies in period $t = 1$ and $t = 2$ and agents view this system as being permanent. Agents are surprised to learn that the model tax system is modified permanently at $t = 3$. They learn this at the start of period $t = 3$. Thus, the model periods $t = 1, ..., 7$ loosely correspond to the years $1991 - 1997$ for the US economy. The tax system is modified by increasing the top tax rate to $\bar{\tau} = 0.52$, which corresponds to the tax rate at the top of the model Laffer curve. This is a partial equilibrium analysis of Reform 1 as factor prices and transfers are assumed to be fixed over time at initial steady-state values.

[18] Saez et al. (2012, Table 2 Panel B) apply these methods to their sample of the top 10 percent of taxpayers and produce point estimates for ϵ that lie in the interval $[-1.669, 2.42]$ when comparing only two years of data. Point estimates lie in the interval $[0.143, 1.395]$ when using all years of the data. Giertz (2008, Table 5 and 6) explores these results in more detail.

We construct 100 different panel data sets from the model. Each data set mimics the structure of the data set used by Saez et al. (2012, Table 2) and Giertz (2008). Their data set is a panel with roughly 30,000 tax units followed over the period 1991-1997.[19]

Table 5 presents estimation results based on this tax reform. Instruments 1-3 used in Table 5 are precisely the model analogs of instruments 1-3 used by Saez et al. (2012) as discussed above. Estimation results are based on the two-stage-least squares regression methods employed by Saez et al. (2012). Taking the results in Table 5 at face value and plugging the mean of the point estimate of the elasticity across the 100 data sets into the tax rate formula produces a top tax rate $\tau^* = 1/(1 + a\epsilon)$ that ranges from a low value of 67 percent to a high value of 98 percent. The true top of the model Laffer curve occurs at a 52 percent top tax rate for Reform 1 and a 49 percent top tax rate for Reform 2.

In summary, an economist who is equipped with the formula $\tau^* = 1/(1+a\epsilon)$ and proxies for (a, ϵ), based on standard techniques from the literature and data drawn from the model, would substantially overstate the true top of the model Laffer curve. This leads us to conclude that the theory and related empirical practices that Diamond and Saez (2011) draw upon to provide quantitative guidance for determining the revenue maximizing tax rate on top earners are not accurate predictors of the tax rate at the top of the Laffer curve when the data is being generated by the human capital model.

7 Discussion

This article assesses the consequences of increasing the marginal tax rate on top earners using a human capital model. We highlight three main findings. First, the top of the model Laffer curve occurs at top tax rates of roughly 49 and 52 percent for the two tax reforms that we analyze. These are well below the 73 percent top tax rate highlighted by Diamond and Saez (2011). Second, we determine the role that human capital accumulation plays in governing the shape of the model Laffer curve. The model Laffer curve is flatter with a substantially lower revenue

[19] To construct a model panel, we first draw $N = 2000$ lifetime shock histories for each initial condition $x = (a, h) \in X_1^{grid}$. We then calculate histories of all model decisions for time periods $t = 1 - 7$ for agents at all ages 55 and below at $t = 1$. The second step is to draw a random sample of 30,000 agents at $t = 1$ whose earnings lie above the 90th percentile of earnings and follow these agents for seven model periods. An agent of age j and initial condition x is drawn from the step 1 sample with probability proportional to $\mu_j \psi(x)$ and kept only if earnings at $t = 1$ is above the 90th percentile. See Appendix A.4 for the construction of $(X_1^{grid}, \psi(x))$.

maximizing top tax rate compared to the Laffer curve that would hold in a similar model that ignores the possibility of skill change in response to a tax reform. The main mechanism behind this is that the mean age-earnings profiles of agents with high learning ability rotate clockwise in response to the tax reform due to a change in skill investments. Third, we argue that a simple, elegant formula for the revenue maximizing top tax rate does not hold in dynamic models. Diamond and Saez (2011) use this formula to provide quantitative guidance for setting the top tax rate. We also show that in practice the formula substantially over predicts the top of the Laffer curve in the human capital model when one applies standard methods from the literature to calculate the relevant elasticity.

We now discuss three issues that provide perspective on these findings:

Issue 1: A common view is that labor hours are not very elastically supplied by prime-age males and that this might imply that the top of the Laffer curve occurs at a high top tax rate. Key evidence for the first part of this view comes from the literature on the Frisch labor hours elasticity that follows MaCurdy (1981) and Altonji (1986). We discipline our human capital model by choosing model parameters to target the regression coefficient estimated by MaCurdy (1981). While the model regression coefficient is close to the point estimate from MaCurdy (1981), the key utility function parameter ν exceeds the model-based regression coefficient.

Issue 2: A compelling analysis of the broad consequences of a tax reform directed at top earners needs a plausible mechanism for how one becomes a top earner. The mechanism that we analyze is based in part on period-by-period investments by agents with high learning ability. In our quantitative model, such agents have a steeply-sloped, mean-earnings profile and become top earners later in life. The patterns in U.S. data in Figure 4 show that individuals with very high lifetime earnings have average profiles that are steeply sloped. Alternative theories of top earners would also need to account for such a strong data feature. In the last few decades general equilibrium models with heterogenous agents have been developed that account for some aspects of earnings, income, labor hours and wealth distribution. Within this class, models where earnings are determined by the product of a common wage rate, an exogenous individual labor productivity component and a labor hours choice (see the literature reviewed by Heathcote, Storesletten and Violante (2009)) are widely used. However, we think that such models should be used with caution in assessing the consequences of increasing tax rates on top earners. First, in the data the vast majority of the variance in log earnings is due to the variance in log labor productivity rather than log work hours. Thus, the labor hours decisions of agents in such models can at best play a secondary role in accounting for who becomes a top earner. Second, these models ignore the posssibility that a tax reform can impact earnings via the labor productivity component. Neglecting this possibility implies a Laffer curve that collects much more revenue and has a much higher tax rate at the top compared to a similar model that allows skill accumulation to respond to a tax reform.

Issue 3: We argue that the revenue maximizing top tax rate formula $\tau^* = 1/(1+a\epsilon)$ is not valid in dynamic models. This is one of the two problems that we highlight with respect to the quantitative guidance provided by Diamond and Saez (2011). Of course, one could ask whether there is a valid top tax rate formula that can be applied to static models and to steady states of dynamic models. Such a formula would potentially be very useful as steady states of dynamic models are the main theoretical tool for interpreting inequality data and analyzing policy reforms.

Badel and Huggett (2014, Theorem 1) provide such a formula and show how it can be applied to static and dynamic models. It includes as a special case the formula $\tau^* = 1/(1 + a\epsilon)$ that is applicable to some static models. They apply the formula to the benchmark model analyzed in this paper and find that it accurately predicts the top of the model Laffer curve when the three elasticities in the formula are each approximated using a difference quotient. They find that the elasticity of labor income for top earners in the human capital model (i.e. one of the three theoretically relevant elasticities as defined by the formula) exceeds the elasticity estimates calculated in Table 5 of this paper. A key challenge for future work is to develop techniques to reliably estimate the three elasticities that enter the formula.

References

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

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

Badel, A. and M. Huggett (2014), The Sufficient Statistic Approach: Locating the Top of the Laffer Curve, manuscript.

Ben-Porath, Y. (1967), The Production of Human Capital and the Life Cycle of Earnings, Journal of Political Economy, 75, 352-65.

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

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

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.

Diaz-Gimenez, J., Glover, A. and V. Rios-Rull (2011), Facts on the Distributions of Earnings, Income and Wealth in the United States: 2007 Update, Federal Reserve Bank of Minneapolis Quarterly Review, 34, 2-31.

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.

Giertz, S. (2008), Panel Data Techniques and the Elasticity of Taxable Income, CBO Working Paper 2008-11.

Guner, N., Kaygusuz, R. and G. Ventura (2013), Income Taxation of U.S. Households: Facts and Parametric Estimates, manuscript.

Guner, N., Lopez-Daneri, M. and G. Ventura (2014), Heterogeneity and Government Revenues: Higher Taxes at the Top?, manuscript.

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 (2013), The Nature of Countercyclical Income Risk, manuscript.

Guvenen, F., Karahan, F., Ozkan, S. and J. Song (2014), 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.

Huggett, M. (1996), Wealth Distribution in Life-Cycle Economies, Journal of Monetary Economics, 38, 469-94.

Huggett, M. and J.C. Parra (2010), How Well Does the U.S. Social Insurance System Provide Social Insurance?, Journal of Political Economy, 118, 76-112.

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.

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

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

Keane, M. and R. Rogerson (2011), Reconciling Micro and Macro Labor Supply Elasticities: A Structural Perspective, NBER Working Paper 17430.

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.

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

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.

Tauchen, G. (1986), Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions, Economic Letters, 20, 177-81.

Wallenius, J. (2011), Human Capital Accumulation and the Intertemporal Elasticity of Substitution of Labor: How Large is the Bias?, Review of Economic Dynamics, 14, 577-91.

Wolff, E. (2010), Recent Trends in Household Wealth in the United States: Rising Debt and Middle-Class Squeeze - an Update 2007, Levy Economics Institute, Working Paper 589.

A Appendix

A.1 Data

SSA Data We use Social Security Administration (SSA) earnings data from Guvenen, Ozkan and Song (2013). 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 using the 2005 Personal Consumption Expenditure deflator. See Guvenen et. al. (2013) 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 Type-1 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 mean earnings for earnings beyond the 99th percentile that is implied by the Pareto distribution for that age and year. The mean follows the formula $E[e|e \geq e_{99}] = \frac{\lambda e_{99}}{\lambda - 1}$.

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 between the ages of 23 and 62 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.

Top-coding and bracketed variables HPV impute a value to top-coded observations of each component of earnings. A Pareto distribution is fitted to the non-top-coded upper end of the observed distribution and the imputation value is the distribution's mean conditional on the earnings component being above the top coding threshold. Also, in some of the early survey years, some of income variables were bracketed. HPV impute the midpoint of the corresponding bracket to these variables and 1.5 times the bottom of the top bracket for observations in the top bracket.

Age-Year Cells We split the dataset into age-year cells, compute the relevant moment within each cell and then collapse the dataset so there is a single observation per age-year cell. We put a PSID observation in the (a, y) cell if the interview was conducted during year $y = 1968, 1970, 1971, ..., 1996$ or $y = 1998, 2000, 2002$ with reported head of household's age a in the interval $[a, a+4]$. The life-cycle profiles we calculate correspond to $(\beta_{23}+d, \beta_{24}+d, \beta_{25}+d, ..., \beta_{63}+d)$, where the β_a are the estimated age coefficients and d is a vertical displacement selected in the manner described in section 3.

Table A1: Tax Rates and Tax Brackets

n	Brackets q_n	Federal Rate $100 \times R_n$	Combined Rate $100 \times R_n$
1	0	0.0	7.5
2	0.05	10.0	17.5
3	0.1	15.0	22.5
4	0.25	25.0	32.5
5	0.45	28.0	35.5
6	0.65	33.0	40.5
7	1.13	35.0	42.5

Note: Tax brackets are expressed as multiples of the 99th percentile of U.S. income distribution in 2010. Tax brackets come from 2010 IRS Form 1040 Instructions (Schedule Y-1, pg. 98). The 99th percentile data comes from the World Top Incomes Database.

A.2 Tax Function

This appendix describes how the first component of the model income tax function, discussed in section 4, is implemented.

Step 1: Specify the empirical tax function $\hat{T}(x)$ using the ordered pairs $\{(q_1, R_1), ..., (q_N, R_N)\}$.

$$\hat{T}(x) = \begin{cases} R_1[x - q_1] & i(x) = 1 \\ \sum_{n=2}^{i(x)} R_{n-1}[q_n - q_{n-1}] + R_{i(x)}[x - q_{i(x)}] & i(x) > 1 \end{cases}$$

$$i(x) \equiv \max n \; s.t. \; n \in \{1, 2, ..., N\} \; and \; q_n \leq x$$

The values of $\{(q_1, R_1), ..., (q_N, R_N)\}$ in Table A1 are set based on the 2010 federal tax brackets and rates for taxable income for married couples filing jointly. Brackets come from Schedule Y-1 in the IRS Form 1040 Instructions for the 2010 tax year. Adding $\$18,700$ to each of the taxable income brackets from Schedule Y-1 generates total income cutoffs that produce these taxable income cutoffs in Schedule Y-1 for joint filers without dependents according to the NBER tax program TAXSIM for the 2010 tax year. We state the brackets q_n as multiples of the 99-th percentile of the U.S. income distribution (including capital gains) for the year 2010.[20] Finally, we uniformly increase these federal tax rates by 7.5 percent so that the combined rate for the highest bracket is 42.5 - the top combined rate calculated by Diamond and Saez (2011).

Step 2: Specify the model tax function $T(x)$ using a 5th order polynomial P, where x denotes the sum of earnings and social security transfers in the model:

$$T(x) = \begin{cases} \kappa P(x/\kappa; \zeta) & x \leq \kappa \\ \kappa P(1; \zeta) + \bar{\tau}[x - \kappa] & x > \kappa \end{cases}$$

We set the coefficients ζ of the polynomial P to minimize the distance $\sum_{x_i \in X^{grid}} (\hat{T}(x_i) - P(x_i; \zeta))^2$ subject to $P(0; \zeta) = 0$ and $P'(1; \zeta) = 0.425$. X^{grid} contains 100 points uniformly distributed on the interval $[0, q_7]$. This implies that $\zeta = (0.0, 0.093, 0.472, -0.341, 0.099)$. We set $\kappa = q_7 \times I_{99}^{model} = 1.13 \times I_{99}^{model}$ and $\bar{\tau} = R_7 = 0.425$. The quantity I_{99}^{model} is the 99th percentile of income in the benchmark model economy. This quantity has to be computed for the benchmark model in an iterative procedure as the model tax system is specified as a function of I_{99}^{model} and $\bar{e}$.

In summary, the model tax function in step 2 approximates the empirical tax function with a polynomial. The polynomial is restricted to produce zero taxes at zero income and to produce a

[20]The World Top Incomes Database reports that the US 99th percentile for income in 2010 was \$365,026 (reported in 2012 dollars). The 99th percentile is then \$348,177 after converting to 2010 dollars using the CPI.

42.5 percent marginal tax rate at the start of the top tax bracket. Beyond the top tax bracket, the model tax function has a marginal tax rate set equal to the empirical top rate $\bar{\tau} = 0.425$. The tax function in the benchmark reform differs from the function specified here only via changes in the top rate $\bar{\tau}$ and the resulting lump-sum transfer. Figure 2 in the main text displays the marginal earnings tax rates arising from the model income tax system.

A.3 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 $(K/L, \bar{T})$. 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. $c + k' \leq whl + k(1 + r) - T_j(whl, kr; \bar{\tau}, \bar{T})$ and $k' \geq 0$

   ii. $h' = H(h, s, z', a)$ and $0 \leq l + s \leq 1$
3. Compute $(K', L', \bar{T}')$ implied by the optimal decision rules in step 2.
4. If $K'/L' = K/L$ and $\bar{T}' = \bar{T}$, then stop. Otherwise, update the guesses and repeat 1-3.

Comments:

Step 1: A guess for the lump-sum transfer $\bar{T}$ corresponding to the top rate $\bar{\tau}$ is needed as the model tax system is specified as a function of these values.
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 bilinear interpolation on (k', h'). To compute expectations, follow Tauchen (1986) and discretize the distribution of the shock variable z' with 11 equi-spaced log shocks lieing 3 standard deviations on each side of the mean.
Step 3: Compute aggregates (K', L') as follows. First, consider initial conditions $x = (h, a) \in X_1^{grid}$. For each $x \in X_1^{grid}$, draw $N = 2000$ random histories z^J from the distribution resulting from applying the Tauchen procedure. Use the decision rules from step 2 to compute lifetime histories. Set $E[k_j(x, z^j)|x] = \frac{1}{N}\sum_{n=1}^{N} k_j(x, z_n^j)$ and $E[h_j(x, z^j)l_j(x, z^j)|x] = \frac{1}{N}\sum_{n=1}^{N} h_j(x, z_n^j)l_j(x, z_n^j)$, where z_n^j is the n-th draw of the shock history. Compute aggregates as indicated below, where $\psi(x)$ is the probability of $x \in X_1^{grid}$. Appendix A.4 describes how $(X_1^{grid}, \psi(x))$ are set. Shock histories are fixed across all iterations in the Main Algorithm. The lump-sum transfer condition $\bar{T}' = \bar{T}$ holds when aggregate taxes implied from the computed decision rules equal G.

$$K' = \sum_{x \in X_1^{grid}} \sum_{j=1}^{J} \mu_j E[k_j(x, z^j)|x]\psi(x)$$

$$L' = \sum_{x \in X_1^{grid}} \sum_{j=1}^{J} \mu_j E[h_j(x, z^j)l_j(x, z^j)|x]\psi(x)$$

Setting Model Parameters: Following the discussion in section 4, some model parameters are fixed and the remaining model parameters are set based on an iterative procedure that involves guessing the parameter vector, computing equilibria and then revising the guess until the distance between equilibrium model values and data values is minimized. 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 set model parameters. When we set model parameters, the parameters of the tax system $(\bar{e}, I_{99})$ need to be chosen in an iterative way as the tax system is specified as a function of these endogenous values.

The algorithm to compute the Laffer curve for the model economy where the human capital process is exogenous 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 $(K/L, \bar{T})$. 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. $c + k' \leq w\bar{h}l + k(1 + r) - T_j(w\bar{h}l, kr; \bar{\tau}, \bar{T})$ 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 $(K', L', \bar{T}')$ implied by the optimal decision rules in step 2.
4. If $K'/L' = K/L$ and $\bar{T}' = \bar{T}$, then stop. Otherwise, update the guesses and repeat 1-3.

A.4 Initial Conditions

Construct a bivariate distribution from a univariate distribution. Assume that learning ability is distributed according to a Pareto-Lognormal distribution and then construct a bivariate distribution based on assumptions A1-2 below.

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 $z \equiv y^{\beta_1}$.

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

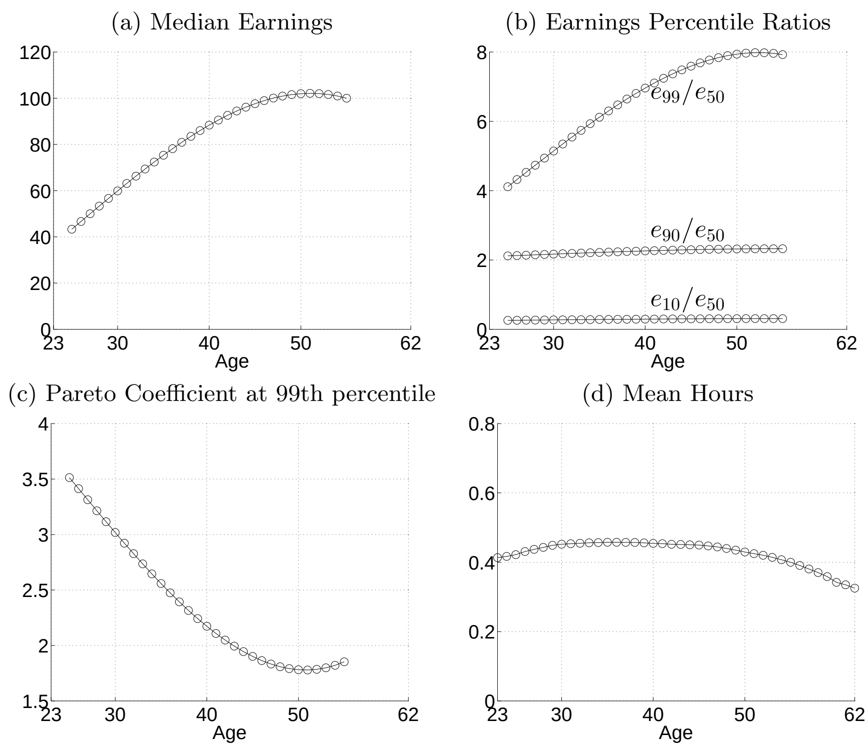
The second term on the right hand side is distributed Type-1 Pareto with scale parameter 1 and shape parameter λ_a/β_1. $\|$

We now discretize this bivariate distribution. First, construct a discrete approximation (a_i, P_i) for $i = 1, 2, 3, ..., 9$. Set $(P_1, ..., P_9) = (0.225, ..., 0.225, 0.06, 0.03, 0.005, 0.004, 0.001)$. Given the probabilities, set learning ability levels to equal conditional means implied by the marginal distribution $F(a)$ implied by $PLN(\mu_a, \sigma_a^2, \lambda_a)$.

$$\begin{aligned} a_1 &= E[a|a \leq F^{-1}(P_1)] \\ a_9 &= E[a|a \geq F^{-1}(1 - P_9)] \\ a_i &= E\left[a|F^{-1}(\sum_{j=1}^{i-1} P_j) \leq a \leq F^{-1}(\sum_{j=1}^{i+1} P_j)\right] \text{ for } 1 < i < 9. \end{aligned}$$

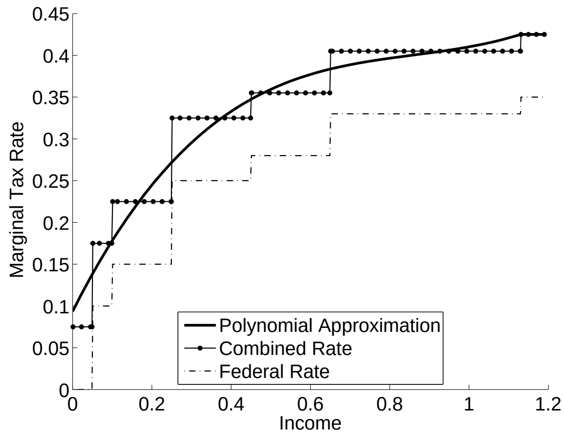
Second, for any $a \in A^{grid} = \{a_1, ..., a_9\}$, specify a 20 point human capital grid that is equi-spaced in log human capital units and that ranges 3 standard deviations above and below the conditional mean implied by a and assumption A2. Probabilities P_j for $j = 1, ..., 20$ are set following Tauchen (1986). This then implies that X_1^{grid} is a 9 by 20 grid and that $\psi(x) = P_i P_j$ for $x = (a_i, h_j) \in X_1^{grid}$.

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



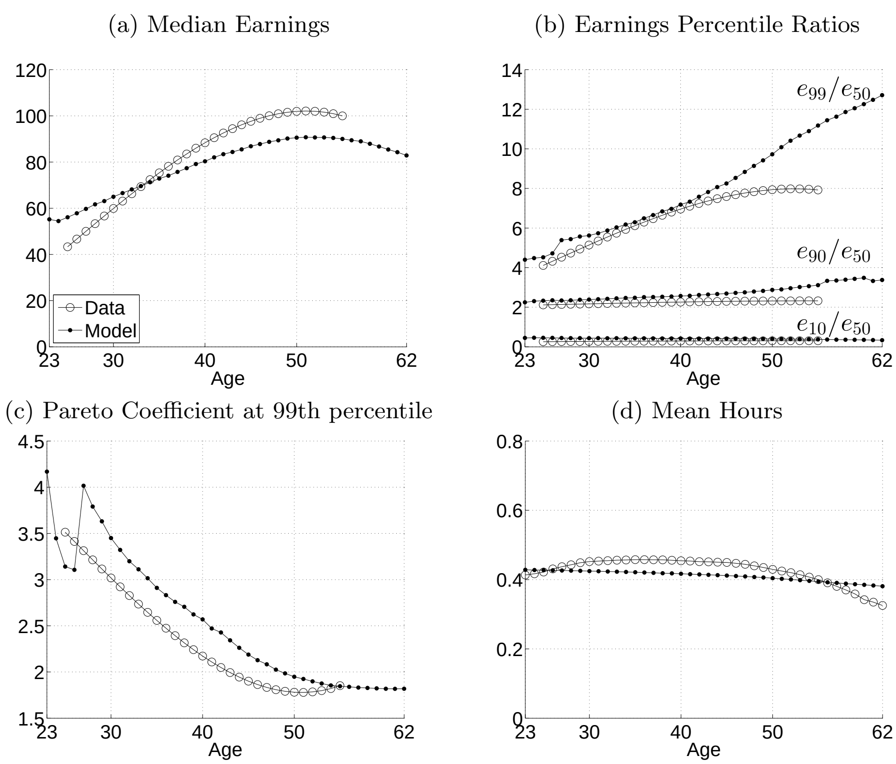
Note: Earnings profiles are based on SSA data. Hours profiles are based on PSID data.

Figure 2: Model Tax System



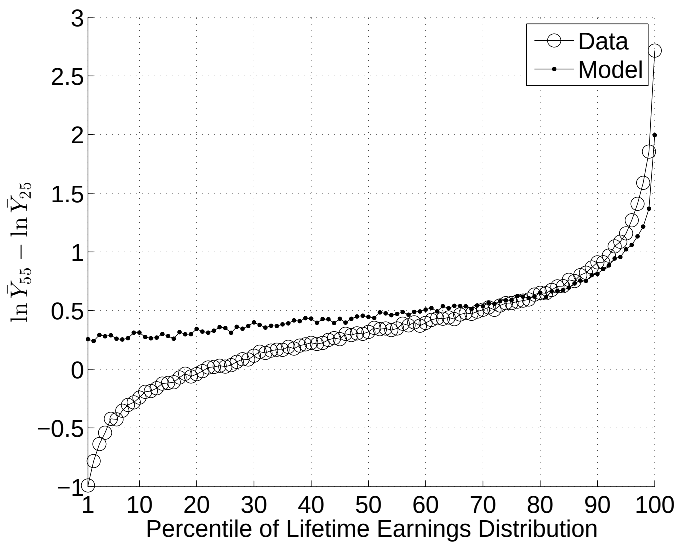
Note: The horizontal axis measures income in multiples of the 99th percentile of income.

Figure 3: Life-Cycle Profiles: Data and Model



Note: Large open circles describe profiles for the U.S. economy. Small solid circles describe profiles for the model economy.

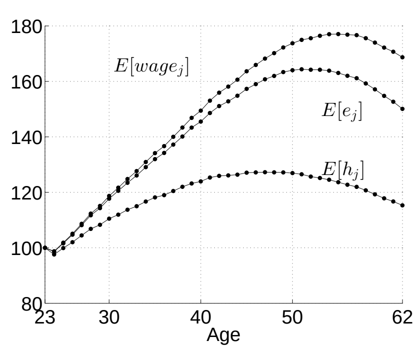
Figure 4: Growth of Mean Earnings by Percentile of Lifetime Earnings



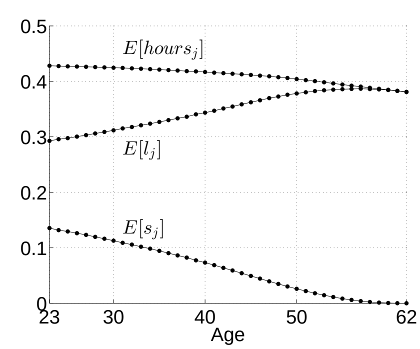
Note: Data results are taken from Guvenen, Karahan, Ozkan and Song (2014) based on SSA data. The vertical axis plots $\ln \bar{Y}_{55} - \ln \bar{Y}_{25}$, where $(\bar{Y}_{25}, \bar{Y}_{55})$ are mean earnings at these ages for groups based on percentiles of the present value of lifetime earnings.

Figure 5: Life-Cycle Mean Model Profiles

(a) Earnings, Wage and Human Capital



(b) Time Learning, Working and Total



Note: Earnings, wages and human capital in Figure 5(a) are all normalized to equal 100 at age 23.

Figure 6: Laffer Curves

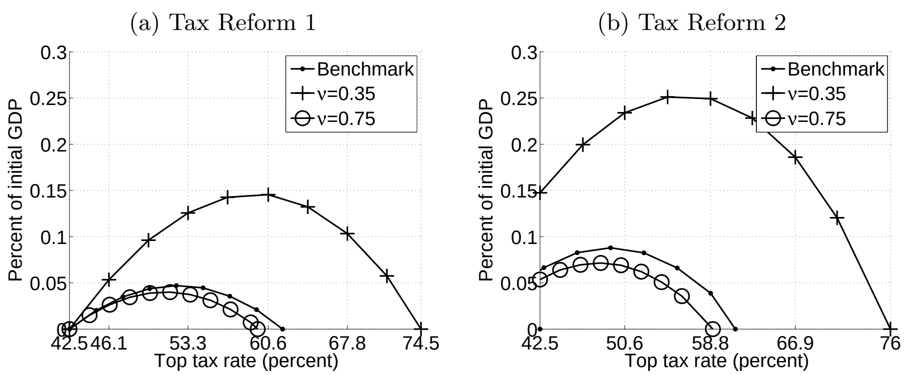
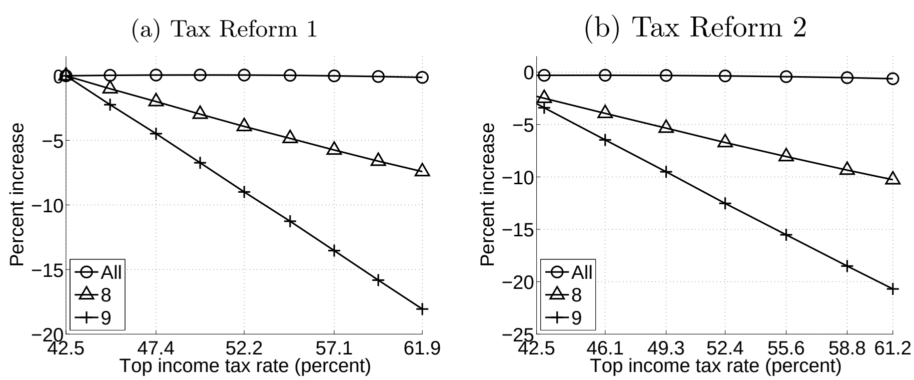
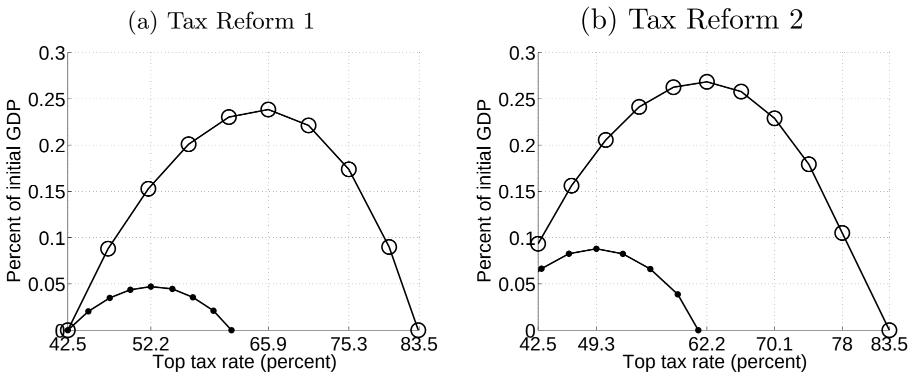


Figure 7: Equivalent Consumption Variation



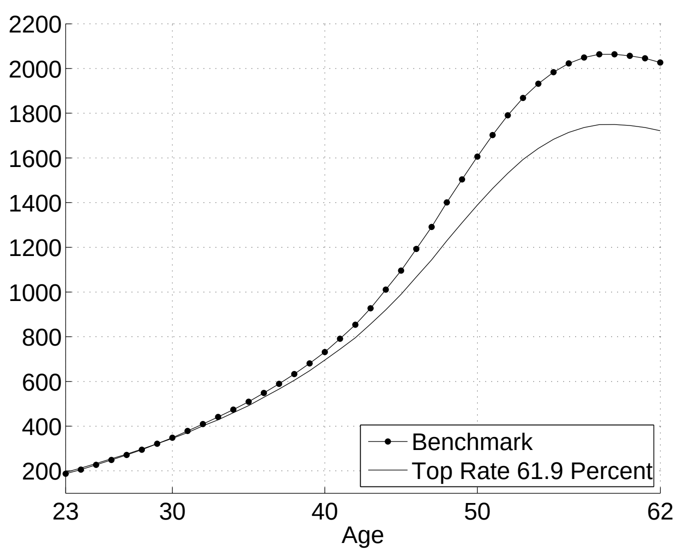
Note: Legend labels 1-9 denote results conditional on learning ability level. Level 1 is the lowest and level 9 is the highest. The legend label ALL denotes unconditional results in that expected utility is calculated averaging across all ability types.

Figure 8: Laffer Curves: Endogenous and Exogenous Human Capital



Note: Dots plot properties of the endogenous human capital model, whereas open circles plot properties of the exogenous human capital model.

Figure 9: Mean Earnings Profiles



Note: Mean earnings by age are taken over all agents with ability levels 7 through 9. These ability levels constitute the top 1 percent of the population by ability level.