Online Appendices

Contents

Appendix A

A.A. *Empirical measurement of progressivity*

Statutory progressivity. It is statutory marginal tax rates that are the relevant tax rates in determining household's labor supply and skill investment choices on the margin. This follows from a standard envelope argument: if an individual is already optimizing on the margin in terms of spending on tax-deductible items or in terms of tax avoidance activities, then the marginal value of working a little harder can be computed holding deductions fixed, i.e., using the statutory marginal rate. We therefore want to estimate the progressivity of statutory tax rates. We begin by proving our claim of Section II. that τ does measure the progressivity of statutory (as opposed to effective) tax rates when using taxable income (i.e., gross income y minus deductible expenses x) in the estimation. Tax liabilities are then given by the generalized formula:

(A1) $$T(y,x) = (y-x) - \lambda(y-x)^{1-\tau},$$

where $y - x$ is taxable income, which implies the log-linear relation between post-government income and pre-government taxable income

(A2) $$\log(y - x - T(y,x)) = \log\lambda + (1-\tau)\log(y-x)$$

that we use to estimate the parameter τ, as discussed in detail in Section II. of the paper. The statutory marginal tax rate implied by (A1) is:

(A3) $$MTR^s \equiv \frac{\partial T(y,x)}{\partial y} = 1 - \lambda(1-\tau)(y-x)^{-\tau},$$

where note that, by definition, x is kept fixed as y varies marginally. Note, instead, that the effective marginal tax rate MTR^e requires computing the total derivative, or:

(A4)
$$MTR^e \equiv \frac{dT(y,x)}{dy} = \frac{\partial T(y,x)}{\partial y} + \frac{\partial T(y,x)}{\partial x}\frac{dx}{dy} = \left[1 - \lambda(1-\tau)(y-x)^{-\tau}\right]\left(1 - \frac{dx}{dy}\right) \leq MTR^s,$$

which is less than or equal to the statutory marginal tax rate MTR^s, as long as x is a normal good. Intuitively, as income grows, so do non-taxable expenditures and thus taxable income rises less than one for one. The average tax rate, expressed as a share of taxable income, is

(A5) $$ATR \equiv \frac{T(y,x)}{y-x} = 1 - \lambda(y-x)^{-\tau}.$$

A natural measure of statutory progressivity is one minus the coefficient of residual income progression: $1 - (1 - MTR^s)/(1 - ATR)$. By combining (A3) and (A5) it is easy to see that this expression is exactly equal to τ, and thus our empirical approach to estimating progressivity captures the progressivity of statutory tax rates, as desired.

Average marginal tax rates. We now show that when assuming a balanced budget, the average income-weighted marginal tax rate is given by eq. (4). To see this, note that budget balance requires $gY = \int y_i - \lambda y_i^{1-\tau} di$. The income-weighted average marginal tax rate is then given by

$$\int \left[1 - \lambda\left(1-\tau\right) y_i^{-\tau}\right] \left(\frac{y_i}{Y}\right) di = 1 - \left(1-\tau\right) \int \lambda y_i^{1-\tau} \left(1/Y\right) di \;=\; 1 - \left(1-\tau\right)\left(1-g\right).$$

Measurement of tax deductions. As discussed in the main text, the source of our data used in Section II. is the sample from the Panel Study of Income Dynamics (PSID) for survey years 2000, 2002, 2004, and 2006 (when the PSID is biannual) described in detail in Heathcote, Perri and Violante (2012). From this data set, we are able to construct, for every household, measures of income and government transfers defined in the main text. Our aim is to obtain an accurate estimate of tax liabilities for household in the sample by running the NBER's TAXSIM program (Feenberg and Coutts, 1993) on every record in the data. Most of the fields required by TAXSIM (marital status, number of children, various types of income and government transfers, see the description of the software at `http://users.nber.org/˜taxsim/taxsim9/`) are are already available in our original sample. What requires more work is constructing, for each household, an estimate of their tax deductions in order to arrive at a reliable measure of taxable income. We begin by calculating the four major sources of itemized deductions in the US tax code: medical expenses, mortgage interest, state taxes paid, and charitable contributions. The PSID contains three comprehensive questions on out-of-pocket medical expenditures for, respectively, (i) nursing home and hospital bills, (ii) doctor, outpatient surgery, and dental bills, and (iii) prescriptions, in-home medical care, special facilities, and other medical services. Overall these three questions cover a very broad range of medical expenses. [46] Median (mean) household expenditures across ages 25-60, the age range of our sample, are \$750 (\$1,970), in line with existing estimates from other sources

46. More precisely for each of survey years 2001-03-05-07, the questions we extracted are: 2001: ER19842, ER19848, ER19854. 2003: ER23279, ER23285, ER23291. 2005: ER27240, ER27246, ER27252. 2007: ER40415, ER40421, ER40447. Since the question asks to report expenditures incurred in the past two years, we divide the household's responses by a factor of two to arrive at annual estimates.

(see, e.g., Jung and Tran, 2013, for evidence based on the Medical Expenditure Panel Survey). We follow the instructions on the NBER TAXSIM website on how to calculate the deductible amount for medical expenses.[47] The PSID contains a question on the residual mortgage debt on the main residence for homeowners.[48] As per the US tax code, we cap this amount at $1,000,000. To calculate interest payments, we multiply that amount by 7 percent, the average 30-year conventional annual mortgage rate over this period (FRED series MORTG). As for state taxes, our data set contains the household state of residence that TAXSIM uses to compute state taxes paid by households in calculating the itemized deduction (we used the Internal Revenue Service Statistics of Income (SOI) state codes, as indicated by the instructions for field 2). Finally, since the PSID has no information on charitable contributions, we use a simple imputation procedure. Based on the SOI data, in 2004 charitable contributions amounted to roughly 3% of income, for taxpayers with income greater than $75,000, where the vast majority of tax returns claim itemized deductions (SOI Table 2.1).[49] List (2011, Table 2) reports, from survey data across all income levels, that over 2/3 of households donate to charity, and those who do contribute about 4 percent of their income. Thus a charitable contribution rate of 3 percent of household income, for all households in the sample, seems an appropriate rule of thumb for our imputation, and this is what we assume.[50] Adding up these four components we obtain a measure of itemized deduction. TAXSIM then calculates whether each household is better off taking an itemized deduction or the standard deduction.[51] After running TAXSIM on the data, 49 percent of households in our sample would have taken the itemized deduction (96 percent of those with income above $100,000). According to SOI Table 2.1, in 2004, around 40 percent of tax returns claimed the itemized deduction (97 percent of those with income above $100,000), which confirms that our imputation procedure, combined with the TAXSIM program, is quite accurate.

Imputation of social security benefits. As explained in the main text, to arrive at a com-

47. See http://users.nber.org/~taxsim/taxsim9/medical_deduction.html

48. For survey years 2001-03-05-07, the question is, respectively: ER17052, ER21051, ER25042, ER36042.

49. The exception is taxpayers with income above $10M, whose charitable contributions exceed 6 percent of income. However, our PSID sample has no observation in that income range.

50. In particular, these survey data show no sizable differences across income levels. For example, List reports that for income levels between $20,000-40,000, the fraction of households who donate is 0.58 with a contribution rate of 5 percent of household income. Thus, even for lower income levels, a 3 percent average contribution rate seems correct.

51. The standard deduction varies by marital status. For example, during 2000-2006, for singles it grows steadily from $4,400 to $5,150, and for married couples from $7,350 to $10,300.

prehensive measure of gross income, we augment income reported in the PSID for each working household member with the employer's share (50%) of the Federal Insurance Contribution Act (FICA) tax, which comprises the Old Age, Survivor and Disability Insurance (OASDI) tax (or Social Security tax), and the Hospital Insurance (HI) tax (Medicare tax). Incidentally, the Congressional Budget Office (CBO) makes this same adjustment when computing marginal tax rates. For the years 2000-2006, the OASDI tax rate was 12.4% (and the employer's share half of that) applicable up to an earnings ceiling which varied by year. For details, see the tables at `https://www.ssa.gov/oact/COLA/cbb.html#Series`). The HI tax rate was 2.9% (again, the employer's share being half) and there was no limit to the earnings subject to this tax. To derive our measure of post-government income, we subtract from gross income the entire FICA tax liability. For consistency, we need to make an imputation for the corresponding gain in social security benefits accruing to the household member because of the additional year of work (as explained in the text we make no analogous adjustment for Medicare benefits because Medicare eligibility is based on age rather than tied to lifetime earnings). We compute the present value Δ_{ij} of the extra social security benefits that individual i will receive by working at age j relative to the counterfactual where i does not work in that year, but where her past and the future earnings are unchanged. We then add Δ_{ij} to government transfers for that individual-year observation.

This calculation is implemented as follows. For every individual in the sample, we compute an age-earnings profile conditional on gender g and education e (less than high-school, high-school degree, and college degree) using a cubic polynomial in age j. Call these functions $h_{ge}(j)$. Since we observe income y_{ij} for every individual i only once at age j, our best estimate of age j^* earnings for this individual is

$$\hat{y}_{ij^*} = \frac{h_{ge}(j^*)}{h_{ge}(j)} y_{ij}.$$

Let $\bar{Y}_i$ be the Average Index of Monthly Earnings (AIME) under the assumption the individual works from age 0 until retirement age $J^{ret} = 35$ (36 years, as in our sample), i.e.,

$$\bar{Y}_i = \frac{1}{12} \cdot \left(\frac{1}{J^{ret}} \sum_{m=0}^{J^{ret}} \hat{y}_{im} \right)$$

and let

$$\bar{Y}_{ij}^{-} = \bar{Y}_i - \frac{1}{12} \cdot \frac{y_{ij}}{J^{ret}}$$

be the counterfactual AIME in the absence of age j earnings. The implied annualized social security benefit gain from working at age j is:

$$\pi_{ij} = 12 \cdot \left[P\left(\bar{Y}_i\right) - P\left(\bar{Y}_{ij}^{-}\right) \right],$$

where P is the formula that determines monthly benefits as a function of AIME, see the explanation at `https://www.ssa.gov/oact/COLA/piaformula.html`.

The marginal gain to be imputed to post-government earnings is the present value of this pension gain discounted back to age j accounting for the fact that this additional pension income is paid in every year following retirement age, conditional on survival:

$$\Delta_{ij} = \left(\frac{1}{R^{J^{ret}-j}} \right) \pi_{ij} \sum_{m=J^{ret}}^{J} \frac{s_{j,m}}{R^{m-J^{ret}}},$$

where $s_{j,m}$ is the probability of surviving from age j to age m (i.e., the probability of collecting at age m), which is gender-specific and computed based on U.S. Life Tables). In this calculation, we assume that $J = 100$, so $s_{j,J} = 0$, and $R = 1.04$ annually.

Appendix B

This Appendix proves all the results in the main body of the paper.

B.1. Proof of Proposition 1 [hours and consumption]

We follow the proof in Heathcote, Storesletten, and Violante (2014), simplified by the absence of risk-free bonds in the economy. Since the only securities that are traded are insurance claims against ε shocks, without loss of generality we can think of our economy as an island economy where each island is populated by agents indexed by their fixed individual characteristic (φ, s) and their uninsurable wage component α. On each island, there are complete markets with respect to ε, so the competitive equilibrium allocation can be computed as the outcome of an island-specific planner problem. Since agents on an island are ex ante identical, the planner weights must be equal across agents. Moreover, since each island transfers zero net financial wealth between periods (by assumption) and preferences are time separable, the island-specific planner problem is static. The island social planner's problem, taking the aggregate fiscal variables (G, λ, τ) and the skill price $p(s)$ as given, is

$$\max_{\{c(\varepsilon),h(\varepsilon)\}} \int_E \left\{ \log c(\varepsilon) - \frac{\exp\left[(1+\sigma)\varphi\right]}{1+\sigma} h(\varepsilon)^{1+\sigma} + \chi \log G \right\} dF_\varepsilon$$

subject to the resource constraint

$$\text{(B1)} \qquad \int_E c(\varepsilon)\, dF_\varepsilon = \lambda \int_E \exp\left[(1-\tau)\left(p(s)+\alpha+\varepsilon\right)\right] h(\varepsilon)^{1-\tau}\, dF_\varepsilon.$$

It is immediate to see that all agents on each island consume the same amount. Exploiting this perfect risk-sharing outcome, and substituting the resource constraint into the intratemporal first-order condition (FOC), one obtains hours worked by individuals with shock ε on each island

$$\text{(B2)} \qquad \log h(\varepsilon) = \frac{1}{(1+\sigma)} \log(1-\tau) - \varphi + \frac{1-\tau}{\sigma+\tau}\varepsilon - \frac{1}{\sigma+\tau}\left[\frac{(1-\tau)(1-2\tau-\sigma\tau)}{\sigma+\tau}\frac{v_\varepsilon}{2}\right],$$

where the first term is hours worked by the representative agent, and the term in the square bracket is the constant $\mathcal{M}(v_\varepsilon;\tau)$. Substituting (B2) into (B1) one obtains

$$\text{(B3)} \qquad \log c = \log\lambda + \frac{1-\tau}{1+\sigma}\log(1-\tau) + (1-\tau)\left[\log p(s) + \alpha - \varphi\right] + \mathcal{M}(v_\varepsilon;\tau).$$

In the proof of Proposition 4, we solve for λ as a function of (τ, g) and other structural parameters using the government budget constraint. Note that in eq. 21,

$$\vartheta(\tau) = \log \lambda - \left(\log(1-g) + \frac{\tau}{1+\sigma} \log(1-\tau) \right).$$

B.2. Proof of Proposition 2 [skill price and skill choice]

The education cost is given by $v(s) = \frac{\kappa^{-1/\psi}}{1+1/\psi}(s)^{1+1/\psi}$, where κ is exponentially distributed, $\kappa \sim \eta \exp(-\eta\kappa)$. Recall from eq. (14) in the main text that the optimality condition for skill investment is

$$\text{(B4)} \qquad v'(s) = \left(\frac{s}{\kappa}\right)^{\frac{1}{\psi}} = (1-\beta\delta)\,\mathbb{E}_0 \sum_{a=0}^{\infty} (\beta\delta)^a \frac{\partial u\left(c\left(\varphi, \alpha, s; g, \tau\right), h\left(\varphi, \varepsilon; \tau\right), g\right)}{\partial s}.$$

The skill level s affects only the consumption allocation (not the hours allocation) and only through the price $p(s;\tau)$, which is fixed over time. Hence, using (B3), (B4) can be simplified as

$$\left(\frac{s}{\kappa}\right)^{\frac{1}{\psi}} = (1-\tau)\frac{\partial \log p(s;\tau)}{\partial s}.$$

We now guess that the log-price function has the form

$$\text{(B5)} \qquad \log p(s;\tau) = \pi_0(\tau) + \pi_1(\tau) \cdot s,$$

which implies that the skill allocation has the form

$$\text{(B6)} \qquad s(\kappa;\tau) = \left[(1-\tau)\,\pi_1(\tau)\right]^{\psi} \cdot \kappa.$$

Since the exponential distribution is closed under scaling, skills inherit the exponential density shape from κ, with parameter $\zeta \equiv \eta\left[(1-\tau)\,\pi_1(\tau)\right]^{-\psi}$, and its density is $m(s) = \zeta \exp(-\zeta s)$. We now turn to the production side of the economy. Effective hours worked N are independent of skill type s (see Proposition 1). Aggregate output is therefore

$$Y = \left\{ \int_0^{\infty} \left[N \cdot m(s)\right]^{\frac{\theta-1}{\theta}} ds \right\}^{\frac{\theta}{\theta-1}}.$$

The (log of the) hourly skill price $p(s)$ is the (log of the) marginal product of an extra effective hour supplied by a worker with skill s, or

$$\begin{aligned} \text{(B7)} \qquad \log p(s) &= \log\left[\frac{\partial Y}{\partial\left[N \cdot m(s)\right]}\right] = \frac{1}{\theta}\log Y - \frac{1}{\theta}\log\left[N \cdot m(s)\right] \\ &= \frac{1}{\theta}\log\left(\frac{Y}{N}\right) - \frac{1}{\theta}\log\zeta + \frac{\zeta}{\theta}s. \end{aligned}$$

Equating coefficients across equations (B5) and (B7) implies $\pi_1(\tau) = \frac{\zeta}{\theta} = \frac{1}{\theta}\frac{\eta}{[(1-\tau)\pi_1(\tau)]^{\psi}}$, which yields

$$\pi_1(\tau) = \left(\frac{\eta}{\theta}\right)^{\frac{1}{1+\psi}} (1-\tau)^{-\frac{\psi}{1+\psi}} \tag{B8}$$

and thus

$$m(s) = (\eta)^{\frac{1}{1+\psi}} \left(\frac{\theta}{1-\tau}\right)^{\frac{\psi}{1+\psi}} \exp\left(-(\eta)^{\frac{1}{1+\psi}} \left(\frac{\theta}{1-\tau}\right)^{\frac{\psi}{1+\psi}} s\right). \tag{B9}$$

Similarly, the base skill price is

$$\pi_0(\tau) = \frac{1}{\theta}\log\left(\frac{Y}{N}\right) - \frac{\log\left(\frac{\eta}{\theta}\right)}{\theta(1+\psi)} + \frac{\psi}{\theta(1+\psi)}\log(1-\tau). \tag{B10}$$

We derive a fully structural expression for $\pi_0(\tau)$ below in the proof of Corollary 2.2 when we solve for Y and N explicitly.

B.3. Proof of Corollary 2.1 [distribution of skill prices]

The log of the skill premium for an agent with ability κ is

$$\pi_1(\tau) \cdot s(\kappa;\tau) = \pi_1(\tau) \cdot [(1-\tau)\,\pi_1(\tau)]^{\psi} \cdot \kappa = \frac{\eta}{\theta} \cdot \kappa$$

where the first equality uses (B6), and the second equality follows from (B8). Thus, log skill premia are exponentially distributed with parameter θ. The variance of log skill prices is

$$var\left(\log p(s;\tau)\right) = var\left(\pi_0(\tau) + \pi_1(\tau)\cdot s(\kappa;\tau)\right) = \left(\frac{\eta}{\theta}\right)^2 var(\kappa) = \frac{1}{\theta^2}.$$

Since log skill premia are exponentially distributed, the distribution of skill prices in levels is Pareto. The scale (lower bound) parameter is $\exp(\pi_0(\tau))$ and the Pareto parameter is θ.

B.4. Proof of Corollary 2.2 [aggregate quantities]

Aggregate hours and aggregate effective hours are given, respectively, by

$$\begin{aligned} H(\tau) &= \int\int h(\varphi,\varepsilon;\tau)\, dF_{\varphi}dF_{\varepsilon}, \\ N(s;\tau) &= N(\tau) = \int\int\int \exp(\alpha+\varepsilon)h(\varphi,\varepsilon;\tau)\, dF_{\varphi}dF_{\alpha}dF_{\varepsilon}. \end{aligned}$$

Using the expression for individual hours in Proposition 1 and integrating over the normal distributions for φ, α, and ε gives

$$
\begin{aligned}
H(\tau) &= (1-\tau)^{\frac{1}{1+\sigma}} \exp\left(\frac{\tau\,(1+\hat{\sigma})-\hat{\sigma}}{\hat{\sigma}^2}\frac{v_\varepsilon}{2}\right) \\
N(\tau) &= (1-\tau)^{\frac{1}{1+\sigma}} \exp\left(\frac{\tau\,(1+\hat{\sigma})+\hat{\sigma}}{\hat{\sigma}^2}\frac{v_\varepsilon}{2}\right) = H\left(\tau\right)\cdot \exp\left(\frac{1}{\hat{\sigma}}v_\varepsilon\right).
\end{aligned}
$$

Aggregate output is equal to aggregate labor earnings

$$
\begin{aligned}
Y\left(\tau\right) &= \int\int\int\int p(s;\tau)\exp(\alpha+\varepsilon)h(\varphi,\varepsilon;\tau)\,dF_s dF_\varphi dF_\alpha dF_\varepsilon \\
&= \int p\left(s;\tau\right)dF_s \cdot N\left(\tau\right) \\
&= \frac{\theta}{\theta-1}\cdot \exp\left(\pi_0\left(\tau\right)\right)\cdot N\left(\tau\right),
\end{aligned} \tag{B11}
$$

where the last line follows from the fact that skill prices are Pareto distributed with scale $\exp(\pi_0(\tau))$ and Pareto parameter θ. Aggregate labor productivity is

$$
\frac{Y(\tau)}{H(\tau)} = \frac{Y(\tau)}{N(\tau)}\cdot\frac{N(\tau)}{H(\tau)} = \mathbb{E}\left[p\left(s;\tau\right)\right]\cdot \exp\left(\frac{1}{\hat{\sigma}}v_\varepsilon\right).
$$

Finally, one can solve for the base log skill price $\pi_0\left(\tau\right)$. From the production function, eq. (5) in the main text, we have that

$$
\begin{aligned}
Y &= \left\{\int_0^\infty \left[N(\tau)\cdot\zeta\exp\left(-\zeta s\right)\right]^{\frac{\theta-1}{\theta}}ds\right\}^{\frac{\theta}{\theta-1}} \\
&= N(\tau)\cdot\left(\frac{\theta}{\theta-1}\right)^{\frac{\theta}{\theta-1}}\left(\eta\right)^{-\frac{1}{(\theta-1)(1+\psi)}}\left(\frac{1-\tau}{\theta}\right)^{\frac{\psi}{(\theta-1)(1+\psi)}}.
\end{aligned} \tag{B12}
$$

Comparing this equation to eq. (B11) it is immediate that

$$
\pi_0(\tau) = \frac{1}{(\theta-1)(1+\psi)}\left[\psi\log\left(\frac{1-\tau}{\theta}\right)-\log\left(\eta\right)\right]+\frac{1}{\theta-1}\log\left(\frac{\theta}{\theta-1}\right),
$$

which is the expression reported in Proposition 2.

B.5. Proof of Proposition 3 [efficiency with $\chi = v_\omega = \tau = 0$]

With $\chi = 0$ there is no desire for the publicly provided good, and thus the absence of a private market for this good is irrelevant. With $v_\omega = 0$ the absence of private markets for insuring shocks to α is similarly irrelevant: such shocks are simply assumed away. Recall that there are competitive markets for consumption, for the labor supply of each skill type, and competitive insurance markets for shocks to ε. Thus, given $\chi = 0$ and $v_\omega = 0$ and absent government intervention (i.e., with $\tau = 0$ and $\lambda = 1$), the first welfare theorem applies and competitive equilibrium allocations are Pareto efficient and correspond to the solution to a planner's problem. We now derive the Pareto weights such that the solution to the planner's problem corresponds to the competitive equilibrium allocations. Here, we take as given a result that we'll formally prove in Corollary 4.6: when the weights the planner puts on future generations equals the agent's discount factor β, then social welfare is equal to average period utility in cross section. Moreover, absent uninsurable life-cycle shocks, average period utility is independent of age. Thus, the planner chooses allocations $c(\varphi,\kappa,\varepsilon)$, $h(\varphi,\kappa,\varepsilon)$, and $s(\varphi,\kappa,\varepsilon)$ to solve

$$\max \int\int\int \zeta(\varphi,\kappa,\varepsilon)\left\{\log c(\varphi,\kappa,\varepsilon) - \frac{\exp\left[(1+\sigma)\varphi\right]}{1+\sigma} h(\varphi,\kappa,\varepsilon)^{1+\sigma} - \frac{(\kappa)^{-1/\psi}}{1+1/\psi}\left(s(\varphi,\kappa,\varepsilon)\right)^{1+1/\psi}\right\} dF_\kappa dF_\varphi dF_\varepsilon$$

subject to

$$\int\int\int c(\varphi,\kappa,\varepsilon) F_\kappa dF_\varphi dF_\varepsilon = \left[\int_0^\infty \mathbb{E}(z)^{\frac{\theta-1}{\theta}} dz\right]^{\frac{\theta}{\theta-1}},$$

where effective hours by skill type z is given by

$$\mathbb{E}(z) = \int\int\int h(\varphi,\kappa,\varepsilon)\exp(\varepsilon) I_{\{s(\varphi,\kappa,\varepsilon)=z\}} dF_\kappa dF_\varphi dF_\varepsilon.$$

The first-order condition with respect to consumption is

$$\frac{\zeta(\varphi,\kappa,\varepsilon)}{c(\varphi,\kappa,\varepsilon)} = \mu,$$

where μ is the multiplier on the resource constraint. The competitive equilibrium consumption allocation is given by

$$\begin{aligned}
\log c(\varphi,\kappa,\varepsilon) &= \log\lambda + \pi_0(\tau=0) + \pi_1(\tau=0)\cdot s(\kappa;\tau=0) - \varphi + \frac{1}{\sigma}\frac{v_\varepsilon}{2} \\
&= \log\lambda + \pi_0(\tau=0) + \frac{\eta}{\theta}\kappa - \varphi + \frac{1}{\sigma}\frac{v_\varepsilon}{2}.
\end{aligned}$$

It follows immediately that the Pareto weights must take the form

$$\log\zeta(\varphi,\kappa) = \frac{\eta}{\theta}\kappa - \varphi + \varpi,$$

where ϖ is a constant. Now the average Pareto weight must equal one:

$$\int\int \zeta(\varphi,\kappa) dF_\varphi dF_\kappa = 1$$

Thus,

$$\begin{aligned}\int\int \exp\left(\frac{\eta}{\theta}\kappa-\varphi+\varpi\right)dF_\varphi dF_\kappa &= \exp(\varpi)\int\exp\left(\frac{\eta}{\theta}\kappa\right)dF_\kappa\\ &= \exp(\varpi)\frac{\theta}{(\theta-1)}\\ &= 1,\end{aligned}$$

which implies $\exp(\varpi)=\frac{\theta-1}{\theta}$ and thus

$$\log\zeta(\varphi,\kappa)=\frac{\eta}{\theta}\kappa-\varphi-\log\frac{\theta}{\theta-1}.$$

Therefore, we have shown that given the candidate Pareto weights, the planner's consumption allocation aligns with the competitive equilibrium allocation. We now verify that given the same Pareto weights, the equilibrium allocation for skill investment corresponds to the skill investment rule preferred by the planner. To simplify the analysis, we abstract from flexible labor supply and preference heterogeneity, so that agents are heterogeneous only with respect to κ, and the planner's skill investment rule must take the form $s(\kappa)$. Thus,

$$Y=N\cdot\left[\int f_s(z)^{\frac{\theta-1}{\theta}}dz\right]^{\frac{\theta}{\theta-1}}$$

where N is effective hours worked per capita and $f_s(z)$ is the density of skill type z. Let F_s denote the unknown CDF for skills. We know that

$$F_s(s(\kappa))=F_\kappa(\kappa).$$

By the chain rule

$$f_s(s(\kappa))s'(\kappa)=f_\kappa(\kappa).$$

So

$$\begin{aligned}Y &= N\cdot\left[\int_0^\infty f_s(z)^{\frac{\theta-1}{\theta}}dz\right]^{\frac{\theta}{\theta-1}} = N\cdot\left[\int_0^\infty\left(\frac{f_\kappa(s^{-1}(z))}{s'(s^{-1}(z))}\right)^{\frac{\theta-1}{\theta}}dz\right]^{\frac{\theta}{\theta-1}}\\ &= N\cdot\left[\int_0^\infty\left(\frac{f_\kappa(\kappa)}{s'(\kappa)}\right)^{\frac{\theta-1}{\theta}}s'(\kappa)d\kappa\right]^{\frac{\theta}{\theta-1}},\end{aligned}$$

where the substitutions in the last line use $\kappa=s^{-1}(z)$ and $s'(\kappa)d\kappa=dz$ and also exploit the fact that the limits of integration do not change because $s(0)=0$ and $s(\infty)=\infty$. Thus, the planner's problem can be formulated in Lagrangian form as follows:

$$\begin{aligned}\max_{\{c(\kappa),s(\kappa),s'(\kappa),\lambda,\xi(\kappa)\}}\Lambda &= \int\zeta(\kappa)\left\{\log c(\kappa)-\frac{(\kappa)^{-1/\psi}}{1+1/\psi}s^{1+1/\psi}\right\}dF_\kappa\\ &+\lambda\left\{N\cdot\left[\int_0^\infty f_\kappa(\kappa)^{\frac{\theta-1}{\theta}}s'(\kappa)^{\frac{1}{\theta}}d\kappa\right]^{\frac{\theta}{\theta-1}}-\int c(\kappa)dF_\kappa\right\}\\ &+\int_0^\infty\xi(\kappa)\left[s(\kappa)-\left(\int_0^\kappa s'(x)dx+s(0)\right)\right]d\kappa,\end{aligned}$$

where the first line is the objective, the second is the resource constraint, and the third is a set of constraints linking skill investment levels and derivatives. We know that $s(0)=0$. The first-order conditions for $s(\kappa)$ and $s'(\kappa)$ are

$$-\left(\frac{s(\kappa)}{\kappa}\right)^{\frac{1}{\psi}}\zeta(\kappa)f(\kappa)+\xi(\kappa)=0$$

and

$$\lambda N\frac{\theta}{\theta-1}\left[\int_0^\infty f_\kappa(\kappa)^{\frac{\theta-1}{\theta}}s'(\kappa)^{\frac{1}{\theta}}d\kappa\right]^{\frac{1}{\theta-1}}f(\kappa)^{\frac{\theta-1}{\theta}}\frac{1}{\theta}s'(\kappa)^{\frac{1-\theta}{\theta}}-\int_\kappa^\infty\xi(x)dx=0.$$

The last first-order condition can be rewritten as

$$\lambda N^{1-\frac{1}{\theta}}\frac{\theta}{\theta-1}Y^{\frac{1}{\theta}}f_\kappa(\kappa)^{\frac{\theta-1}{\theta}}\frac{1}{\theta}s'(\kappa)^{\frac{1-\theta}{\theta}}=\int_\kappa^\infty\xi(x)dx.$$

Combining the two first-order conditions yields

$$\text{(B13)}\qquad \frac{1}{\theta-1}\left(\frac{Y}{N}\frac{s'(\kappa)}{f_\kappa(\kappa)}\right)^{\frac{1-\theta}{\theta}}=\int_\kappa^\infty\left(\frac{s(\kappa)}{\kappa}\right)^{\frac{1}{\psi}}\zeta(x)f_\kappa(x)dx.$$

Now the planner weights and competitive equilibrium skill investment rule with $\tau=0$ are

$$\log\zeta(\kappa)=\frac{\eta}{\theta}\kappa-\log\left(\theta/(\theta-1)\right)$$

$$s(\kappa)=\left(\frac{\eta}{\theta}\right)^{\frac{\psi}{1+\psi}}\cdot\kappa.$$

Substituting these into the first-order condition (B13) gives

$$\frac{1}{\theta-1}\left(\frac{Y}{N}\frac{\left(\frac{\eta}{\theta}\right)^{\frac{\psi}{1+\psi}}}{f_\kappa(\kappa)}\right)^{\frac{1-\theta}{\theta}}=\int_\kappa^\infty\left\{\left(\frac{\left(\frac{\eta}{\theta}\right)^{\frac{\psi}{1+\psi}}x}{x}\right)^{\frac{1}{\psi}}\exp\left(\frac{\eta}{\theta}x-\log\left(\frac{\theta}{\theta-1}\right)\right)\eta\exp\left(-\eta x\right)\right\}dx,$$

which simplifies to

$$\frac{Y}{N}=\theta\left(\theta-1\right)^{\frac{\theta}{1-\theta}}\left(\frac{\eta}{\theta}\right)^{-\frac{1}{(\psi+1)(\theta-1)}},$$

which is exactly productivity per efficiency unit of labor supplied in the competitive equilibrium. This can be easily verified by substituting eq. (24) for π_0 into eq. (28) for aggregate output. Thus the planner's first-order condition is satisfied at the equilibrium allocation.

B.6. *Proof of Proposition 4 [closed-form social welfare]*

We prove this proposition in two steps. First, we show how to derive a closed-form solution for the residual fiscal instrument λ. Second, we substitute the allocations into the social welfare function and show how to obtain eq. (30).

Step 1. If we let $\tilde{Y} = \int y_i^{1-\tau} di$, we have

$$\lambda = \frac{(1-g)\,Y}{\tilde{Y}}. \tag{B14}$$

To compute $\tilde{Y}$, it is useful to aggregate by age group. Let $\tilde{Y}^a$ denote average per capita disposable income for agents of age a:

$$\begin{aligned}
\tilde{Y}^a &= \int \left[y\left(s,\varphi,\varepsilon,\alpha\right)\right]^{1-\tau} m\left(s\right) ds dF_\alpha^a dF_\varphi dF_\varepsilon \\
&= \int \left[h\left(\varepsilon\right)\exp\left(p\left(s\right)+\alpha_a+\varepsilon\right)\right]^{1-\tau} m\left(s\right) ds dF_\alpha^a dF_\varphi dF_\varepsilon.
\end{aligned}$$

Substituting in the hours allocation (20) the expression for the skill price (22), the density function $m\left(s\right)$ in eq. (B9), and integrating, we arrive at

$$\tilde{Y}^a = \mathcal{K}\times\exp\left(-\tau\left(1-\tau\right)\frac{v_\alpha^a}{2}\right),$$

where

$$\begin{aligned}
\mathcal{K} &= (1-\tau)^{\frac{1-\tau}{1+\sigma}}\exp\left(-\frac{1}{\widehat{\sigma}}\mathcal{M}\right)\exp\left(\left(\frac{\left(1-\tau\right)\left(1+\widehat{\sigma}\right)}{\widehat{\sigma}}\left(\frac{\left(1-\tau\right)\left(1+\widehat{\sigma}\right)}{\widehat{\sigma}}-1\right)\right)\frac{v_\varepsilon}{2}\right) \\
&\quad\times\exp\left(-\tau(1-\tau)\frac{v_\varphi}{2}\right)\exp\left[(1-\tau)\pi_0(\tau)\right]\int\exp\left[\left(1-\tau\right)\pi_1(\tau)s\right]m\left(s\right)ds.
\end{aligned}$$

Note that

$$\begin{aligned}
&\int_0^\infty \exp\left[\left(1-\tau\right)\pi_1(\tau)s\right]m\left(s\right)ds \\
&= \frac{\theta}{\theta-1+\tau}\int_0^\infty\frac{\theta-1+\tau}{\theta}\zeta\exp\left(-\frac{\theta-1+\tau}{\theta}\zeta s\right)ds \\
&= \frac{\theta}{\theta-1+\tau},
\end{aligned}$$

and recall that

$$\begin{aligned}
\pi_0(\tau) &= \frac{1}{\theta-1}\left\{\frac{1}{1+\psi}\left[\psi\log(1-\tau)-\log\eta-\psi\log\theta\right]+\log\left(\frac{\theta}{\theta-1}\right)\right\} \\
\mathcal{M} &= \frac{(1-\tau)\left[1-\tau(1+\widehat{\sigma})\right]}{\widehat{\sigma}}\frac{v_\varepsilon}{2}.
\end{aligned}$$

Now sum across age groups to obtain

$$\begin{aligned}
\tilde{Y} &= (1-\delta)\sum_{a=0}^{\infty}\delta^a\tilde{Y}^a=\mathcal{K}\times(1-\delta)\sum_{a=0}^{\infty}\delta^a\exp\left(-\tau(1-\tau)\frac{v_\alpha^a}{2}\right) \\
&= \mathcal{K}\times\frac{(1-\delta)\exp\left(-\tau(1-\tau)\frac{v_\alpha^0}{2}\right)}{1-\delta\exp\left(\frac{-\tau(1-\tau)}{2}v_\omega\right)}. \qquad \text{(B15)}
\end{aligned}$$

Substituting $(B12)$ and $(B15)$ into $(B14)$ and simplifying, we arrive at a solution for the equilibrium value of λ which, in logs, is

$$\begin{aligned}
\log\lambda &= \log(1-g)+\frac{\tau(1-\tau)}{\sigma+\tau}\left(\frac{1+\sigma}{\sigma+\tau}+2+\sigma\right)\frac{v_\varepsilon}{2}+\frac{\tau}{1+\sigma}\log(1-\tau) \\
&\quad+\tau(1-\tau)\frac{v_\varphi}{2} \\
&\quad-\log(1-\delta)+\tau(1-\tau)\frac{v_\alpha^0}{2}+\log\left[1-\delta\exp\left(\frac{-\tau(1-\tau)}{2}v_\omega\right)\right] \\
&\quad+\left(\frac{\psi}{1+\psi}\right)\frac{\tau}{\theta-1}\log\left(\frac{1-\tau}{\theta}\right)-\left(\frac{1}{1+\psi}\right)\frac{\tau}{\theta-1}\log\eta \\
&\quad+\left(\frac{\theta+\tau-1}{\theta-1}\right)\log\left(\frac{\theta}{\theta-1}\right)+\log\left(\frac{\theta-1+\tau}{\theta}\right). \qquad \text{(B16)}
\end{aligned}$$

Step 2. Substituting the equilibrium allocations into period utility at age $a\geq 0$, we have

$$\begin{aligned}
u(c_a,h,G) &= \log\lambda+\frac{1-\tau}{1+\sigma}\log(1-\tau)-(1-\tau)\varphi \\
&\quad+\frac{1-\tau}{(\theta-1)(1+\psi)}\left[\psi\log(1-\tau)+\log\left(\frac{1}{\eta}\frac{\theta}{(\theta-1)^{(1+\psi)}}\right)\right]+\mathcal{M} \qquad \text{(B17)} \\
&\quad-\exp\left(-\frac{1+\sigma}{\widehat{\sigma}(1-\tau)}\mathcal{M}\right)\exp\left(\frac{1+\sigma}{\widehat{\sigma}}\varepsilon\right)\left(\frac{1-\tau}{1+\sigma}\right)+(1-\tau)\kappa\frac{\eta}{\theta}+\chi\log G+(1-\tau)\alpha_a
\end{aligned}$$

The disutility cost from investing in education is

$$v(s(\kappa))=-\frac{\kappa^{-1/\psi}}{1+1/\psi}\cdot s(\kappa;\tau)^{1+1/\psi}=-(1-\tau)\kappa\frac{\eta}{(1+1/\psi)\theta}. \qquad \text{(B18)}$$

Average cross-sectional utility (excluding skill investment costs) at age a, which we denote by $\bar{u}_a$, is

$$\begin{aligned}
\bar{u}_a &= \int\int\int\int u(c_a, h, G) dF_\kappa dF_\varepsilon dF_\varphi dF_\alpha^a \\
&= \bar{u} - (1-\tau)\frac{v_\alpha^a}{2} \\
&= \bar{u} - (1-\tau)\frac{a v_\omega}{2},
\end{aligned}$$

where

$$\begin{aligned}
\bar{u} &= \log\lambda + \frac{1-\tau}{1+\sigma}\log(1-\tau) - (1-\tau)\frac{v_\varphi}{2} + \frac{1-\tau}{(\theta-1)(\psi+1)}\left[\psi\log(1-\tau) + \log\left(\frac{1}{\eta}\frac{\theta}{(\theta-1)^{(1+\psi)}}\right)\right] \\
&\quad \mathcal{M} - \frac{1-\tau}{1+\sigma} + \frac{1-\tau}{\theta} + \chi\log(gY) \\
&\quad \text{(B19)}
\end{aligned}$$

and where the derivation of the expression for $\bar{u}$ exploits the facts that

$$\begin{aligned}
\int \exp\left(-\frac{1+\sigma}{\widehat{\sigma}(1-\tau)}\mathcal{M}\right)\exp\left(\frac{1+\sigma}{\widehat{\sigma}}\varepsilon\right)dF_\varepsilon &= 1, \\
\int (1-\tau)\kappa\frac{\eta}{\theta}dF_\kappa &= \frac{1-\tau}{\theta}.
\end{aligned}$$

Substituting in eq. (B19) expression (B16) for λ and (B12) for Y gives

$$\begin{aligned}
\text{(B20)}\,\bar{u} &= \log(1-g) + \chi\log g + (1+\chi)\frac{\log(1-\tau)}{(1+\hat{\sigma})(1-\tau)} - \frac{1}{(1+\hat{\sigma})} \\
&\quad + (1+\chi)\left[\frac{1}{\theta-1}\frac{\psi}{1+\psi}\log(1-\tau) + \frac{1}{\theta-1}\frac{1}{1+\psi}\log\left(\frac{1}{\eta\theta^\psi}\left(\frac{\theta}{\theta-1}\right)^{\theta(1+\psi)}\right)\right] \\
&\quad - \left[-\log\left(1-\left(\frac{1-\tau}{\theta}\right)\right) - \left(\frac{1-\tau}{\theta}\right)\right] - (1-\tau)^2\frac{v_\varphi}{2} \\
&\quad + \log\left(\frac{1-\delta\exp\left(\frac{-\tau(1-\tau)}{2}v_\omega\right)}{1-\delta}\right) + (1+\chi)\left[\frac{1}{\hat{\sigma}}v_\varepsilon - \sigma\frac{1}{\hat{\sigma}^2}\frac{v_\varepsilon}{2}\right].
\end{aligned}$$

Average skill investment costs for agents born after the tax reform are

$$\bar{v}_Y = \mathbb{E}\left[-(1-\tau)\kappa\frac{\eta}{(1+1/\psi)\theta}\right] = -\left(\frac{\psi}{1+\psi}\right)\frac{1}{\theta}(1-\tau),$$

whereas average net costs for those born prior to the reform are

$$\bar{v}_O = \frac{\psi}{1+\psi}\frac{1-\tau}{\theta} - \frac{\psi}{1+\psi}\frac{1-\tau_{-1}}{\theta} = -\left(\frac{\psi}{1+\psi}\right)\frac{1}{\theta}(\tau-\tau_{-1}).$$

Now we are in a position to add up across cohorts to compute social welfare defined as

$$\begin{aligned}\mathcal{W}_0(g,\tau;\tau_{-1}) &= (1-\gamma)\frac{\gamma-\beta\delta}{\gamma(1-\beta\delta)}\sum_{j=-\infty}^{\infty}\gamma^j U_{j,0}(g,\tau;\tau_{-1}) \\ &= (1-\gamma)\frac{\gamma-\beta\delta}{\gamma(1-\beta\delta)}\sum_{j=-\infty}^{-1}\gamma^j U_j + (1-\gamma)\frac{\gamma-\beta\delta}{\gamma(1-\beta\delta)}\sum_{j=0}^{\infty}\gamma^j U_j,\end{aligned}$$

where the second line partitions the population into cohorts born before and after the tax reform. Starting with the agents born after the tax reform,

$$\begin{aligned}(1-\gamma)\frac{\gamma-\beta\delta}{\gamma(1-\beta\delta)}\sum_{j=0}^{\infty}\gamma^j U_j &= (1-\gamma)\frac{\gamma-\beta\delta}{\gamma(1-\beta\delta)}\sum_{j=0}^{\infty}\gamma^j\left((1-\beta\delta)\sum_{a=0}^{\infty}(\beta\delta)^a\bar{u}_a - \bar{v}_Y\right) \\ &= \frac{\gamma-\beta\delta}{\gamma}\sum_{a=0}^{\infty}(\beta\delta)^a\bar{u}_a - \frac{\gamma-\beta\delta}{\gamma(1-\beta\delta)}\bar{v}_Y.\end{aligned}$$

Now

$$\bar{u}_a = \bar{u} - (1-\tau)\, a\frac{v_\omega}{2},$$

so

$$\begin{aligned}\sum_{a=0}^{\infty}(\beta\delta)^a\bar{u}_a &= \frac{\bar{u}}{1-\beta\delta} - (1-\tau)\frac{v_\omega}{2}\left\{\beta\delta + 2(\beta\delta)^2 + ...\right\} \\ &= \frac{\bar{u}}{1-\beta\delta} - \frac{\beta\delta}{(1-\beta\delta)^2}(1-\tau)\frac{v_\omega}{2}\end{aligned}$$

and thus

$$(1-\gamma)\frac{\gamma-\beta\delta}{\gamma(1-\beta\delta)}\sum_{j=0}^{\infty}\gamma^j U_j = \frac{\gamma-\beta\delta}{\gamma(1-\beta\delta)}\bar{u} - \frac{\gamma-\beta\delta}{\gamma(1-\beta\delta)^2}\beta\delta(1-\tau)\frac{v_\omega}{2} - \frac{\gamma-\beta\delta}{\gamma(1-\beta\delta)}\bar{v}_Y.$$

Now consider agents born before the reform (the youngest of which are age 1 at the time of reform):

$$\begin{aligned}(1-\gamma)\frac{\gamma-\beta\delta}{\gamma(1-\beta\delta)}\sum_{j=-\infty}^{-1}\gamma^j U_j &= (1-\gamma)\frac{\gamma-\beta\delta}{\gamma(1-\beta\delta)}\times \\ &\quad \gamma^{-1}(\beta\delta)^1\left\{(1-\beta\delta)\left(\bar{u}_1 + (\beta\delta)\bar{u}_2 + (\beta\delta)^2\bar{u}_3 + ...\right) - \bar{v}_o\right\} \\ &\quad +\gamma^{-2}(\beta\delta)^2\left\{(1-\beta\delta)\left(\bar{u}_2 + (\beta\delta)\bar{u}_3 + ...\right) - \bar{v}_o\right\} + ...\end{aligned}$$

Adding the pieces here involving $\bar{u}$ and $\bar{v}_o$ gives

$$(1-\gamma)\frac{\beta\delta}{\gamma(1-\beta\delta)}\left(\bar{u} - \bar{v}_o\right).$$

The term in v_ω is

$$
\begin{aligned}
& (1-\gamma)\frac{\gamma-\beta\delta}{\gamma(1-\beta\delta)} \times \left(-(1-\tau)\frac{v_\omega}{2}\right) \times (1-\beta\delta) \times \\
& \left\{ \left(\frac{\beta\delta}{\gamma}\right)\left(1+2(\beta\delta)+3(\beta\delta)^2+...\right) + \left(\frac{\beta\delta}{\gamma}\right)^2 \left(2+3(\beta\delta)+4(\beta\delta)^2+...\right)+...\right\} \\
= \; & (1-\gamma)\frac{\gamma-\beta\delta}{\gamma} \times \left(-(1-\tau)\frac{v_\omega}{2}\right) \times \left\{ \frac{\beta\delta}{\gamma}\frac{1}{(1-\beta\delta)^2} + \left(\frac{\beta\delta}{\gamma}\right)^2\left(\frac{1}{1-\beta\delta}+\frac{1}{(1-\beta\delta)^2}\right)+...\right\} \\
= \; & (1-\gamma)\frac{\gamma-\beta\delta}{\gamma} \times \left(-(1-\tau)\frac{v_\omega}{2}\right) \times \left\{ \frac{\frac{\beta\delta}{\gamma}}{(1-\beta\delta)(1-\frac{\beta\delta}{\gamma})}\left[\frac{1}{(1-\beta\delta)}+\frac{\left(\frac{\beta\delta}{\gamma}\right)}{\left(1-\frac{\beta\delta}{\gamma}\right)}\right]\right\}.
\end{aligned}
$$

Now we can add together the contributions to social welfare from agents born before and after the reform. The two terms involving $\bar{u}$ add up exactly to $\bar{u}$. The terms in $\bar{v}_Y$ and $\bar{v}_0$ simplify to give $-\left(\frac{\psi}{1+\psi}\right)\frac{1}{\theta}\left[(1-\tau)+\frac{\beta\delta}{\gamma}\frac{(1-\gamma)}{(1-\beta\delta)}(1-\tau_{-1})\right]$. The term in v_ω is $-\frac{\beta\delta}{\gamma-\beta\delta}(1-\tau)\frac{v_\omega}{2}$. Collecting all these terms gives the expression for social welfare in Proposition 4. In particular, collecting the terms in v_ω, we obtain

$$
-\left[(1-\tau)\frac{\beta\delta}{\gamma-\beta\delta}\frac{v_\omega}{2} - \log\left(\frac{1-\delta\exp\left(\frac{-\tau(1-\tau)}{2}v_\omega\right)}{1-\delta}\right)\right].
$$

When $\gamma=\beta$, the first term in square brackets simplifies to $(1-\tau)\frac{\delta}{1-\delta}\frac{v_\omega}{2}$. The second term can be approximated as follows:

$$
\log\left(\frac{1-\delta\exp\left(\frac{-\tau(1-\tau)}{2}v_\omega\right)}{1-\delta}\right) \approx \log\left(1+\frac{\delta}{1-\delta}\tau(1-\tau)\frac{v_\omega}{2}\right) \approx \tau(1-\tau)\frac{\delta}{1-\delta}\frac{v_\omega}{2}.
$$

Adding the two pieces, we have

$$
(1-\tau)\frac{\delta}{1-\delta}\frac{v_\omega}{2} - \tau(1-\tau)\frac{\delta}{1-\delta}\frac{v_\omega}{2} = (1-\tau)^2\frac{\delta}{1-\delta}\frac{v_\omega}{2} = (1-\tau)^2\frac{v_\alpha}{2},
$$

where the last equality reflects the fact that v_α is the cross-sectional variance of the cumulated innovations ω, and the δ in the numerator reflects our assumption that wage shocks start realizing at age $a=1$. We use this approximate result when we interpret the various components of social welfare in Section V.B. Note that this approximation is extremely accurate for plausible parameter values. For example, when evaluated at $\beta=\gamma$, the empirical $\tau^{US}=0.181$, and the calibrated values $\delta=0.971$ and $v_\omega=0.003$, the accurate and the approximated values for the welfare cost if uninsurable life-cycle risk are 0.03371 and 0.03368, respectively.

B.7. *Proofs of Corollaries 4.1, 4.2, 4.3, 4.4, 4.5*

4.1: In eq. (30), the term in τ_{-1} is additively separable from all the others containing g and τ . **4.2:** Differentiating the expression for social welfare twice with respect to τ, it is straightforward to show that each term except the last one involving insurable risk is strictly concave in τ. The term in insurable risk has a second derivative equal to

$$-\left(1+\chi\right)\frac{\sigma-2\tau}{\left(\sigma+\tau\right)^{4}}\left(1+\sigma\right)^{2}v_{\varepsilon},$$

which is less than or equal to zero if $\sigma \geq 2$. Thus, $\sigma \geq 2$ is a (weak) sufficient condition for global concavity of social welfare with respect to τ. It is straightforward to verify that the social welfare expression is concave in g. **4.3:** Differentiating eq. (30) with respect to τ, the first-order condition has no terms involving g. Thus, the optimal choice for τ is independent of g. **4.4:** Differentiating eq. (30) with respect to g, the first-order condition is

$$\frac{-1}{1-g}+\frac{\chi}{g}=0,$$

which immediately gives the expression for g^* in eq. (31). **4.5:** The parameter η only appears in an additively separable constant in eq. (30). Thus, this parameter does not appear in the first-order conditions defining the optimal choices for g and τ.

B.8. *Proof of Corollary 4.6 [$\gamma = \beta$ case]*

In eq. (29) when $\gamma = \beta$, the constant term Γ simplifies to $\frac{1-\delta}{1-\beta\delta}$. Let $\mathbb{E}\left[u_0\right]$ denote expected period utility for newborn agents from consumption and leisure. The contribution to social welfare from newborn agents is then

$$(1-\beta)\frac{1-\delta}{1-\beta\delta}\cdot\left(1+\beta+\beta^{2}+...\right)\cdot(1-\beta\delta)\mathbb{E}\left[u_{0}\right]=(1-\delta)\cdot\mathbb{E}\left[u_{0}\right].$$

where $(1+\beta+\beta^2+...)$ reflects the weights the planner puts on current and future cohorts of age zero. Note that $(1-\delta)$ is the size of the population at age zero. Similarly, the age 1 component is given by

$$(1-\beta)\frac{1-\delta}{1-\beta\delta}\cdot\left(\beta^{-1}+1+\beta+\beta^{2}+...\right)\cdot(1-\beta\delta)\cdot\beta\delta\mathbb{E}\left[u_{1}\right]=\left(1-\delta\right)\delta\mathbb{E}\left[u_{1}\right]$$

where the term $(1-\delta)\,\delta$ is the size of the population at age 1. And so on. Now we need to compute how skill acquisition costs factor into social welfare. Education costs for the newborn and future cohorts are

$$(1-\beta)\frac{1-\delta}{1-\beta\delta}\cdot\left(1+\beta+\beta^2+...\right)\cdot\mathbb{E}\left[v\left(s\left(\kappa,\tau\right),\kappa\right)\right]=\frac{1-\delta}{1-\beta\delta}\cdot\mathbb{E}\left[v\left(s\left(\kappa,\tau\right),\kappa\right)\right]$$

If skill accumulation decisions are fully reversible, the net skill investment cost for an agent of type κ given a new progressivity value τ and a past progressivity value τ_{-1} is $v\left(s(\kappa;\tau),\kappa\right)-v\left(s(\kappa;\tau_{-1}),\kappa\right)$. Thus, with $\gamma=\beta$, the contribution to social welfare from net skill investments from cohorts who entered the economy in the past (ages $a=1,2,...$) is

$$\begin{aligned}&(1-\beta)\frac{1-\delta}{1-\beta\delta}\cdot\sum_{a=1}^{\infty}\left(\frac{\beta\delta}{\beta}\right)^{a}\mathbb{E}\left[v\left(s(\kappa;\tau),\kappa\right)-v\left(s(\kappa;\tau_{-1}),\kappa\right)\right]\\ =\;&\frac{\delta(1-\beta)}{1-\beta\delta}\cdot\mathbb{E}\left[v\left(s(\kappa;\tau),\kappa\right)-v\left(s(\kappa;\tau_{-1}),\kappa\right)\right]\end{aligned}$$

Adding the two pieces, and ignoring the term in τ_{-1}, since it is separable from τ and thus irrelevant for optimization, gives

$$\mathbb{E}\left[v\left(s\left(\kappa,\tau\right),\kappa\right)\right]=\left(\frac{\psi}{1+\psi}\right)\frac{1-\tau}{\theta}$$

Adding up these various welfare components gives the expression for social welfare in eq. (32).

B.9. Proof of Proposition 5 [efficiency in RA model]

We will prove this proposition for general CRRA utility with risk aversion coefficient $\gamma>0$. The baseline model corresponds to the special case $\gamma=1$. Absent household heterogeneity, the planner's problem (for a planner with access to lump-sum taxes) is

$$\max_{H,C,G}\left\{\frac{C^{1-\gamma}}{1-\gamma}-\frac{H^{1+\sigma}}{1+\sigma}+\chi^{\gamma}\frac{G^{1-\gamma}}{1-\gamma}\right\}$$

such that

$$C+G=H.$$

The first-order conditions give the following solution

$$\begin{aligned}H^{*}&=(1+\chi)^{\frac{\gamma}{\gamma+\sigma}}\\ g^{*}&=\frac{G^{*}}{H^{*}}=\frac{\chi}{1+\chi}.\end{aligned}$$

We now show that a planner with access only to the tax schedule in (1) can replicate the same allocations by setting $\tau = -\chi$ and $\lambda = (1+\chi)^{-1-\frac{\chi\gamma}{\gamma+\sigma}}$. To verify this, consider the representative household's problem:

$$\max_{C,H} \left\{ \frac{C^{1-\gamma}}{1-\gamma} - \frac{H^{1+\sigma}}{1+\sigma} \right\}$$
$$s.t.$$
$$C = \lambda H^{1-\tau}.$$

with FOC

$$\left(\lambda H^{1-\tau}\right)^{-\gamma} \lambda(1-\tau)H^{-\tau} = H^{\sigma}.$$

Substituting in the candidate expressions for τ and λ gives

$$H = (1+\chi)^{\frac{\gamma}{\sigma+\gamma}} = H^{*}.$$

Substituting this expression for hours plus the candidate expressions for τ and λ into the government budget constraint gives

$$G = H - \lambda H^{1-\tau} = \frac{\chi}{1+\chi} H^{*}$$

so that $G/H = G^{*}/H^{*}$.

B.10. Derivation of welfare cost of between-skill consumption dispersion

The skill-related component of consumption, $p\,(s)^{1-\tau}$, is Pareto distributed with parameter $P = \theta/(1-\tau)$. If consumption is Pareto distributed with Pareto parameter P, the expected value for consumption is $P/(P-1)$. Log consumption is then exponentially distributed, with exponential parameter P and the expected value for log consumption is $1/P$. Let F_c denote the Pareto cumulative distribution function (CDF) for this consumption, and let F_z denote the Exponential CDF for $z = \log c$. The welfare cost ϖ of consumption dispersion (assuming logarithmic preferences) can then be calculated as the fraction ϖ by which safe consumption must be reduced to deliver the same expected utility as risky consumption:

$$\begin{aligned}
\log\left(\frac{P}{P-1}(1-\varpi)\right) &= \int \log c\, dF_c \\
\log\left(\frac{P}{P-1}(1-\varpi)\right) &= \int z\, dF_z \\
\log\left(\frac{P}{P-1}\right) + \log(1-\varpi) &= \frac{1}{P}
\end{aligned}$$

$$\varpi \approx -\log(1-\varpi) = \log\left(\frac{P}{P-1} - \frac{1}{P}\right)$$

Since in our economy $P = \theta/(1-\tau)$ and ϖ is small, ϖ is approximately equal to the expression in eq. (36).

B.11. Proof of Proposition 6 [condition for optimal progressivity]

Assume $\gamma = \beta$ and approximate the sixth line of the social welfare expression (eq. 30) by $-(1-\tau)^2 \frac{v_\alpha}{2}$. Then the derivative of the social welfare expression with respect to τ is

$$\frac{\partial \mathcal{W}(g,\tau;\tau_{-1})}{\partial \tau}|_{\tau=0} = -\left(\frac{1}{1+\sigma} + \frac{\psi}{1+\psi}\frac{1}{\theta-1}\right)(1+\chi) - \frac{1}{1+\psi}\frac{1}{\theta} + \frac{1}{1+\sigma} + \frac{1}{\theta-1} + v_\varphi + v_\alpha.$$

It is immediate that this derivative is positive if and only if the condition in Proposition 6 is satisfied.

B.12. Proof of Proposition 7 [inequality aversion]

We begin by computing expected utility (excluding skill investment costs) for an agent with states $(\kappa, \varphi, \alpha)$ prior to ε being drawn. Substituting $(B16)$ into $(B17)$ and taking the expected value of this prior to ε being drawn yields, after some simplifications:

$$\tilde{u}(\kappa,\varphi,\alpha;g,\tau) = \mathcal{C}(g,\tau) + (1-\tau)\,\alpha - (1-\tau)\,\varphi + (1-\tau)\frac{\eta}{\theta}\kappa - \frac{1-\tau}{1+\sigma} + \chi G(g,\tau)$$

where $\mathcal{C}(g,\tau)$ captures the components showing up in the log consumption allocation that are common across all individuals

$$\begin{aligned}
\mathcal{C}(g,\tau) \;\; = \;\; & \log(1-g) + \frac{(1-\tau)}{(\sigma+\tau)^2}(\sigma + 2\tau + \sigma\tau)\frac{v_\varepsilon}{2} \\
& + \left(\frac{1}{1+\sigma} + \frac{\psi}{1+\psi}\frac{1}{(\theta-1)}\right)\log(1-\tau) \\
& + \tau(1-\tau)\frac{v_\varphi}{2} + \tau(1-\tau)\frac{v_\alpha^0}{2} + \log\left(\frac{1-\delta\exp\left(\frac{-\tau(1-\tau)}{2}v_\omega\right)}{1-\delta}\right) \\
& - \frac{1}{(\theta-1)(\psi+1)}\log(\eta) + \frac{\theta}{(\theta-1)}\log\left(\frac{\theta}{\theta-1}\right) - \frac{\psi}{(1+\psi)(\theta-1)}\log(\theta) \\
& + \log\left(\frac{\theta-1+\tau}{\theta}\right)
\end{aligned}$$

and where

$$\begin{aligned} G(g,\tau) &= \log g + \log Y(\tau) \\ &= \log g + \left(\frac{\tau(1+\hat{\sigma})+\hat{\sigma}}{\hat{\sigma}^2}\frac{v_\varepsilon}{2}\right) + \log\left(\left(\frac{\theta}{\theta-1}\right)^{\frac{\theta}{\theta-1}}\right) + \log\left(\eta^{\frac{1}{1+\psi}\frac{1}{1-\theta}}\theta^{\frac{\psi}{\psi+1}\frac{1}{1-\theta}}\right) \\ &\quad + \log\left((1-\tau)^{\frac{1}{1+\sigma}-\frac{\psi}{\psi+1}\frac{1}{1-\theta}}\right). \end{aligned}$$

Now, let's add the net cost of skill investment:

$$v(\kappa;\tau) = \begin{cases} [(1-\tau)-(1-\tau_{-1})]\frac{\psi}{(1+\psi)}\frac{\eta}{\theta}\kappa & \text{if } a>0 \\ (1-\tau)\frac{\psi}{(1+\psi)}\frac{\eta}{\theta}\kappa & \text{if } a=0. \end{cases}$$

Next, we compute the value for constant consumption forever that, assuming equilibrium hours and skill investment, gives an agent lifetime utility equal to what would accrue, in expectation, to an agent with type $(\kappa,\varphi,\alpha;g,\tau)$. Define the answer as

$$\bar{c}(\kappa,\varphi,\alpha;g,\tau) = \exp\left(\mathcal{C}(g,\tau) + (1-\tau)\alpha - (1-\tau)\varphi + (1-\tau)\frac{\eta}{\theta}\kappa - (1-\tau)\frac{v_\omega}{2}\frac{\beta\delta}{1-\beta\delta}\right).$$

Now suppose that the contribution to social welfare from the cohort of age a at the time of the tax reform is an age-dependent weight times

$$\begin{aligned} \mathcal{V}_a(g,\tau) = \quad & \log\left(\int\int\int \bar{c}(\kappa,\varphi,\alpha;g,\tau)^{1-\nu}dF_\kappa dF_\varphi dF_\alpha^a\right)^{\frac{1}{1-\nu}} - \frac{1-\tau}{1+\sigma} + \chi G(g,\tau) \\ & - \frac{\psi}{(1+\psi)\theta}\left[(1-\tau) - \mathbb{I}_{\{a>0\}}(1-\tau_{-1})\right]. \end{aligned}$$

Integrating the first term and rearranging yields:

$$\begin{aligned} \mathcal{V}_a(g,\tau) &= \mathcal{C}(g,\tau) - (1-\tau)\frac{v_\omega}{2}\frac{\beta\delta}{1-\beta\delta} + (1-\tau)\left((1-\tau)(1-\nu)-1\right)\frac{(v_\varphi + a v_\omega)}{2} \\ &\quad + \frac{1}{\nu-1}\log\left((\nu-1)\frac{1}{\theta}(1-\tau)+1\right) \\ &\quad - \frac{1-\tau}{1+\sigma} + \chi G(g,\tau) - \frac{\psi}{(1+\psi)\theta}\left[(1-\tau) - \mathbb{I}_{\{a>0\}}(1-\tau_{-1})\right]. \end{aligned}$$

The planner's objective is

$$\text{(B21)} \quad \mathcal{W}_0(g,\tau) = (1-\gamma)\frac{\gamma-\beta\delta}{\gamma(1-\beta\delta)}\sum_{a=1}^{\infty}\gamma^{-a}(\beta\delta)^a\mathcal{V}_a(g,\tau) + (1-\gamma)\frac{\gamma-\beta\delta}{\gamma(1-\beta\delta)}\sum_{j=0}^{\infty}\gamma^j\mathcal{V}_0(g,\tau).$$

Aggregating across cohorts using $(B21)$ with $\gamma = \beta$ yields, after some manipulations, the expression in Proposition 7. To conclude, it is useful to explain why lines (4)-(6) in Proposition 7 measure the cost of consumption inequality reflecting skill (κ) differentials, preference heterogeneity, and past uninsurable shocks. Define ω_κ as the solution to

$$\log\left(\int \exp\left((1-\tau)\frac{\eta}{\theta}\kappa\right)^{1-\nu} dF_\kappa\right)^{1/(1-\nu)} = \log[(1+\omega_\kappa)\theta/(\theta-(1-\tau))]$$

where $\exp[(1-\tau)\eta/\theta \cdot \kappa]$ is the factor in consumption attributable to κ and $\theta/(\theta-(1-\tau))$ is the expected value for this factor. Given our exponential distribution for κ, $\omega_\kappa \approx \log(1-(1-\tau)/\theta)+$ $1/(\nu-1)\cdot\log(1+(\nu-1)\cdot(1-\tau)/\theta)$. Similarly, for the welfare cost of consumption inequality due to heterogeneity in the taste for leisure, line (5), define ω_φ as the solution to

$$\log\left(\int \exp\left(-(1-\tau)\varphi\right)^{1-\nu} dF_\varphi\right)^{1/(1-\nu)} = \log\left((1+\omega)\exp\left((1-\tau)\left((1-\tau)-1\right)v_\varphi/2\right)\right)$$

where $\exp\left(-(1-\tau)\varphi\right)$ is the factor in consumption attributable to φ, and $\exp\left((1-\tau)\left((1-\tau)-1\right)v_\varphi/2\right)$ is the expected value for this factor. Given our Normal distribution for φ, $\omega_\varphi = -(1-\tau)^2\nu\cdot v_\varphi/2$. Line (6) in the welfare expression in Proposition 7 is obtained analogously.

B.13. Proof of Propositions 8-9 [median voter]

To understand Proposition 8, note that the preferred value for g for agent i obeys the first-order condition

$$\chi\frac{1}{g} = \frac{c_i}{1-g}\cdot\frac{1}{c_i}. \tag{B22}$$

The left-hand side is the benefit from a marginal increase in the share of output devoted to publicly provided goods, which, given separable preferences, is identical across agents. The right-hand side is the cost associated with a marginal increase in g. Since c_i can be expressed as $c_i(g,\tau) = \lambda(g)\,\bar{c}_i(\tau) = (1-g)\,\bar{\Lambda}\bar{c}_i(\tau)$, where the terms $\bar{\Lambda}$ and $\bar{c}_i(\tau)$ are independent of g, the derivative of individual consumption with respect to g is (minus) the first term on the right-hand side of (B22). The second term is the marginal utility of private consumption.

We now move to the proof of Proposition 9. Substituting the allocations $s(\kappa;\tau)$, $h(\varphi,\varepsilon;\tau)$, and $c(\varphi,\alpha,s;g,\tau)$ into expected utility (7) yields

$$\begin{aligned}
U &= -\frac{(1-\tau)\,\eta}{(1+1/\psi)\,\theta}\kappa+(1-\beta\delta)\,\mathbb{E}\sum_{j=0}^{\infty}(\beta\delta)^j\left\{\log\lambda\,(g,\tau)+\frac{\log(1-\tau)}{(1+\hat{\sigma})}+\mathcal{M}+(1-\tau)\left[\alpha_j-\varphi+\kappa\frac{\eta}{\theta}\right]\right.\\
&\quad\left.+\frac{1-\tau}{(\theta-1)\,(\psi+1)}\left(\psi\log\,(1-\tau)+\log\left(\frac{1}{\eta}\frac{\theta}{(\theta-1)^{(1+\psi)}}\right)\right)-\frac{1-\tau}{1+\sigma}\exp\left(-\frac{1+\sigma}{\widehat{\sigma}(1-\tau)}\mathcal{M}+\frac{1+\sigma}{\widehat{\sigma}}\varepsilon\right)\right\}\\
&\quad+\chi\log G.
\end{aligned}$$

Now, suppose that the choice of τ is made before observing ε. Then, the term in ε becomes $-\frac{1-\tau}{1+\sigma}$ (see the proof of Corollary 6.1). In addition,

$$(1-\beta\delta)\,\mathbb{E}\sum_{j=0}^{\infty}(\beta\delta)^j\,(1-\tau)\,\alpha_j=(1-\tau)\left(\alpha-\frac{\beta\delta}{1-\beta\delta}\frac{v_\omega}{2}\right).$$

Thus,

$$\begin{aligned}
\int U\,(\kappa,\varphi,\alpha,\varepsilon;g,\tau)\,dF_\varepsilon &= \log\lambda\,(g,\tau)+\frac{\log(1-\tau)}{(1+\hat{\sigma})}+\mathcal{M}\\
&\quad+\frac{1-\tau}{(\theta-1)\,(1+\psi)}\left(\psi\log\,(1-\tau)+\log\left(\frac{1}{\eta}\frac{\theta}{(\theta-1)^{(1+\psi)}}\right)\right)\\
&\quad-(1-\tau)\,\varphi+\frac{1-\tau}{1+\psi}\frac{\eta}{\theta}\kappa-\frac{1-\tau}{1+\sigma}\\
&\quad+(1-\tau)\left(\alpha-\frac{\beta\delta}{1-\beta\delta}\frac{v_\omega}{2}\right)+\chi\log G.
\end{aligned}$$

Recall that the baseline social welfare function is

$$\begin{aligned}
\mathcal{W}(g,\tau) &= \log\lambda\,(g,\tau)+\frac{\log(1-\tau)}{(1+\hat{\sigma})}+\mathcal{M}+\frac{1-\tau}{(\theta-1)\,(1+\psi)}\left(\psi\log\,(1-\tau)+\log\left(\frac{1}{\eta}\frac{\theta}{(\theta-1)^{(1+\psi)}}\right)\right)\\
&\quad-(1-\tau)\,\frac{v_\varphi}{2}+\frac{(1-\tau)}{(1+\psi)\,\theta}-\frac{1-\tau}{1+\sigma}\\
&\quad-(1-\tau)\left(\frac{\beta\delta}{\gamma-\beta\delta}\frac{v_\omega}{2}\right)+\chi\log G.
\end{aligned}$$

Therefore, we can express expected utility, conditional on the state (κ,φ,α), as

$$\begin{aligned}
U\,(\kappa,\varphi,\alpha;g,\tau) &= \mathcal{W}(g,\tau)+(1-\tau)\frac{v_\varphi}{2}-\frac{(1-\tau)}{(1+\psi)\,\theta}\\
&\quad+(1-\tau)\left(\frac{\beta\delta}{\gamma-\beta\delta}\frac{v_\omega}{2}-\frac{\beta\delta}{1-\beta\delta}\frac{v_\omega}{2}\right)\\
&\quad+(1-\tau)\left(\alpha-\varphi+\frac{1}{1+\psi}\frac{\eta}{\theta}\kappa\right).
\end{aligned}$$

Note that $U(\kappa, \varphi, \alpha; g, \tau)$ is strictly concave in τ, since $\mathcal{W}(g, \tau)$ is concave in τ and the additional terms in $U(\kappa, \varphi, \alpha; g, \tau)$ are linear in τ. We need to determine the median voter. A useful property is that the three individual states $(\kappa, \alpha, \varphi)$ enter as a linear combination. Let

$$x = \alpha - \varphi + \frac{1}{1+\psi}\frac{\eta}{\theta}\kappa.$$

The median voter is the agent with the median value for x. Since α and φ are normally distributed and κ is exponentially distributed, the random variable x follows an Exponentially Modified Gaussian distribution.

B.14. Skill investment constraints

B.14.1 Consumption for skilled and unskilled. We begin by calculating the consumption of skilled and unskilled workers. This is needed to evaluate the probabilities that consumption of skilled and unskilled parents exceed the threshold $\underline{c}$. These probabilities are in turn inputs to the transition of ξ, see equation (46). Consumption of a skilled person is

$$\begin{aligned} \log c &= \log \lambda(\xi) + \frac{1-\tau}{1+\sigma}\log(1-\tau) + \frac{(1-\tau)[1-2\tau-\sigma\tau]}{(\sigma+\tau)}\frac{v_\varepsilon}{2} + (1-\tau)[\log p(s) + \alpha - \varphi] \\ &= \log \lambda(\xi) + \frac{1-\tau}{1+\sigma}\log(1-\tau) + \frac{(1-\tau)[1-2\tau-\sigma\tau]}{(\sigma+\tau)}\frac{v_\varepsilon}{2} + (1-\tau)[\alpha - \varphi + \pi_0 + \pi_1 s] \\ &= C(\tau, \xi) + (1-\tau)\left[\alpha - \varphi + \frac{\eta}{\theta}\kappa + \pi_0\right], \end{aligned}$$

where $\lambda(\xi)$ is defined below and the constant $C(\tau, \xi)$ is defined as

$$C(\tau, \xi) = \log \lambda(\xi) + \frac{1-\tau}{1+\sigma}\log(1-\tau) + \frac{(1-\tau)[1-2\tau-\sigma\tau]}{(\sigma+\tau)}\frac{v_\varepsilon}{2}$$

Moreover, the random variable $\alpha - \varphi + \frac{\eta}{\theta}\kappa$ is distributed according to an Exponentially Modified Gaussian distribution with an exponential parameter θ and Gaussian parameters $\left(-\frac{1}{2}(v_\alpha + v_\varphi), v_\alpha + v_\varphi\right)$. The probability that the child of a skilled parent will become unskilled is therefore

$$\begin{aligned} \Pr(c < \underline{c}|\text{skilled}) &= \Pr\left(\alpha - \varphi + p(s) < \frac{\ln \underline{c} - C(\tau, \xi)}{1-\tau} - \pi_0\right) \\ &= \Pr\left(\alpha - \varphi + \frac{\eta}{\theta}\kappa < \frac{\ln \underline{c} - C(\tau, \xi)}{1-\tau} - \pi_0\right). \end{aligned}$$

Consumption of an unskilled person is

$$\begin{aligned} \ln \underline{c}(\alpha, \varphi, \xi) &= \log \lambda(\xi) + \frac{1-\tau}{1+\sigma}\log(1-\tau) + \frac{(1-\tau)[1-2\tau-\sigma\tau]}{(\sigma+\tau)}\frac{v_\varepsilon}{2} + (1-\tau)[\alpha - \varphi + \ln \underline{w}] \\ &= C(\tau, \xi) + (1-\tau)[\alpha - \varphi + \ln \underline{w}] \end{aligned}$$

The probability that the child of an unskilled parent will be unskilled is

$$\begin{aligned}\Pr\left(c<\underline{c}|\text{unskilled,}\xi\right) &= \Pr\left(C\left(\tau,\xi\right)+\left(1-\tau\right)\left[\alpha-\varphi+\ln\underline{w}\right]<\ln\underline{c}\right)\\ &= \Phi\left(\frac{1}{\sqrt{v_\alpha+v_\varphi}}\left(\frac{\ln\underline{c}-C\left(\tau,\xi\right)}{1-\tau}-\ln\underline{w}+\frac{v_\alpha+v_\varphi}{2}\right)\right)\end{aligned}$$

where Φ is a c.d.f. of a standard Normal distribution.

B.14.2 Calculating $\lambda(\xi)$. Recall that effective hours worked N is independent of the skill level. Aggregate output per effective hour worked is therefore

$$\begin{aligned}\frac{Y\left(\xi\right)}{N} &= \xi\frac{Y_U}{N}+\left(1-\xi\right)\frac{Y_S}{N}\\ &= \xi\underline{w}+\left(1-\xi\right)\frac{\theta}{\theta-1}\cdot\exp\left(\pi_0\right).\end{aligned}$$

Define $\tilde{Y}$ as $\tilde{Y}=\int y_i^{1-\tau}di$. Budget balance and $G=gY$ then implies $\lambda=\left(1-g\right)Y/\tilde{Y}$. To compute $\tilde{Y}$, it is useful to aggregate by age group. Let $\tilde{Y}^a$ denote average per capita disposable income for agents of age a:

$$\begin{aligned}\tilde{Y}^a\left(\xi\right) &= \int\left[y\left(s,\varphi,\varepsilon,\alpha\right)\right]^{1-\tau}m\left(s\right)dsdF_\alpha^a dF_\varphi dF_\varepsilon\\ &= \xi\int\left[h\left(\varepsilon,\varphi\right)\exp\left(\alpha_a+\varepsilon\right)\underline{w}\right]^{1-\tau}dF_\alpha^a dF_\varphi dF_\varepsilon\\ &\quad+\left(1-\xi\right)\int\left[h\left(\varepsilon\right)\exp\left(p\left(s\right)+\alpha_a+\varepsilon\right)\right]^{1-\tau}m\left(s\right)dsdF_\alpha^a dF_\varphi dF_\varepsilon\\ &= \left(\xi\left(\underline{w}\right)^{1-\tau}+\left(1-\xi\right)\int\left[\exp\left(\left(1-\tau\right)\left(\pi_0+\pi_1 s\right)\right)\right]^{1-\tau}m\left(s\right)ds\right)\int\left[h\left(\varepsilon,\varphi\right)\exp\left(\alpha_a+\varepsilon\right)\right]^{1-\tau}dF_\alpha^a dF_\varphi dF_\varepsilon.\end{aligned}$$

Substituting in the hours allocation (20), the expression for the skill price (22), the density function $m\left(s\right)$ (B9), and integrating, we arrive at

$$\ln\tilde{Y}^a\left(\xi\right)=\mathcal{L}-\tau\left(1-\tau\right)\frac{v_\alpha^a}{2}+\ln\left[\xi\left(\underline{w}\right)^{1-\tau}+\left(1-\xi\right)\exp\left(\left(1-\tau\right)\pi_0\right)\frac{\theta}{\theta-1+\tau}\right],$$

where the constant $\mathcal{L}$, common for all cohorts, is defined as

$$\mathcal{L}=\frac{1-\tau}{1+\sigma}\ln\left(1-\tau\right)-\tau(1-\tau)\frac{v_\varphi}{2}+\mathcal{M},$$

where $\mathcal{M}$ is defined in Proposition 1. Now sum across age groups to obtain

$$\begin{aligned}\tilde{Y}\left(\xi\right) &= \left(1-\delta\right)\sum_{a=0}^{\infty}\delta^a\tilde{Y}^a=\exp\left(\mathcal{L}\right)\times\left[\xi\left(\underline{w}\right)^{1-\tau}+\left(1-\xi\right)\exp\left(\left(1-\tau\right)\pi_0\right)\frac{\theta}{\theta-1+\tau}\right]\times\\ &\quad\frac{\left(1-\delta\right)\exp\left(-\tau(1-\tau)\frac{v_\alpha^0}{2}\right)}{1-\delta\exp\left(\frac{-\tau(1-\tau)}{2}v_\omega\right)}.\end{aligned}$$

Substituting (B12) and (B15) into (B14) and simplifying, we arrive at a solution for the equilibrium value of λ which, in logs, is

$$\begin{aligned}
\ln\lambda(\xi) &= \ln(1-g)+\ln Y-\ln\tilde{Y}\\
&= \ln(1-g)+\ln\left[\xi\underline{w}+(1-\xi)\frac{\theta}{\theta-1}\cdot\exp(\pi_0)\right]+\frac{\ln(1-\tau)}{1+\sigma}+\frac{\tau(1+\hat{\sigma})+\hat{\sigma}}{\hat{\sigma}^2}\frac{v_\varepsilon}{2}\\
&\quad -\left(\frac{1-\tau}{1+\sigma}\ln(1-\tau)-\tau(1-\tau)\frac{v_\varphi}{2}+\frac{(1-\tau)(1-\tau(1+\widehat{\sigma}))}{\widehat{\sigma}}\frac{v_\varepsilon}{2}\right)\\
&\quad -\ln\left[\xi(\underline{w})^{1-\tau}+(1-\xi)\exp((1-\tau)\pi_0)\frac{\theta}{\theta-1+\tau}\right]\\
&\quad -\ln(1-\delta)+\tau(1-\tau)\frac{v_\alpha^0}{2}+\ln\left(1-\delta\exp\left(\frac{-\tau(1-\tau)}{2}v_\omega\right)\right)
\end{aligned}$$

It follows that the model with unskilled workers and investment constraints changes λ relative to its equivalent in the benchmark economy analyzed in Sections III.–VI., i.e., λ_{BM}, as follows,

$$\ln\lambda(\xi)=\ln\lambda_{BM}+\ln\left(\frac{\theta-1}{\theta-1+\tau}\right)-\tau\pi_0+\ln\left(\frac{\xi\underline{w}+(1-\xi)\frac{\theta}{\theta-1}\cdot\exp(\pi_0)}{\xi(\underline{w})^{1-\tau}+(1-\xi)\frac{\theta}{\theta-1+\tau}\exp((1-\tau)\pi_0)}\right).$$

Note that $\lambda(0)=\lambda_{BM}$.

B.14.3 Social welfare in the presence of investment constraints. Consider now the expression for social welfare when the share of unskilled, ξ_t, is changing over time. It is convenient to express social welfare by adding up current utilities, appropriately discounted, at each point in time;

$$\begin{aligned}
\mathcal{W}(g,\tau;\{\xi_t\}_{t=0}^{\infty}) &= \frac{(1-\gamma)(\gamma-\beta\delta)}{\gamma(1-\beta\delta)}\left[\sum_{t=0}^{\infty}\left(\gamma^t(\beta\delta)^0\bar{u}_{0,t}+\gamma^{t-1}(\beta\delta)^1\bar{u}_{1,t}+\gamma^{t-2}(\beta\delta)^2\bar{u}_{2,t}+...\right)\right.\\
&\quad\left.-\sum_{t=0}^{\infty}\gamma^t(1-\xi_t)\bar{\nu}-(1-\xi_0)\sum_{j=-\infty}^{-1}\gamma^j(\beta\delta)^{-j}\bar{\nu}\right],
\end{aligned}$$

where $\bar{u}_{a,t}$ denotes the expected current utility (excluding the education costs) of an individual who is a years old in period t. The last term captures the education costs of the individuals alive at the time of the reform. Recall that education costs are fully reversible, so we can ignore the investments these individuals made before the reform. Note that the age component of $\bar{u}_{a,t}$ does not depend on ξ_t and is therefore time invariant. We can therefore express it as $\bar{u}_{a,t}=\bar{u}_t+\bar{u}^a$, where $\bar{u}^a=(1-\beta\delta)(1-\tau)jv_\omega/2$ (see the proof of Proposition 1). It follows that $\mathcal{W}$ can be expressed, after some algebra, as

$$\begin{aligned}\mathcal{W}\left(g,\tau;\{\xi_t\}_{t=0}^{\infty}\right) &= (1-\gamma)\sum_{t=0}^{\infty}\gamma^t\frac{\bar{u}_t}{1-\beta\delta}-(1-\tau)\frac{v_\omega}{2}\frac{\beta\delta}{\gamma-\beta\delta}\\ &\quad -\left(\frac{\beta\delta}{\gamma}(1-\xi_0)+\frac{\gamma-\beta\delta}{\gamma}\sum_{t=0}^{\infty}\gamma^t(1-\xi_t)\right)\frac{(1-\gamma)}{(1-\beta\delta)}\bar{\nu}.\end{aligned}$$

We must now calculate expected period utility excluding education costs, $\bar{u}_t$. Recall that the current utility – excluding the investment cost – for a skilled agent with state $(\kappa, a, \alpha, \varphi, \varepsilon, \xi)$ is Recall that the current utility – excluding the investment cost – for a skilled agent with state $(\kappa, a, \alpha, \varphi, \varepsilon, \xi)$ is

$$\begin{aligned}u(c_a,h,G) &= \chi\left(\ln g+\ln\left(\frac{Y}{N}\right)+\ln N\right)+\log\lambda+\frac{1-\tau}{1+\sigma}\log(1-\tau)\\ &\quad -(1-\tau)\varphi+(1-\tau)\alpha_a-\exp\left(-\frac{1+\sigma}{\widehat{\sigma}(1-\tau)}\mathcal{M}\right)\exp\left(\frac{1+\sigma}{\widehat{\sigma}}\varepsilon\right)\frac{(1-\tau)}{(1+\sigma)}+\mathcal{M}\\ &\quad +(1-\tau)\kappa\frac{\eta}{\theta}+(1-\tau)\pi_0,\end{aligned}$$

The utility for an unskilled person is the same, except that the last line is replaced by $(1-\tau)\ln\underline{w}$. Take the cross-sectional expectation over $u(c_a, h, G)$ w.r.t. ε, φ, and α_a in equation (B23). The first two lines of the welfare then become

$$\begin{aligned}&\chi\ln g+\chi\ln\left(\xi\underline{w}+(1-\xi)\frac{\theta}{\theta-1}\cdot\exp(\pi_0)\right)+\chi\left(\frac{1}{1+\sigma}\ln(1-\tau)+\frac{\tau(1+\hat{\sigma})+\hat{\sigma}}{\hat{\sigma}^2}\frac{v_\varepsilon}{2}\right)+\log\lambda\\ &+\frac{1-\tau}{1+\sigma}\log(1-\tau)-\frac{1-\tau}{1+\sigma}+\mathcal{M}-(1-\tau)\frac{v_\varphi+v_{\alpha_a}}{2}\end{aligned}$$

Now take the expectation over κ in equation (B23), which implies that the third line yields

$$E\left\{(1-\tau)\kappa\frac{\eta}{\theta}+(1-\tau)\pi_0\right\}=(1-\tau)\pi_0+\frac{1-\tau}{\theta}$$

Calculate now the cross-sectional utility cost of skilled investment:

$$E\left\{-v\left(s\left(\kappa;\tau\right)\right)\right\}=-\frac{\psi}{1+\psi}\frac{1}{\theta}(1-\tau)$$

When adding up current utility across generations, the term $-(1-\tau)^2 v_{\alpha_a}/2$ becomes, as in Proposition 3,

$$-\left[(1-\tau)\frac{\beta\delta}{\gamma-\beta\delta}\frac{v_\omega}{2}-\log\left(\frac{1-\delta\exp\left(\frac{-\tau(1-\tau)}{2}v_\omega\right)}{1-\delta}\right)\right]$$

It follows that social welfare can be expressed as

$$\begin{aligned}\mathcal{W}\left(g,\tau;\{\xi_t\}_{t=0}^{\infty}\right) &= (1-\gamma)\sum_{t=0}^{\infty}\gamma^t\frac{\bar{u}_t}{1-\beta\delta}-(1-\tau)\frac{v_\omega}{2}\frac{\beta\delta}{\gamma-\beta\delta}\\ &\quad-\left(\frac{\beta\delta}{\gamma}(1-\xi_0)+\frac{\gamma-\beta\delta}{\gamma}\sum_{t=0}^{\infty}\gamma^t(1-\xi_t)\right)\frac{(1-\gamma)}{(1-\beta\delta)}\frac{\psi}{1+\psi}\frac{1-\tau}{\theta}.\end{aligned}$$

where

$$\begin{aligned}\frac{\bar{u}_t}{1-\beta\delta} &= \ln(1-g)+\chi\ln g+(1+\chi)\frac{\tau}{1+\sigma}\ln(1-\tau)-\frac{1-\tau}{1+\sigma}\\ &\quad+(1+\chi)\frac{\tau(1+\hat{\sigma})+\hat{\sigma}}{\hat{\sigma}^2}\frac{v_\varepsilon}{2}\\ &\quad-(1-\tau)^2\frac{v_\varphi+v_\alpha^0}{2}-\ln(1-\delta)+\ln\left(1-\delta\exp\left(\frac{-\tau(1-\tau)}{2}v_\omega\right)\right)\\ &\quad+(1+\chi)\ln\left[\xi_t\underline{w}+(1-\xi_t)\frac{\theta}{\theta-1}\cdot\exp(\pi_0)\right]\\ &\quad-\ln\left[\xi_t(\underline{w})^{1-\tau}+(1-\xi_t)\exp((1-\tau)\pi_0)\frac{\theta}{\theta-1+\tau}\right]\\ &\quad+(1-\xi_t)(1-\tau)\left(\pi_0+\frac{1}{\theta}\right)+\xi_t(1-\tau)\ln\underline{w},\end{aligned}$$

Note that, as in the benchmark case, the public good provision g does not interact with neither τ nor ξ, so it will not matter for the optimal τ. With $\gamma=\beta$ the welfare expression simplifies to

$$\begin{aligned}\mathcal{W} &= \frac{1-\beta}{1-\beta\delta}\sum_{t=0}^{\infty}\beta^t\bar{u}(\xi_t)-(1-\tau)\frac{\delta}{1-\delta}\frac{v_\omega}{2}-\frac{\psi}{1+\psi}\frac{1-\tau}{\theta}\frac{(1-\beta)}{(1-\beta\delta)}\left(1-\xi_0+(1-\delta)\sum_{t=1}^{\infty}\gamma^t(1-\xi_t)\right)\\ &= \frac{(1-\beta)}{(1-\beta\delta)}\left[\sum_{t=0}^{\infty}\beta^t\bar{u}(\xi_t)-\left(1-\xi_0+(1-\delta)\sum_{t=1}^{\infty}\gamma^t(1-\xi_t)\right)\frac{\psi}{1+\psi}\frac{1-\tau}{\theta}\right]\end{aligned}$$

For each possible τ, we solve numerically for the sequence $\{\xi_t\}_{t=0}^{\infty}$ and the associated social welfare $\mathcal{W}$.

APPENDIX C

This Appendix provides additional detail to Section VIII in the main paper. We begin by describing the construction of the data set, and next we present a sensitivity analysis for our regressions.

C.1. Cross-country dataset construction

Tax progressivity. The Andrew Young School World Tax Indicator (WTI) database (Andrew Young School of Policy Studies, 2010), contains measures of personal income tax rates and tax progressivity for 189 countries during the period 1981-2005 (for many countries data are missing in the first part of the sample). The enclosed documentation states that tax rates adjust for allowances/deductions, tax credits, significant local taxes and other main rules of the tax code. For each country-year pair (c, t), the dataset reports an index of marginal rate progression (MRP_t^c), i.e. the slope coefficient from regressing marginal tax rates on the log of gross income. Recall that in our model, $MTR\,(y) = 1 - \lambda\,(1 - \tau)\,y^{-\tau}$, thus using the approximation $\log\,(1 - MTR\,(y)) \simeq -MTR\,(y)$ reveals that this slope coefficient approximately equals τ.

The data also contain average and marginal tax rates at $y, 2y, 3y$, and y where y is average per-capita income. Thus, we can also estimate exactly the parameter τ in our tax function for country c at time t as

$$\tau_t^c\,(y) = 1 - \frac{1 - MTR_t^c\,(y)}{1 - ATR_t^c\,(y)}, \tag{C1}$$

and averaging over the four levels of y available, we obtain an estimate of τ_t^c. Since information on marginal and average tax rates below average income is used to calculate the MRP_t^c index, but it is not available in the public data, and thus we cannot use it for this alternative strategy, the MRP_t^c index is a more comprehensive measure of progressivity and is the one we use in the baseline regressions below. The correlation between the two measures is 0.92 and significant at 1 pct level, and our results are robust to using the alternative estimate of τ_t^c as we show below.

The crucial feature of these data is that estimates of progressivity are comparable across country since they are computed with the exact same method and similar data sources.

Proxy for χ. The parameter χ directly determines the government expenditure share of output, $g = G/Y$, equal to $\chi/\,(1+\chi)$ in the model, so use this variable as a proxy for χ. Data on government consumption as a fraction of GDP come from the Penn World Tables version 8.1 (Feenstra, Inklaar, and Timmer, 2015). We use the share of government consumption at current PPPs (variable name *csh_g*). From the PWT, we also obtain country GDP and population, which we use to construct our weights in the regressions.

Proxy for θ. From the model, we know that θ is the Pareto-Lorenz coefficient of the income distribution. The World Wealth and Income Database (Alvaredo et al, 2016) provides estimates of this parameter for a number of countries.

Proxy for variances of earnings. In our model, the variance of labor income risk $(v_\alpha, v_\varepsilon)$ and of preference heterogeneity v_φ determine income inequality, over and above the role played by θ. Measures of cross-sectional gross income inequality are therefore good proxies for the size of these variances. The most comprehensive cross-country dataset on income inequality is the Standardized World Income Inequality Database (SWIID) (Solt, 2016) which contains estimates of Gini coefficients for equivalized household market income for 173 countries going back, for some countries like the United States, all the way to 1960. The data are available at http://fsolt.org/swiid/. Each country/year observation has 100 different estimates for the Gini that reflect the degree of uncertainty in estimates and imputation procedures: from these we construct an average and use the average as our estimate of Gini for that country/year pair.

The income distribution in our model is Pareto-LogNormal, i.e. it is a lognormal distribution with a Pareto upper tail. Griffiths and Hajargasht (2013) provide a closed form expression for the Gini coefficient of such distribution $\left(G^{P-LN}\right)$ as a function of only two parameters, the Pareto coefficient (θ in the model) and the variance of the lognormal distribution ($\bar{v} = v_\alpha + v_\varphi + \left(\frac{1+\hat{\sigma}}{\hat{\sigma}}\right)^2 v_\varepsilon$ in the model):

$$G^{P-LN} = \frac{2\exp\left(\theta\left(\theta-1\right)\bar{v}\right)}{2\theta-1}\Phi\left(\left(1-2\theta\right)\sqrt{\frac{\bar{v}}{2}}\right) + 2\Phi\left(\sqrt{\frac{\bar{v}}{2}}\right) - 1$$

where Φ is the cdf of the lognormal distribution. Thus, given our country-specific estimates for the Gini and for θ, we can recover $\bar{v}$.

In our main analysis, we start the sample in 1990 for two reasons: (i) in the earlier years many countries have missing data, and (ii) we can include the countries in the Eastern block in all our analysis. The last year of the sample is 2005. Once we merge all the information together, our final

TABLE C1 Empirical determinants of progressivity across countries

	(1)	(2)	(3)	(4)	(5)	(6)
	τ	τ	τ	τ	τ	τ
G/Y	-0.0978***	-0.326***	-0.0970***	-0.443***	0.0703***	-0.214**
	(-4.19)	(-4.24)	(-4.01)	(-5.18)	(3.39)	(-2.83)
Income Gini	0.0861**		0.132***		0.0900***	
	(3.30)		(5.26)		(3.68)	
θ		0.285***		0.251***		0.0274
		(8.62)		(8.18)		(0.80)
θ^2		-0.0551***		-0.0455***		-0.00340
		(-8.79)		(-8.37)		(-0.53)
$\bar{\nu}$		0.0328**		-0.0117		0.0195
		(2.61)		(-0.71)		(1.45)
Regional Dummies	N	N	Y	Y	Y	Y
Development Dummies	N	N	N	N	Y	Y
N	1585	351	1585	351	1585	351
adj. R^2	0.018	0.285	0.377	0.500	0.574	0.644

t statistics in parentheses: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

dataset comprises of 1585 country-years observations in its largest configuration that includes data on progressivity, inequality and government consumption share of output. The number of countries grows from 65 in 1990 to over 103 in the latest years, and 135 countries are present for at least one year. The median number of observations per country is 14. The dataset includes 351 country-year observations in its smallest configuration that also includes data on Pareto coefficients, and thus exogenous variances of income $\bar{v}$, i.e. 26 countries per year, most of which are present for all 16 years.

C.2. Sensitivity

In Table C1 we report the counterpart of Table V where, instead of the marginal rate of progressivity, we use the measure of τ in equation (C1) as our dependent variable. As clear from a comparison of the two tables, results are very robust.

C.3. Theoretical counterparts to the empirical regression coefficients

We now describe how we compute the theoretical analogues of the regression coefficients reported in Table V. We start with our baseline expression for social welfare, eq. 30. Recall that for each country-year in the data we observe government purchases as a share of output, g_t^c, the variance of the normal component of earnings, $\bar{v}_t^c$, and the Pareto coefficient, θ_t^c. The theoretical expression for social welfare instead involves the structural parameters $(\chi, v_\alpha, v_\varphi, v_\varepsilon, \theta)$. We use the optimality condition for government purchases, $g = \chi/(1+\chi)$ to translate the sensitivity of optimal progressivity with respect to χ to the implied sensitivity with respect to g_t^c. Since we do not have country-specific empirical counterparts for v_α, v_φ or v_ε we will assume that the ratios of these variances to $\bar{v}_t^c$ are common across countries, and equal to the ratios implied by our calibration to the United States. Let δ_α, δ_φ and δ_ε denote these ratios, where, for example, $\delta_\alpha = v_\alpha/(v_\alpha + v_\varphi + ((1+\hat{\sigma})^2/\hat{\sigma}^2)v_\varepsilon$. Given $(v_\alpha)_t^c = \delta_a \bar{v}_t^c$, $(v_\varphi)_t^c = \delta_\varphi \bar{v}_t^c$, $(v_\varepsilon)_t^c = \delta_\varepsilon \bar{v}_t^c$, we can compute the sensitivity of optimal progressivity with respect to $\bar{v}_t^c$.

Differentiating eq. 30 with respect to τ gives the condition that implicitly defines optimal progressivity, as a function of structural parameters:

$$F((1-\tau), \chi, \theta, \bar{v}, \sigma, \psi) = 0$$

where

$$\begin{aligned} F((1-\tau), \chi, \theta, \bar{v}, \sigma, \psi) \;\; = \;\; & \frac{-(1+\chi)}{(1+\sigma)(1-\tau)} + \frac{1}{1+\sigma} - (1+\chi)\frac{\psi}{(1+\psi)(\theta-1)}\frac{1}{(1-\tau)} + \frac{\psi}{(1+\psi)\theta} \\ & + \frac{\frac{1}{\theta}}{1-\left(\frac{1-\tau}{\theta}\right)} - \frac{1}{\theta} \\ & + \left((1-\tau)(\delta_\alpha + \delta_\varphi) - \frac{(1-(1-\tau))(\sigma+1)^2(1+\chi)}{((\sigma+1)-(1-\tau))^3}\delta_\varepsilon \right)\bar{v} \end{aligned}$$

Using the implicit function theorem, we can compute the theoretical sensitivity of progressivity τ with respect to the variance of the normally distributed component of earnings $\bar{v}$:

$$\begin{aligned} \frac{d\tau}{d\bar{v}} \;\; &= \;\; -\frac{d(1-\tau)}{d\bar{v}} \\ &= \;\; \frac{(1-\tau)(\delta_\alpha + \delta_\varphi) - \tau(\sigma+1)^2 \frac{(1+\chi)}{(\sigma+\tau)^3}\delta_\varepsilon}{\frac{(1+\chi)}{(1+\sigma)(1-\tau)^2} + (1+\chi)\frac{\psi}{(1+\psi)(\theta-1)}\frac{1}{(1-\tau)^2} + \frac{1}{(\theta-(1-\tau))^2} + \left((\delta_\alpha + \delta_\varphi) + \frac{(-2\tau+\sigma)}{(\sigma+\tau)^4}(1+\sigma)^2(1+\chi)\delta_\varepsilon \right)\bar{v}}. \end{aligned}$$

Evaluated at our calibration to the United States, and at the value for τ that is optimal given that calibration (0.084), this derivative is equal to 0.175.

Now consider the sensitivity of progressivity to government purchases. By the Chain Rule,

$$\frac{d\tau}{dg} = \frac{d\tau}{d\chi}\frac{d\chi}{dg} = \frac{d\tau}{d\chi}\frac{1}{(1-g)^2}.$$

Using the implicit function theorem to compute $\frac{d(1-\tau)}{d\chi}$ gives

$$\begin{aligned}
\frac{d\tau}{dg} &= -\frac{d(1-\tau)}{d\chi}\frac{1}{(1-g)^2} \\
&= \frac{1}{(1-g)^2}\frac{-\frac{1}{(1+\sigma)(1-\tau)}-\frac{\psi}{(1+\psi)(\theta-1)}\frac{1}{(1-\tau)}-\tau\frac{(\sigma+1)^2}{(\sigma+\tau)^3}v\delta_\varepsilon}{\frac{1+\chi}{(1+\sigma)(1-\tau)^2}+\frac{(1+\chi)\psi}{(1+\psi)(\theta-1)}\frac{1}{(1-\tau)^2}+\frac{1}{(\theta-(1-\tau))^2}+\left((\delta_\alpha+\delta_\varphi)+\frac{(-2\tau+\sigma)}{(\sigma+\tau)^4}(1+\sigma)^2(1+\chi)\delta_\varepsilon\right)\bar{v}}
\end{aligned}$$

Evaluating this expression at our baseline parameter values and at $g = 0.189$ and $\tau = 0.084$, we find $d\tau/dg = -0.690$.

APPENDIX D: ADDITIONAL MATERIAL

D.1. CES skill aggregator in our model versus Benabou (2005)

Our paper studies optimal taxation within the class of log-linear tax functions in eq. (1) in the main text, assuming a production technology with a constant elasticity of substitution between different skill inputs. As discussed in the main text, Benabou (2002, 2005) studies optimal taxation within the same class of tax functions and he too assumes a CES production function over various skill inputs. However, as it turns out, the link between the optimal tax progressivity parameter τ and the elasticity of substitution between output from different skill types, θ, is quite different in the two models. In our model, the optimal progressivity τ is hump-shaped in θ: rising in θ for low levels of θ and falling in θ for large θ (see Figure 2 of the main text). In contrast, in Benabou (2002, 2005), the optimal progressivity falls monotonically with the elasticity of substitution θ. We now explain why the link between θ and τ in Benabou (2002, 2005) is so different from the link in our paper.

Consider the formulation in Benabou (2005). Section 3.2 of his paper develops an extension of the baseline model to a case in which skills are aggregated in a CES fashion to produce a final good. Equation (30) in his paper describes this technology (in equilibrium, with constant labor supply). Note that the distribution of agents across what he labels "skills", $g(i)$, is uniform by assumption. Re-expressed in our notation, his technology is

$$\frac{Y}{N} = \left(\int_0^1 [k(i) \cdot g(i)]^{\frac{\sigma-1}{\sigma}} \, di \right)^{\frac{\sigma}{\sigma-1}},$$

where $g(i) = 1$ and $k(i)$ is human capital of individual i.

Contrast this to our baseline production technology (our eq.5), which is

$$Y = \left(\int_0^\infty [N(s) \cdot m(s)]^{\frac{\theta-1}{\theta}} \, ds \right)^{\frac{\theta}{\theta-1}}$$

where $m(s)$ is (in equilibrium) an exponential distribution. In our model skill is a choice and due to the cost of choosing a higher skill, the equilibrium distribution of skills $m(s)$ has density falling with s.

These two technologies may appear to be very similar, but the way the two models work is quite different. In Benabou's model, the distribution $g(i)$ is exogenous and uniform, and progressivity effects the distribution of $k(i)$. Given a uniform distribution $g(i)$, productivity is maximized when human capital is evenly distributed, which calls for progressive taxation in his model. Moreover, this force is larger the smaller is θ.

In our model, in contrast, effective hours $N(s)$ (our formal analogue to Benabou's human capital) are, in equilibrium, equal across skill types. The way in which the tax system impacts the distribution of effective hours by different skill types is by affecting the shape of the distribution $m(s)$. Now the more progressive is the tax system, the more clustered towards zero is the distribution $m(s)$, since agents do less skill investment, and high skill types are more scarce. Because a uniform distribution for $m(s)$ maximizes productivity, progressivity is therefore productivity-reducing, and this force is larger the smaller is θ.

D.2. Comparison of actual and optimal tax systems

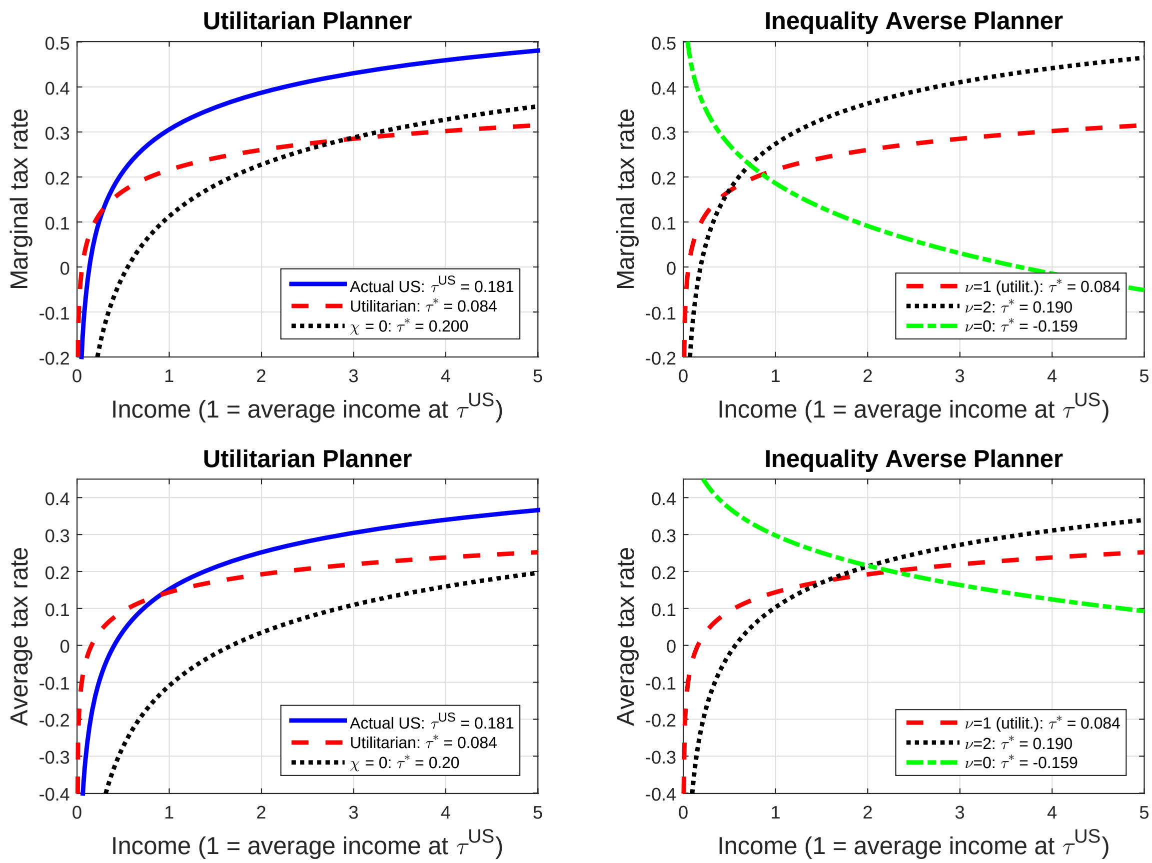

FIGURE D2 Left panels: the US tax system compared to the utilitarian optimum with valued government expenditures, and to the utilitarian optimum when government expenditures are not valued $(\chi = 0)$. Right panels: Utilitarian optimum compared to the optimum for a planner with more inequality aversion that the utilitarian planner $(\nu = 2)$ and for an inequality-neutral planner $(\nu = 0)$.

D.3. Comparison between fixed and flexible skill investment models

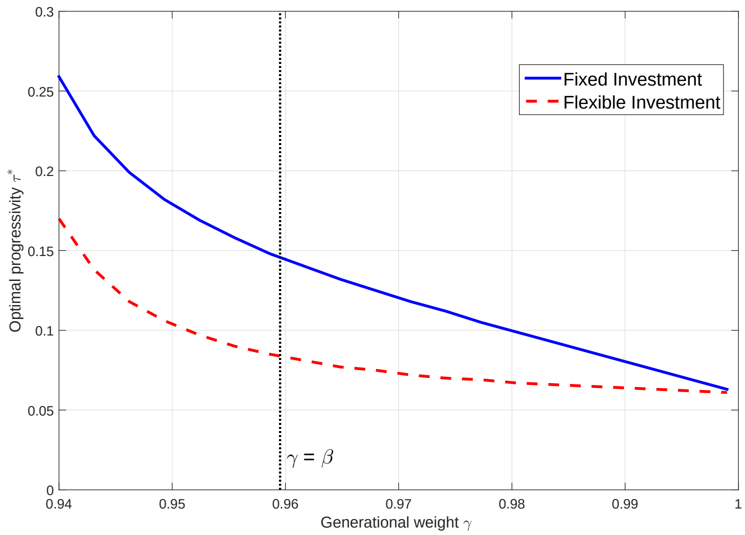

FIGURE D3 Optimal degree of progressivity τ^* as a function of the generational weight γ in the fixed investment model and the baseline flexible investment model.

References

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

[2] Benabou, R (2005) *Inequality, Technology, and the Social Contract*, in Handbook of Economic Growth, P. Aghion and S. Durlauf eds., North-Holland: Vol 1B, Chapter 25, 1595-1638.

[3] Feenberg, D., and E. Coutts (1993): "An Introduction to the TAXSIM Model," *Journal of Policy Analysis and Management*, 12(1), 189-194.

[4] Heathcote, J., F. Perri, and G. L. Violante (2010): "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967-2006," *Review of Economic Dynamics*, 13(1), 15-51.

[5] Heathcote, J., K. Storesletten, and G. L. Violante (2014): "Consumption and Labor Supply with Partial Insurance: An Analytical Framework," *American Economic Review*, 104(7), 2075-2126.

[6] Jung, J., and C. Tran (2013): "Medical Consumption over the Life Cycle: Facts from a US Medical Expenditure Panel Survey," UNSW Australian School of Business Research Paper.

[7] List, J. A. (2011): "The Market for Charitable Giving," *Journal of Economic Perspectives*, 25(2), 157-180.

[8] Hajargasht, G., and W.E. Griffiths (2013): "Pareto–Lognormal distributions: Inequality, Poverty, and Estimation from Grouped Income Data,"*Economic Modelling*, 33, 593–604.