

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: SEPTEMBER 9, 2015

Abstract

We assess the consequences of 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) there is a formula featuring three elasticities that accurately predicts the top of the model Laffer curve and (3) standard empirical methods underestimate the long-run model earnings elasticity that enters this formula.

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

Diamond and Saez (2011) argue that the revenue maximizing marginal earnings tax rate on top earners in the U.S. is approximately 73 percent, but calculate that the top rate in 2010 is roughly 43 percent. They also argue that the top rate that maximizes revenue will approximate the top rate that maximizes welfare for some welfare measures. Therefore, they propose a 73 percent top tax rate.¹

There is a simple formula for the revenue maximizing top tax rate in some static models that depends on only two inputs: an earnings elasticity for top earners in response to a change in the top tax rate and a statistic of the upper tail of the earnings distribution. The formula does not rely on making parametric assumptions on the primitive elements of a specific static model. Diamond and Saez (2011) employ this formula and estimates of these two inputs to calculate that 73 percent is the revenue maximizing top rate.

From the perspective of human capital theory, we argue that there are two main problems with their analysis and that both of these lead to a lower revenue maximizing top rate. First, there is a theoretical problem: the simple formula is not valid in dynamic models. However, there is a formula (see Badel and Huggett (2015, Theorem 1)), featuring three elasticities rather than one elasticity, that applies to static models and to steady states of dynamic models. We show that accounting for the two additional elasticities in the Badel-Huggett formula reduces the value of the revenue maximizing top tax rate within a human capital model.

The second problem is that the earnings elasticity estimate that Diamond and Saez employ is a short-run estimate. It is widely agreed that the long-run elasticity is of greatest interest for policy making. This is because the central question involves the consequences of a permanent change in the top tax rate. We show that when one applies standard empirical methods for estimating the short-run elasticity (e.g. the methods developed by Gruber and Saez (2002)) to data produced by a human capital model, then the estimated short-run elasticity is systematically below the true long-run model elasticity that enters the Badel-Huggett formula and determines the top of the model Laffer curve. This occurs because skills fall after an increase in the top tax rate but the fall in skills takes a long time to be fully realized. A larger earnings elasticity implies a smaller revenue maximizing tax rate in either the simple formula or the Badel-Huggett formula.

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. The model

¹The top federal tax rate in the US in 2010 was 35 percent. It applies to taxable income levels above \$373,650 for joint filers. The 99th percentile of the U.S. income distribution in 2010 was \$365,026 with capital gains and \$353,632 without capital gains when stated in 2012 dollars, according to the World Top Incomes Database. Thus, Diamond and Saez (2010) propose a substantial tax rate increase on roughly the top 1 percent.

is calibrated to match features of the U.S. age-earnings distribution. The model Laffer curve relates the value of the top tax rate to the resulting steady-state, lump-sum transfer and holds government spending constant. The peak of the model Laffer curve occurs at a 52 percent top tax rate.

We show that the Laffer curve would peak at a 66 percent top tax rate absent skill change in response to an increase in the top tax rate. Thus, the endogenous fall in skill in response to an increase in the top rate is quantitatively very important within the model. Why do skills fall in the steady state associated with the higher top tax rate? The key mechanism is that a higher top rate reduces the marginal benefits of skill investment received later in life without changing the marginal cost of skill investment earlier in life. For this mechanism to work it is key that top earners typically become top earners late in life. There is strong support for this in U.S. data: males with high lifetime earnings have on average a very steep mean earnings profile.

The paper is organized as follows. Section 2 presents the model framework. Section 3 documents properties of the U.S. age-earnings distribution. Section 4 and 5 set model parameters and describe model properties. Section 6 assesses the consequences of increasing the marginal tax rate on top earners and examines all the relevant elasticities in the Badel-Huggett formula within the human capital model. Section 7 concludes.

2 Framework

The model that we employ is closest to the human capital model developed by Huggett, Ventura and Yaron (2011).²

Decision Problem In Problem P1 an agent maximizes expected utility which is determined by consumption $c = (c_1, \dots, c_J)$, work time decisions $l = (l_1, \dots, l_J)$ and learning time decisions $s = (s_1, \dots, s_J)$. Consumption c_j , work time l_j and learning time s_j decisions at age j are functions of initial conditions $x = (h_1, a) \in X$ and shock histories $z^j = (z_1, \dots, z_j)$. 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.

²Heckman, Lochner and Taber (1998), Erosa and Koreshkova (2007) and Guvenen, Kuruscu and Ozkan (2014) analyze implications of varying income tax progression using a human capital model. Unlike our paper, none of these papers focus on tax reforms directed at the extreme upper tail. Altig and Carlstrom (1999) Guner, Lopez-Daneri and Ventura (2014) and Kindermann and Krueger (2014) analyze tax reforms that are directed at the upper tail of the income distribution. A key difference from our work 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.

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

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

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 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 of agents, population fractions $(\mu_1, \mu_2, \dots, \mu_J)$ 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.
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. The parameter χ in the utility function affects the growth rate of total labor time ($l_j + s_j$) over the working lifetime. The utility function parameter ϕ affects the mean of the total labor time. The law of motion for human capital H takes the functional form used in Ben-Porath (1967). The shocks z to an agent's stock of human capital are independent and identically distributed across periods and are normally distributed. Shocks are idiosyncratic in that a known fraction of agents receive a shock lying in any particular set of interest. The distribution ψ of initial conditions has the property that the marginal distributions for learning ability and initial human capital are both Pareto-Log-Normal (PLN) distributions - see Appendix B.2. This bivariate distribution is characterized by 6 parameters.³

Benchmark Model Functional Forms:

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

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

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

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

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 (2014) and PSID male earnings and hours data from Heathcote, Perri and Violante (2010). These data sets are described in Appendix A.1.

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

³We 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.

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. A similar age and time effects regression is applied to mean hours in PSID data and is described in Appendix A1. Hours are stated as a fraction of total discretionary time which is set to 14 hours per day times 365 days per year.

The Pareto statistic at the 99th percentile for age group j in year t is the mean earnings $\bar{e}_{j,t}^{99}$ for observations above this percentile divided by $\bar{e}_{j,t}^{99}$ less the 99th percentile $e_{j,t}^{99}$. 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 enters the revenue maximizing tax rate formula used by Diamond and Saez (2011) and the tax rate formula in Badel and Huggett (2015).

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

Figure 1 highlights the results. Median earnings more than double over the working lifetime. 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. It doubles from a ratio of near 4 at age 25 to a ratio of near 8 at age 55. Thus, earnings dispersion increases with age in the upper half of the distribution. The Pareto statistic decreases with age and is below 2.0 after age 45. Figure 1 also shows that the mean work hours profile is hump-shaped but fairly flat with age.

We examine 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 greater than for the time effects case. Second, we analyze earnings using PSID data rather than SSA data. The age effects based on PSID data display the same qualitative behavior as the age effects based on SSA. However, measures of earnings dispersion display greater dispersion at a given age in SSA data than in PSID data.

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

Table 1 - Parameter Values

Category	Functional Forms	Parameter Values
Demographics	$\mu_{j+1} = \mu_j / (1 + n)$	$Retire = 41, n = 0.012$ $j = 1, \dots, 63$ (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	$H(h, s, z, a) = \exp(z) [h + a(hs)^\alpha]$ 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.

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). Their estimate is based on log wage rate variation towards the end of the working lifetime. Third, the remaining model parameters are set so that equilibrium properties of the model best match empirical targets, including those displayed in Figure 1.⁵ Appendix B.1 describes the computation of an equilibrium.

Demographics 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 $(1 + n)$.

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).

⁵Describe here the function that we minimize.

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 $T_j^{ss}(e_j) = \tau^{ss} \min[e_j, e_{max}]$ 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 $b = 0.4$.⁶ 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 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).⁷ 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.⁸ 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.

⁶We 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).

⁷Their calculation accounts for federal and state income taxes, uncapped medicare taxes, average sales taxes and rules on the deductibility of various taxes.

⁸Guner, Kaygusuz and Ventura (2013) document this fact using Internal Revenue Service data.

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

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, 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 the literature on the Frisch elasticity of labor supply that regresses the change in log labor hours on the change in a log wage measure and a constant term. 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. The small value of this regression coefficient is often viewed as suggesting that the revenue maximizing top tax rate will be quite large.

$$\Delta \log \text{hours} = \alpha_0 + \alpha_1 \Delta \log \text{wage} + \epsilon$$

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. Section 5 and Appendix B.3 discusses the results of the estimation of the regression equation and the construction of model data sets.

5 Properties of the Model Economy

5.1 Age-Earnings Distribution

Figure 3 highlights model properties for a number of statistics that were directly targeted in setting model parameters. The model produces a hump-shaped median and mean earnings profile by a standard human capital mechanism. 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 levels fall. This occurs because time allocated to learning goes to zero towards the end of the working lifetime and because the mean of the multiplicative shock to human capital is below one (i.e. $E[\exp(z)] = \exp(\mu_z + \frac{\sigma_z^2}{2}) < 1$). Thus, on average skills depreciate.

Measures of earnings dispersion increase with age in U.S. data. Specifically, 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 using 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.

5.2 Distribution of Initial Conditions

Simple summary measures of the distribution of initial conditions are given in Table 2. The distribution is based on the discrete approximation described in Appendix B.2. Human capital follows a right-skewed distribution with a mean-median ratio of 1.24 and a coefficient of variation equal to 0.73. The coefficient of variation of learning ability is 0.34. Thus, the model requires a source for increasing earnings dispersion with age beyond that from idiosyncratic risk to produce the increase in the 90-50 and 99-50 ratio observed in U.S. data.

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

Log learning ability and log human capital are positively correlated at age 23. The correlation in levels, rather than log units, is 0.60. The positive correlation is consistent with Huggett, Ventura and Yaron (2006, 2011) who argue that a zero correlation would tend to produce

a U-shaped earnings dispersion pattern with age not found in U.S. data and that a positive correlation eliminates such counterfactual implications.

The positive correlation has two important model implications: (i) high learning ability agents will tend to have high lifetime earnings and (ii) agents with high lifetime earnings will have very high earnings growth rates over the working lifetime. This implies that top earners are disproportionately older agents with very high learning ability. For agents with very high learning ability, increasing the top tax rate acts as an increased tax on the benefits of skill investments while leaving the marginal cost of these skill investments made earlier in life unchanged. This is the key mechanism behind the fall in skills. For this mechanism to work it is key that top earners typically become top earners later in life. Section 6 shows that males with very high lifetime earnings in US data have extremely large average earnings growth rate over the working lifetime.

5.3 Mean Earnings, Wage and Human Capital Profiles

Figure 4 highlights the mean profiles for earnings, wage rates, human capital and hours. The mean wage rate profile is steeper than the mean human capital profile. This occurs because the total hours profile is flatter than the work time profile. 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. Wallenius (2011) also makes this assumption.

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. For example, MaCurdy (1981) uses PSID data for white males age 25-55 and finds a regression coefficient of 0.125.⁹ 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. Domeij and Floden (2006) find similar results based on PSID data over a longer time period. This type of evidence is behind the view that labor hours are not very elastically supplied by prime-age males.¹⁰ This evidence has also been used to support the view that very high top tax rates may be revenue maximizing.

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

⁹This is the average of the point estimates from MaCurdy (1981, Table 1 row 5-6).

¹⁰See Keane (2011) and Keane and Rogerson (2012) for recent reviews of the literature that examines this regression equation. MaCurdy (1981) argues that within exogenous-wage models the regression coefficient α_1 is an estimate of the preference parameter ν under appropriate conditions.

We set the parameters of the human capital model to minimize the distance between data statistics and model statistics. One of these data statistics is the regression coefficient $\alpha_1 = 0.125$. The model counterpart to the empirical regression coefficient is based on the 25-55 age group when $wage_j = e_j/(l_j + s_j)$. The model produces a regression coefficient of $\alpha_1 = 0.114$. Appendix B.3 offers an interpretation for why the regression coefficients in Table 3 lie below the value of the utility function parameter $\nu = .551$ from Table 1 and relates our work to findings in Domeij and Floden (2006).

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 $\nu = 0.551$. Appendix B.3 describes the instrumental variables regression method and the construction of the model data sets.

5.5 Earnings, Income and Wealth Distributions

Table 4 compares statistics of the distribution of earnings, income and wealth in the model economy with those from the U.S. economy. The model produces an income distribution that does not concentrate as much income in the upper tail as 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 excluding capital gains and average each statistic over the period 1978-2011. We measure the income facts averaged over the 1978-2011 period since the earnings process in the model targets earnings statistics calculated from data over the same period. Over this period the top income share has a strong upward trend. The model produces more than half of the fraction of wealth held by the top 1 percent of U.S. households. It is well known, see Huggett (1996), that life-cycle models that are calibrated to match features of the U.S. age-earnings distribution have difficulty matching the wealth held by the top 1 percent.

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	12.7	27.4	-	2.03
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 from tabulations of the SSA data set constructed by Guvenen et al. (2013). (2) US income distribution facts are averages over the years 1978-2011 based on income data from World Top Incomes Database that exclude capital gains. (3) US wealth distribution facts are from Diaz-Gimenez, Glover and Rios-Rull (2011) based on the 2007 Survey of Consumer Finances.

6 Assessing the Tax Reform

6.1 Laffer Curve

We analyze the Laffer curve implied by a reform that 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, given that government spending is held constant.

Figure 5 displays the 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 output. The Laffer curve in the benchmark model peaks at a tax rate of roughly 52 percent. The transfer is below 0.05 percent of the initial steady-state output. Thus, the Laffer curve in the benchmark model is flat in that little additional revenue is raised. The top of the Laffer curve occurs at a tax rate that is well below the 73 percent top rate that Diamond and Saez (2011) highlight as revenue maximizing.

Figure 5 also displays Laffer curves for different values of the utility function parameter ν . We consider two alternative values $\nu = 0.35$ and $\nu = 0.75$ and repeat the estimation procedure from section 4, choosing the remaining parameters to best match targets.¹¹ The revenue maximizing top tax rate increases as the parameter ν decreases.

Figure 6 plots a measure of welfare gains associated with the tax reform 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

¹¹The 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).

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 steady state under the new tax system. The equivalent consumption gain is a small fraction of 1 percent and is presented in Figure 6 under the legend label ALL. Intuitively, the gain is small because the transfer is small and factor prices change very little as both aggregate capital and labor fall by approximately the same percentage.

Figure 6 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.

For agents with learning ability at or below median ability, the equivalent consumption welfare measure has the same qualitative pattern as the Laffer curve. 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 steady states because aggregate capital and labor inputs fall by a similar percentage.

6.2 Understanding the Role of Human Capital Accumulation

What role does skill change play in accounting for the shape of the model Laffer curve? To answer this question, consider an alternative economy that has the same preferences, technology, initial conditions and tax system as the benchmark model. The alternative economy and the human capital model are observationally equivalent in steady state under the benchmark tax system in terms of consumption, wealth, earnings and income. The key difference is that when the tax system changes then human capital investments change in the benchmark model but remain unchanged in the exogenous human capital model.

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 B.1 describes the computation of equilibria in this model.

Figure 7 plots the Laffer curve in the two models. The top of the Laffer curve for the exogenous human capital model raises roughly five times as much extra revenue compared to the benchmark human capital model. Thus, endogenous skill change flattens out the Laffer curve

compared to an otherwise similar model that ignores the possibility of skill change in response to changes in the tax system. The top of the Laffer curve in the exogenous human capital model occurs at a much higher top tax rate equal to 66 percent under the benchmark reform. Intuitively, the Laffer curves differ because aggregate labor input L is more elastic with respect to a change in the top rate in the human capital model. When we decompose the change in the aggregate labor input across steady states, more than half of the fall in the aggregate labor input from the original steady state to the steady state with top rate set to 52 percent is due to skill change as opposed to changes in work time at fixed skills. By far the largest percentage fall in skills comes from agents with high learning ability late in the working lifetime.

$$wh_j(1 - \tau'_j) = \sum_{k=j+1}^{Retire-1} \left(\frac{1}{1 + \hat{r}}\right)^{k-j} \frac{dh_k}{ds_j} w l_k (1 - \tau'_k)$$

A mechanism behind the fall in skill is easily grasped from the Euler equation for skill investment above. Abstracting from idiosyncratic risk for simplicity, at a best choice an agent equates the marginal cost of an extra unit of time spent in skill production at age j to the discounted marginal benefit of the extra skill production $\frac{dh_k}{ds_j}$ in future periods, where \hat{r} is the after-tax real interest rate. Now consider an increase in the top tax rate. Absent any adjustment, the left-hand side of the Euler equation does not change for an agent with earnings below the top tax rate but some of the marginal net-of-tax-rate terms $(1 - \tau'_k)$ decrease for an agent that will be above the threshold in the future. Thus, some adjustment must occur. A decrease in time investment in skill production increases the future marginal product terms $\frac{dh_k}{ds_j}$. If future labor hours l_k decrease in response to the increase in the top tax rate, consistent with model behavior for agents with high learning ability, then an even larger fall in skill investment occurs at age j .

6.3 A Sufficient Statistic Formula

Badel and Huggett (2015, Theorem 1) derive a formula that states the revenue maximizing tax rate τ^* in terms of three elasticities. Their formula applies to static models and to steady states of dynamic models. It is derived based on three model elements: (i) a probability space of agent types (X, \mathcal{X}, P) , (ii) functions (y_1, \dots, y_n) that map agent type $x \in X$ and a top tax rate τ into income and expenditure decisions and (iii) a tax function T . The tax function is assumed to be separable in that $T(y_1, \dots, y_n; \tau) = T_1(y_1; \tau) + T_2(y_2, \dots, y_n)$. In addition, T_1 has a constant top tax rate τ that applies to income levels y_1 beyond a threshold \underline{y} .

The Badel-Huggett formula is stated below. It differs from the widely-used formula (i.e. $\tau^* = 1/(1 + a\epsilon)$) that Diamond and Saez employ in that there are two extra terms in the numerator.

The formula is based on aggregate variables that are calculated by integrating individual-level variables over the sets $X_1 \equiv \{x \in X : y_1(x, \tau^*) > \underline{y}\}$ and $X_2 \equiv \{x \in X : y_1(x, \tau^*) \leq \underline{y}\}$.

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

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

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

The Badel-Huggett formula applies to static models and to steady states of dynamic models, whereas the widely-used formula applies to some static models. The Badel-Huggett formula handles two issues that are not accounted for by the widely-used formula. First, it allows the total taxes $\int_{X_2} T(y_1, \dots, y_n; \tau) dP$ paid by agent types below the threshold to vary as the top tax rate varies. The term $a_2\epsilon_2$ in the formula is non-zero when this occurs. Agent types below the threshold are in the set X_2 . In the human capital model $\epsilon_2 > 0$ because agents with high learning ability, who are below the threshold early in life, anticipate crossing the threshold with positive probability later in life. Their investments in human capital fall as $(1 - \tau)$ decreases. Second, the Badel-Huggett formula accounts for the possibility that agent types above the threshold pay other taxes $\int_{X_1} T_2(y_2, \dots, y_n) dP$ which vary as the top tax rate varies. The term $a_3\epsilon_3$ is non-zero when this occurs. In the human capital model $\epsilon_3 > 0$ because top earners accumulate less wealth and pay less in capital income taxes when the net-of-tax rate $(1 - \tau)$ decreases.¹²

We now calculate all the terms in the top tax rate formula. We do so by mapping equilibrium variables in the human capital model into the three model elements used in deriving the top tax rate formula. To apply the formula, define an agent type $x = (h_1, a, j, z^j)$ to be initial conditions (h_1, a) , age j and (partial) shock history z^j . Let y_1 be labor income, y_2 be social security transfer income and y_3 be capital income.¹³

¹²In some static models, such as the Mirrlees (1971) model, the two extra terms $a_2\epsilon_2$ and $a_3\epsilon_3$ in the formula are both zero. This issue and other issues related to interpreting and applying this formula are discussed in detail in Badel and Huggett (2015).

¹³The notation employed in defining (y_1, y_2, y_3) emphasizes that equilibrium factor prices and decisions depend on the top tax rate τ .

Table 5 - Revenue Maximizing Top Tax Rate Formula

Terms	endogenous human capital model	exogenous human capital model
$a_1 \times \epsilon_1$	$1.97 \times .396 = .780$	$1.97 \times .240 = .473$
$a_2 \times \epsilon_2$	$3.06 \times .019 = .059$	$3.06 \times .014 = .044$
$a_3 \times \epsilon_3$	$.043 \times .508 = .022$	$.043 \times .459 = .020$
$\tau^* = \frac{1-a_2\epsilon_2-a_3\epsilon_3}{1+a_1\epsilon_1}$.52	.64
τ at peak of Laffer curve	.52	.66

Note: The elasticities ($\epsilon_1, \epsilon_2, \epsilon_3$) are calculated by a difference quotient using the initial steady state value $\tau = 0.425$ and the value $\tau = 0.52$. The sets X_1 and X_2 are defined using the threshold \underline{y} in Figure 1 at which the top tax rate begins and using $\tau = .425$.

$$y_1(x, \tau) \equiv \begin{cases} w(\tau)h_j(h_1, a, z^j; \tau)l_j(h_1, a, z^j; \tau) & j < \textit{Retire} \\ 0 & \textit{otherwise} \end{cases}$$

$$y_2(x, \tau) \equiv \begin{cases} 0 & j < \textit{Retire} \\ b\bar{e} & \textit{otherwise} \end{cases}$$

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

The tax function $T(y_1, y_2, y_3; \tau)$ has two components. The component $T_1(y_1; \tau) = T^{prog}(y_1; \tau) + \tau^{ss} \min\{y_1, \textit{emax}\}$ combines the progressive taxation of earnings and the social security earnings tax in the model. The remaining component $T_2(y_2, y_3) = T^{prog}(y_2) - y_2 + \tau^{cap}y_3$ captures the taxation of capital income and the net taxation of social security income.

Table 5 calculates the coefficients (a_1, a_2, a_3) and the elasticities ($\epsilon_1, \epsilon_2, \epsilon_3$) in the formula at the pre-tax reform steady state.¹⁴ For the moment, we take away three messages from Table 5: (1) the formula accurately predicts the top of the Laffer curve in both models, (2) the two “extra terms” in the numerator act to reduce the revenue maximizing rate compared to the widely-used formula and (3) the earnings elasticity of top earners $\epsilon_1 = 0.396$ is larger in the human capital model than in the exogenous human capital model $\epsilon_1 = 0.240$ and accounts for much of the difference in the peaks of the Laffer curves according to the formula.

6.4 Elasticities: Short-Term Evidence

Evidence on the elasticity of earnings or income in response to a change in the net-of-tax rate comes from the elasticity of taxable income literature. Saez, Slemrod and Giertz (2012)

¹⁴The formula in Badel and Huggett (2015) holds exactly when elasticities and coefficients are calculated at the revenue maximizing top tax rate. In practice, such formulae are used to predict the top of the Laffer curve. Thus, inputs are calculated away from the maximum as in Diamond and Saez (2010).

review this literature. Much of the literature applies the regression framework below, where the parameter ϵ is the elasticity, z_{it} is income of individual i at time t , $\tau_t(z_{it})$ is the marginal tax rate, $f(z_{it}) = \log z_{it}$ is an income control and α_t are time dummy variables.¹⁵

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

Saez et al. (2012) estimate elasticities using U.S. data from 1991-97 and the regression equation above. Their estimation is based on income variation before and after a tax reform from the Clinton administration that occurred in 1993. 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. Their panel consists of annual income histories for taxpayers with incomes in the top 10 percent of the income distribution in 1991. Saez et al. (2012, Table 2) estimates the regression equation above using three different choices of instruments for log net-of-tax-rate changes, different control variables and sample periods and a two-stage-least-squares estimator.¹⁶

In Table 6, we use the same regression equation specification, instrument definitions, sample definitions and estimation techniques (i.e. two-stage-least squares) and apply them to a (partial equilibrium) tax reform in the two models analyzed in Table 5. 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 $t = 1, \dots, 7$ correspond to the years 1991 – 1997 for the US economy. In period 3, the tax system is modified by increasing the top tax rate to $\bar{\tau} = 0.52$.

We take away two messages from Table 6. First, the human capital model produces an empirical elasticity in the range of 0.15 to 0.3. Diamond and Saez (2011) claim that a mid-range estimate for the short-term elasticity of top earners in the U.S. is $\epsilon = 0.25$. Thus, the models that we analyze are consistent with such evidence. While $\epsilon = 0.25$ may be a mid-range estimate for U.S. top earners, one should keep in mind that the estimates in the literature and in Saez et al. (2012, Table 1-2) vary widely. Second, when standard methods from the literature are applied to a tax reform within the human capital model then the estimated elasticity ϵ is systematically below the true long-run model elasticity $\epsilon_1 = 0.396$ that is relevant for determining the top of

¹⁵Saez et al. (2012, footnote 37) state that a static optimization problem with a quasi-linear utility function $u(c, z) = c - z^0 \frac{(\frac{z}{z^0})^{1+1/\epsilon}}{1+1/\epsilon}$ generates income response functions consistent with such a regression equation.

¹⁶Instrument 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 $1_{\{i \in T_{1992} \text{ and } t=1992\}}$. 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.

Table 6 - Elasticity Estimates

(a) Endogenous Human Capital Model

	(1)	(2)	(3)	(4)	(5)	(6)
Mean Elasticity ϵ	0.1536	0.2527	0.2688	0.2888	0.2806	0.2982
S.D.	(0.0254)	(0.0243)	(0.0251)	(0.0245)	(0.0228)	(0.0221)
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: $1_{\{i \in T_2 \text{ and } t=2\}}$			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
Long-run Model Elasticity ϵ_1	0.396	0.396	0.396	0.396	0.396	0.396

(b) Exogenous Human Capital Model

	(1)	(2)	(3)	(4)	(5)	(6)
Mean Elasticity ϵ	0.2128	0.2961	0.2925	0.3131	0.2892	0.3075
S.D.	(0.0238)	(0.0254)	(0.0239)	(0.0241)	(0.0233)	(0.0234)
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: $1_{\{i \in T_2 \text{ and } t=2\}}$			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
Long-run Model Elasticity ϵ_1	0.240	0.240	0.240	0.240	0.240	0.240

Note: (1) We draw 100 balanced panel data sets of 30,000 agents. Each data set mimics the structure of the data set used by Saez et al. (2012, Table 2). The agents in each balanced panel have labor income above the 90th percentile of earnings at $t = 1$. We follow these agents from periods $t = 1$ to $t = 7$. The tax reform occurs at $t = 3$. Factor prices are fixed in all periods. (2) We report means and standard deviations of the point estimates of ϵ across 100 randomly drawn balanced panels.

the Laffer curve.

The proximate explanation for why standard methods underestimate the relevant long-run model elasticity comes from examining the transition path produced by the reform. Figure 8 shows that aggregate labor input in the human capital model drops sharply in the first period of the reform (model period 3) and then gradually declines to the new lower steady-state level. Roughly half of the change in the labor input occurs in the first period of the reform. Thus, regression methods that use earnings or income changes a year or so after a reform will at best pick up an hours and factor price response in the human capital model but not the long-run skill response.

One might conjecture that measuring income and marginal tax rate changes more than a few periods after a reform may reduce the difference between estimated model elasticities and the true long-run model elasticity. To analyze this conjecture, we apply the regression specification and methods in Giertz (2010) to model data. He calculates one-year, three-year and six-year log income differences and adds extra controls to the basic regression equation from this section. Appendix B.4 shows that three-year and six-year differences imply lower and sometimes negative elasticity estimates compared to one-year differences in Table 6. Thus, our findings do not support the conjecture that increasing the time gap between measurements and applying the standard panel regression technique from the literature helps to better estimate the long-run model elasticity ϵ_1 .¹⁷

7 Discussion

This article assesses the consequences of increasing the marginal tax rate on top earners from the perspective of a human capital model. We highlight three points:

1. The top of the Laffer curve in the human capital model occurs at a top tax rate of 52 percent. The extra tax revenue produced at the 52 percent top rate is relatively small - less than one-tenth of one percent of the output in the model.
2. We provide a formula $\tau^* = (1 - a_2\epsilon_2 - a_3\epsilon_3)/(1 + a_1\epsilon_1)$ for the revenue maximizing top tax rate relevant for steady states of dynamic models. This formula accurately predicts the top of the model Laffer curve.

¹⁷The long-run elasticity ϵ_1 in the Badel-Huggett formula is based on how aggregate earnings for a set of “agent types” moves across equilibria as the tax rate parameter moves. Thus, conceptually the elasticity ϵ_1 does not involve measuring how earnings change across time periods. The empirical methods under review examine how an earnings or an income measure for a fixed collection of tax units changes across time periods.

3. There are two main reasons why we find a 52 percent revenue maximizing top rate, whereas Diamond and Saez (2011) argue that it is roughly 73 percent. First, we account for two extra terms ($a_2\epsilon_2$ and $a_3\epsilon_3$) that Diamond and Saez do not consider. The extra terms arise generally in dynamic models and are both positive in the human capital model. There are no estimates for these two extra elasticities in the literature. Second, we disagree on the value of the long-run elasticity ϵ_1 . It is $\epsilon_1 \doteq 0.4$ in the human capital model, whereas the Diamond-Saez view is that $\epsilon_1 = 0.25$ is a mid-range estimate for the short-term elasticity. They employ this short-term value, given the absence of convincing long-run estimates.

These differing views on the plausible value of ϵ_1 do not appear to come from differing views as to what is the central question and what type of responses are relevant to answer this question. The central question is what is the revenue maximizing tax rate for a permanent reform that changes only the top tax rate? The relevant response is then the long-term response. Saez et al. (2012, p. 13) state *“The long-term response is of most interest for policy making ... The empirical literature has primarily focused on short-term (one year) and medium-term (up to five year) responses, and is not able to convincingly identify very long-term responses.”* We see little to disagree with concerning what response is relevant, what the existing empirical literature does and what it does not do.

Our approach for weighing in on the plausible value for the long-run elasticity ϵ_1 is in two parts. First, we follow the quantitative general equilibrium model tradition. An equilibrium model is posed that offers a clear mechanism for how one becomes a top earner. Model parameters are then picked so that the model economy matches relevant statistics of the US economy. The most relevant earnings statistics are the 99th percentile at each age and the Pareto statistic beyond this threshold at each age. Another relevant statistic is the regression coefficient estimated in the literature on the Frisch elasticity of male labor hours. MaCurdy (1982), Altonji (1986) and Domeij and Floden (2006) have all estimated the regression coefficient of the change in log work hours on the change in log wage rates for prime-age males in PSID data. The small value of this regression coefficient has been viewed by some public economists (see Saez et al. (2012, p. 3)) as additional support for the plausibility of a small value for ϵ_1 and, thus, a large value for the revenue maximizing top tax rate. After setting all model parameters, we calculate the model-implied value of ϵ_1 .

Our second approach for weighing in on the plausible value for the long-run elasticity ϵ_1 is to examine whether the short-term model elasticity ϵ is consistent with existing estimates. This can be viewed as an external model validation exercise. If the model produced a counterfactual short-term elasticity, then the long-run elasticity would be suspect. Table 6 shows that the mean value of ϵ that results from applying the regression framework of Gruber and Saez (2002)

to model data lie in the interval $[\.15, .30]$. These are precisely the empirical techniques the literature uses to estimate a short-term elasticity. Thus, the human capital model produces a short-term empirical elasticity consistent with the Diamond-Saez view summarized earlier. Of course, the true long-run model elasticity is $\epsilon_1 = 0.4$ and this value determines the top of the Laffer curve. The long-run model elasticity exceeds the short-term model elasticity because skills of top earner types respond only after a long lag.

Next issue: **other** evidence for the mechanism in the model.

References

- Altig, D. and C. Carlstrom (1999), Marginal Tax Rates and Income Inequality in a Life-Cycle Model, *American Economic Review*, 89, 1197-1215.
- Altonji, J. (1986), Intertemporal Substitution in Labor Supply: Evidence from Micro Data, *Journal of Political Economy*, 94, S176-S215.
- Alvaredo, F., Atkinson, A., Piketty, T. and E. Saez, The World Top Incomes Database, <http://topincomes.g-mond.parisschoolofeconomics.eu/>, 10/10/2013.
- Auten, G, Carroll, R. and G. Gee (2008), The 2001 and 2003 Tax Rate Reductions: An Overview and Estimate of the Taxable Income Response, *National Tax Journal*, 61, 345-64.
- Badel, A. and M. Huggett (2015), The Sufficient Statistic Approach: Predicting 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.
- Feenberg, D. and E. Coutts, An Introduction to the TAXSIM Model, *Journal of Policy Analysis and Management*, 12, 189-94.

- Giertz, S. (2010), The Elasticity of Taxable Income During the 1990's: New Estimates and Sensitivity Analyses, *Southern Economic Journal*, 77, 406-33.
- Goolsbee, A. (2000), What Happens When You Tax the Rich? Evidence from Executive Compensation, *Journal of Political Economy*, 108, 352-78.
- Gruber, J. and E. Saez (2002), The Elasticity of Taxable Income: Evidence and Implications, *Journal of Public Economics*, 84, 1-32.
- 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 (2014), The Nature of Countercyclical Income Risk, *Journal of Political Economy*, 122, 621-60.
- 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.
- Heckman, J, Lochner, L. and C. Taber (1998), Tax Policy and Human Capital Formation, *American Economic Review*, Papers and Proceedings, May 1998.
- 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 (2012), Micro and Macro Labor Supply Elasticities: A Reassessment of Conventional Wisdom, *Journal of Economic Literature*, 50, 464-76.
- Kindermann, F. and D. Krueger (2014), High Marginal Tax Rates on the Top 1% ? Lessons from a Life Cycle Model with Idiosyncratic Income Shocks, manuscript.
- MaCurdy, T (1981), An Empirical Model of Labor Supply in a Life-Cycle Setting, *Journal of Political Economy*, 89, 1059- 85.
- Mertens, K. (2015), Marginal Tax Rates and Income: New Time Series Evidence, manuscript.
- Mirrlees, J. (1971), An Exploration into the Theory of Optimum Income Taxation, *Review of Economic Studies*, 38, 175- 208.
- 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.

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 $t \in \{1978, 1979, \dots, 2011\}$. 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 $y = 1968, 1970, 1971, \dots, 1996$ 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

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.

to $(\beta_{23} + d, \beta_{24} + d, \beta_{25} + d, \dots, \beta_{63} + d)$, where the β_a are the estimated age coefficients and d is a vertical displacement selected in the manner described in section 3.

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 $\{(q_1, R_1), \dots, (q_N, R_N)\}$.

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

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

The values of $\{(q_1, R_1), \dots, (q_N, R_N)\}$ 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 2010.¹⁸ 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 $T(x; \zeta)$ using a 5th order polynomial P , where x denotes the sum of earnings and social security transfers in the model:

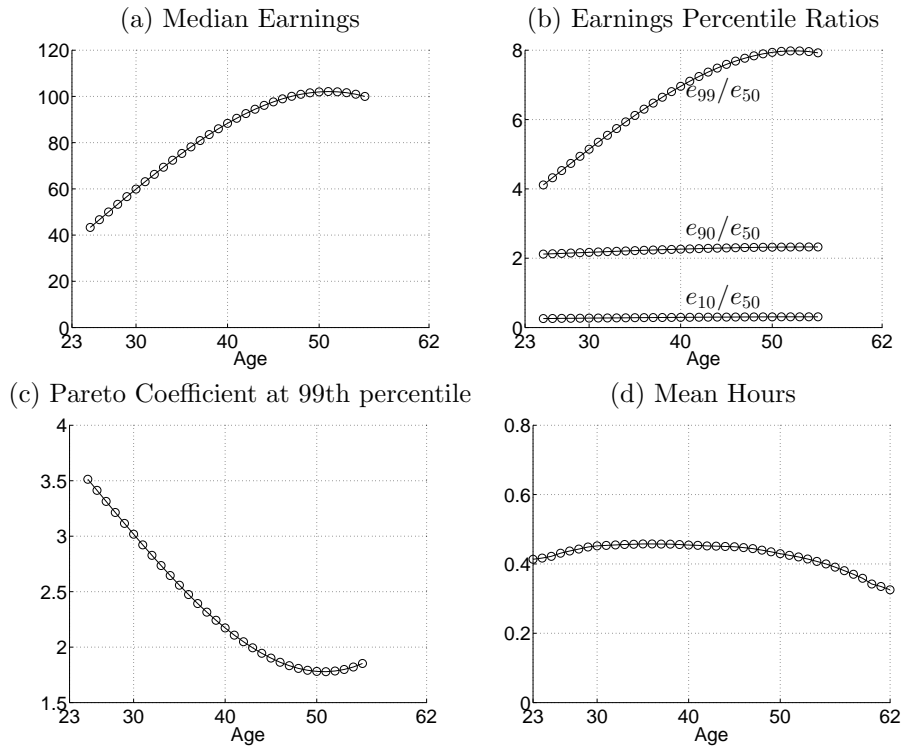
$$T(x; \zeta) = \begin{cases} \kappa P(x/\kappa; \zeta) & x \leq \kappa \\ \kappa P(1; \zeta) + \bar{\tau}[x - \kappa] & x > \kappa \end{cases}$$

¹⁸The 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.

We set the coefficients ζ to minimize the distance $\sum_{x_i \in X^{grid}} (\hat{T}(x_i) - T(x_i; \zeta))^2$ 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} .

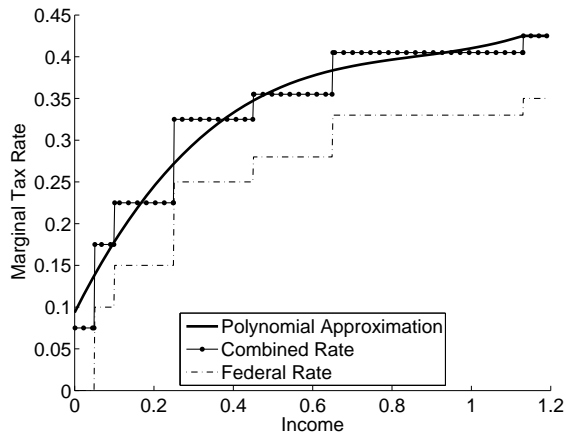
In summary, the model tax function in step 2 approximates the empirical tax function with a polynomial. The polynomial is restricted to produce zero taxes at zero income and to produce a 42.5 percent marginal tax rate at the start of the top tax bracket. Beyond the top tax bracket, the model tax function 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.

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



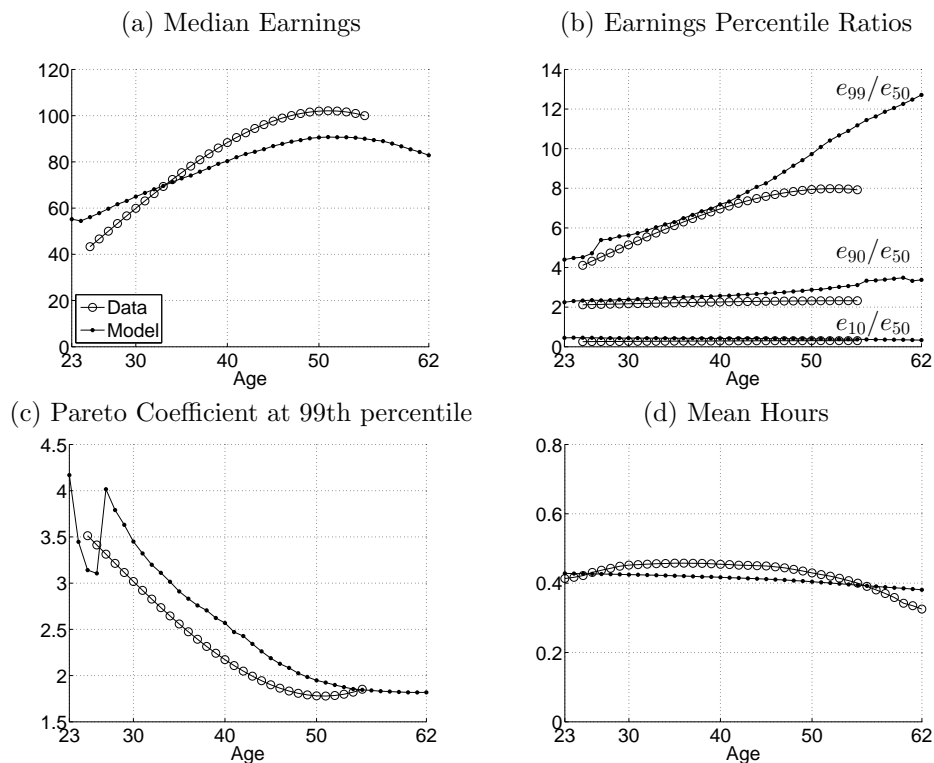
Note: Earnings profiles are based on SSA data. Hours profiles are based on PSID data.

Figure 2: **Model Tax System**



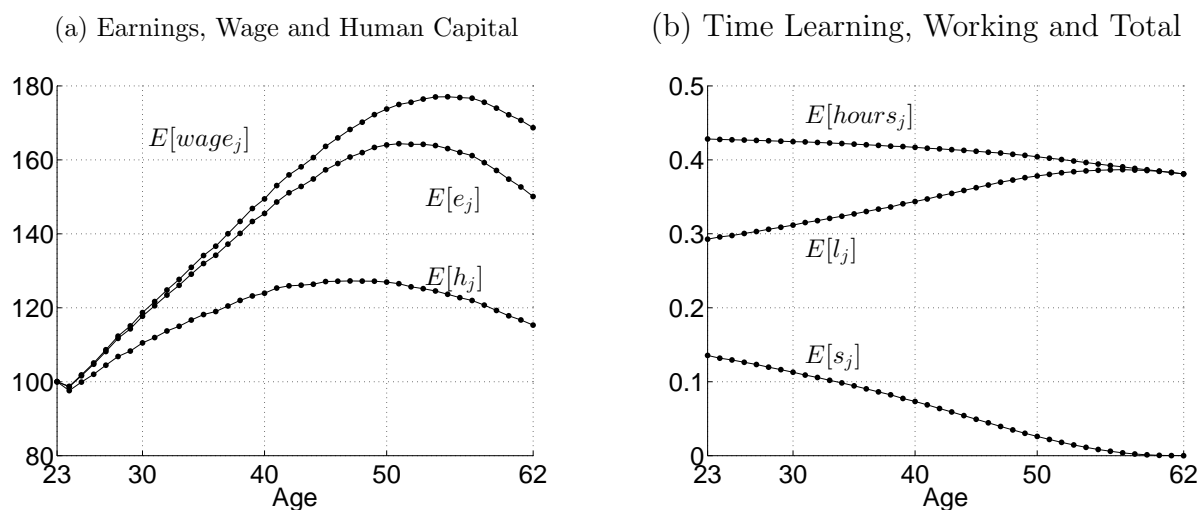
Note: The horizontal axis measures income in multiples of the 99th percentile of income.

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



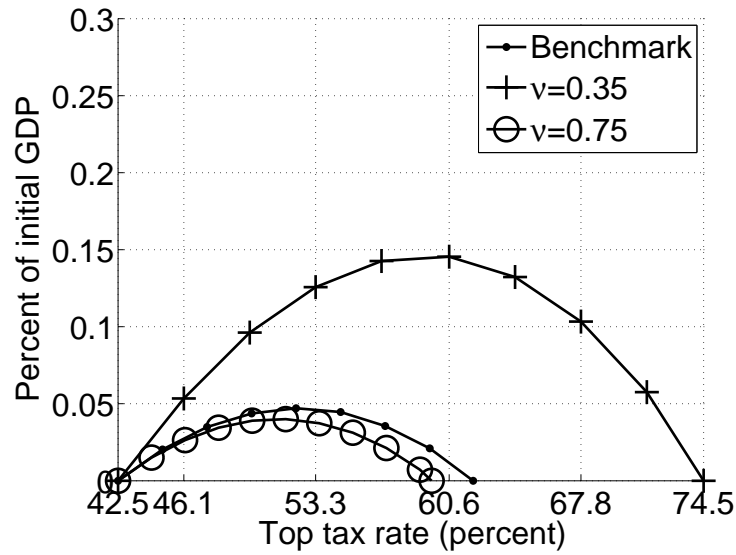
Note: Large open circles describe profiles for the U.S. economy. Small solid circles describe profiles for the model economy.

Figure 4: **Life-Cycle Mean Model Profiles**



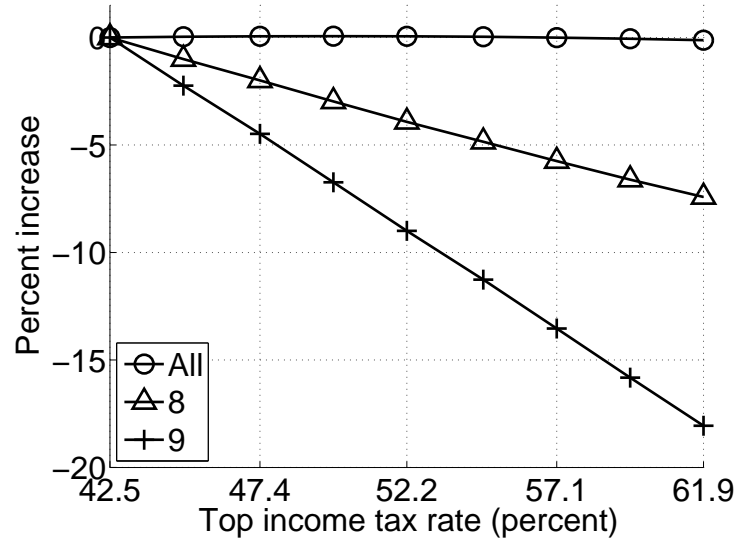
Note: Earnings, wages and human capital in Figure 5(a) are all normalized to equal 100 at age 23.

Figure 5: Laffer Curves



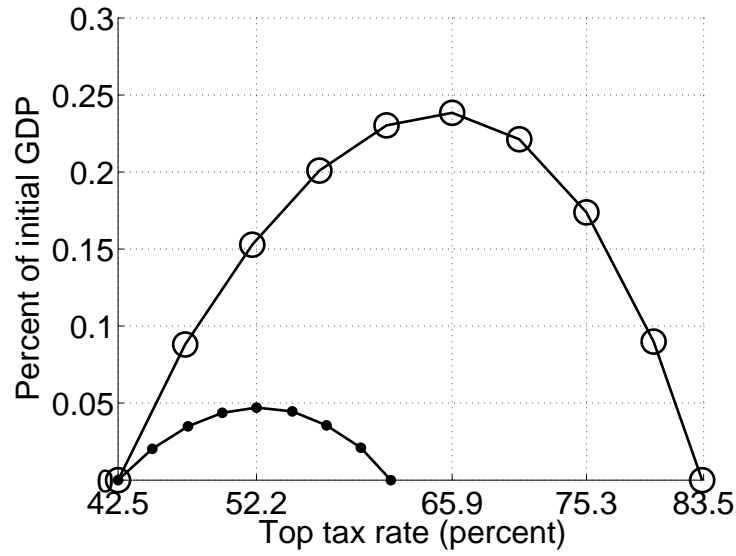
Note: The vertical axis plots the equilibrium lump-sum transfer stated as a percent of initial GDP .

Figure 6: Equivalent Consumption Variation



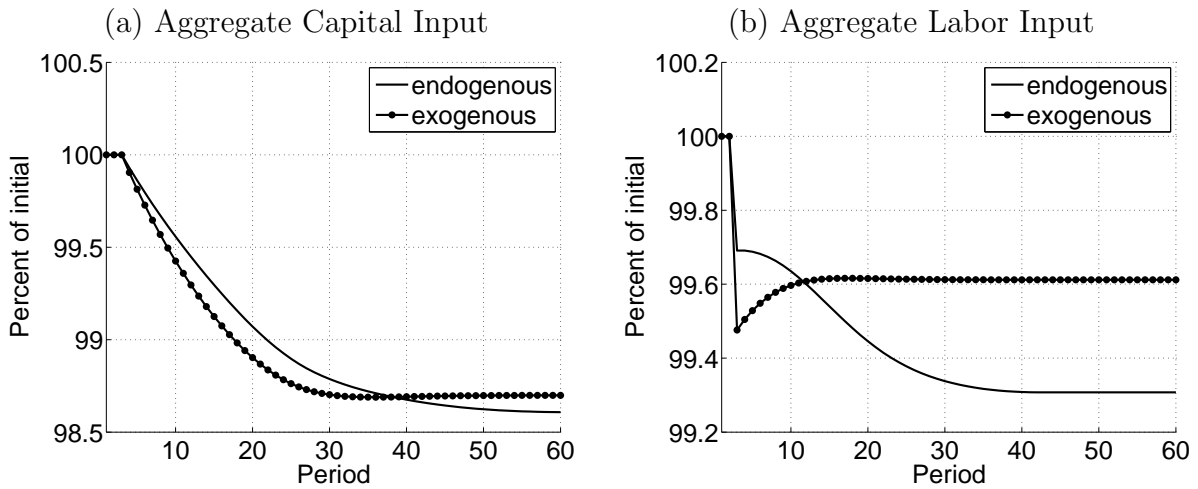
Note: Legend labels 1-9 denote results conditional on learning ability level. Level 1 is the lowest and level 9 is the highest. Level 8 and 9 together make up half of one percent of the population. The legend label ALL denotes unconditional results in that expected utility is calculated averaging across all ability types.

Figure 7: Laffer Curves: Endogenous and Exogenous Human Capital



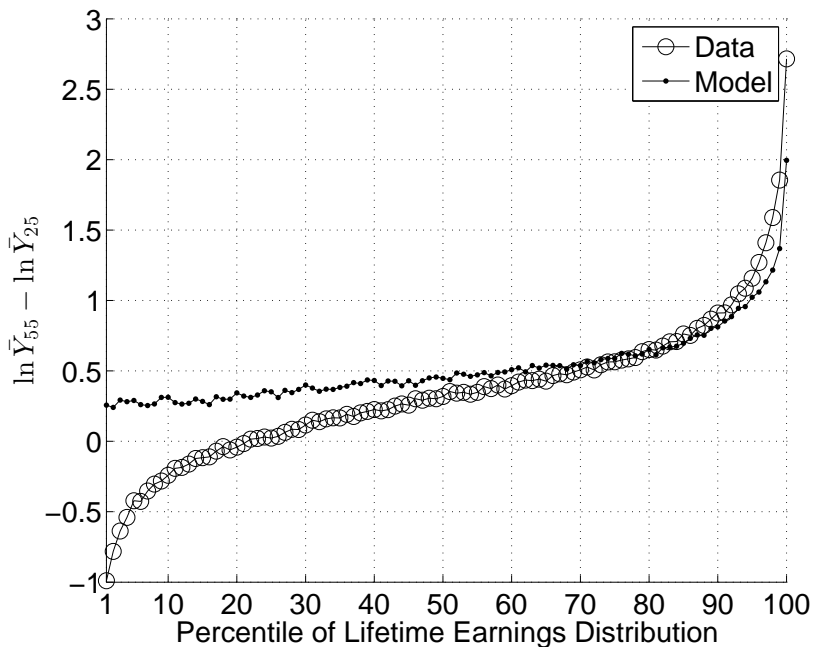
Note: Dots plot properties of the benchmark endogenous human capital model. Open circles plot properties of the exogenous human capital model.

Figure 8: Transition Paths



Note: Aggregate capital and labor input in the endogenous and exogenous human capital models are normalized to equal 100 in period 1. The tax reform occurs in model period 3.

Figure 9: 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.

B Appendix - NOT FOR PUBLICATION

B.1 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) $v_j(x, a) = \max_{(c, l, s, k')} u(c, l + s) + \beta E[v_{j+1}(k', h', a)]$ 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 lying 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 exogenous human capital model is given below. The skills process is by construction exactly the same as in the original benchmark steady-state equilibrium. An equilibrium in this model is defined in the same way as in the benchmark model with the exception that the decision problem differs.

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

1. Given top tax rate $\bar{\tau}$, guess $(K/L, \bar{T})$. 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 .

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

- 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.

B.2 Initial Conditions

We 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, λ_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 $i = 1, 2, 3, \dots, 9$. Set $(P_1, \dots, P_9) = (0.225, \dots, 0.225, 0.06, 0.03, 0.005, 0.004, 0.001)$. Given the probabilities, set learning ability levels to equal conditional means implied by the marginal distribution $F(a)$ implied by $PLN(\mu_a, \sigma_a^2, \lambda_a)$.

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

Second, for any $a \in A^{grid} = \{a_1, \dots, a_9\}$, 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 $j = 1, \dots, 20$ are set following Tauchen (1986). This then implies that X_1^{grid} has 9×20 points and that $\psi(x) = P_i P_j$ for $x = (a_i, h_j) \in X_1^{grid}$.

B.3 MaCurdy Regressions: An Interpretation

We first describe the construction of the data sets underlying the results in Table 3. 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.

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. The last step assumes that the agent is off the corner of the borrowing constraint (i.e. $k_{j+1} > 0$) and that there is no risk. We do so for transparency. 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 (2006)). When there is risk, the last equation is modified by an additive “forecast error” term (see Keane (2011) or Keane and Rogerson (2012)) where the additive term is based on a linear approximation.

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

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

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

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

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. 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 highlighted this point.¹⁹ 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, the regression coefficient is still less than half the value of ν . Domeij and Floden (2006) argue that in exogenous-wage models standard estimation procedures are biased downward. They demonstrate a downward bias due to borrowing constraints and approximation error of the intertemporal Euler equation. When we include in the estimation only agents with substantial assets (more than one quarter of mean assets), then all regression coefficients in Table 3 increase markedly but still remain below the value of ν . Domeij and Floden (2006) find in PSID data that the regression coefficient increases markedly when the sample is restricted to eliminate individuals with low asset holdings.

B.4 Longer-Run Elasticity

The elasticity of taxable income literature acknowledges that comparing measured income a year before and a year after a tax reform may be misleading because the relevant response for policy is the long-run response. One approach to estimate a longer-run elasticity is to use the Gruber and Saez (2002) regression equation but to measure income and net-of-tax rate changes over a longer horizon. For example, Auten, Carroll and Gee (2008) consider three-year and five-year differences, while Giertz (2010) considers one, three and six-year differences. We follow this approach below.

Another approach, used by Goolsbee (2000), argues that some of the measured response may be due to income shifting across years when there is advanced information of a pending reform and provides evidence for such income shifting. As advanced information is not a problem in the model, we will not pursue the modifications of the regression equation suggested by the work of Goolsbee (2000).

Table B1 estimates the regression equation below using $k = 1, 3$ and 6 year horizons. The regression and instrument specification follows those in Giertz (2010 Table 2, row A).

$$\log\left(\frac{z_{it+k}}{z_{it}}\right) = \epsilon \log\left(\frac{1 - \tau_{t+k}(z_{it+k})}{1 - \tau_t(z_{it})}\right) + \beta f(z_{it}) + \alpha_t + X_i' \gamma + \nu_{it+k} \quad (1)$$

The regression equation is a straightforward extension of the regression equation from section 6. The instrument specification is analogous to that used in column (6) of our Table 6. It consists of the counterfactual growth $\log\left(\frac{1 - \tau_{t+k}(z_{it+k}^P)}{1 - \tau_t(z_{it})}\right)$ of the marginal net of tax rate between t and $t+k$. Here z_{it+k}^P equals z_{it} times the growth factor of average earnings in the sample between t and $t+k$ so that earnings are assumed to be constant relative to trend. Columns (4)-(6) of Table B1 report estimations that include a fourth-order polynomial in age in the term X_i and use a weighted estimator where weights are equal to current earnings z_{it} . These practices are advocated by Giertz (2010).

The specification in column (1) of Table B1 resembles the specification used in column (6) of Table 6 in that age controls are not used and earnings weighting is not used. However, the estimates differ because the instrument used in column (1) of Table B1 is based on predicted future earnings z_{it+1}^P while previously earnings z_{it+1} were replaced with z_{it} .

¹⁹This logic for why the regression coefficient is lower than the utility function parameter ν 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 our framework.

Table B1 - Longer-Run Elasticity Estimates

(a) Endogenous Human Capital Model

	(1)	(2)	(3)	(4)	(5)	(6)
Mean Elasticity ϵ	0.1998	0.1114	-0.0250	0.2405	0.1647	-0.1750
S.D.	(0.0223)	(0.0392)	(0.0606)	(0.0290)	(0.0492)	(0.0781)
Difference order (k)	1	3	6	1	3	6
Age Polynomial	no	no	no	yes	yes	yes
z_{it} Weights	no	no	no	yes	yes	yes
Long-run Model Elasticity ϵ_1	0.396	0.396	0.396	0.396	0.396	0.396

(b) Exogenous Human Capital Model

	(1)	(2)	(3)	(4)	(5)	(6)
Mean Elasticity ϵ	0.2053	0.1199	-0.0237	0.2279	0.1571	-0.1706
S.D.	(0.0228)	(0.0320)	(0.0603)	(0.0269)	(0.0430)	(0.0770)
Difference order (k)	1	3	6	1	3	6
Age Polynomial	no	no	no	yes	yes	yes
z_{it} Weights	no	no	no	yes	yes	yes
Long-run Model Elasticity ϵ_1	0.240	0.240	0.240	0.240	0.240	0.240

Notes : (1) All regressions include time effects and income control $f(z_{it}) = \ln(z_{it})$. (2) We draw 100 balanced panel data sets of 30,000 agents following the procedure from Table 6. (3) We report means and standard deviations of the point estimates of ϵ across 100 randomly drawn balanced panels.