OPTIMAL PROGRESSIVITY WITH AGE-DEPENDENT TAXATION

Jonathan Heathcote
Kjetil Storesletten
Giovanni L. Violante

Optimal Progressivity with Age-Dependent Taxation
Jonathan Heathcote, Kjetil Storesletten, and Giovanni L. Violante
NBER Working Paper No. 25617
February 2019
JEL No. E20,H2,H21,H31,H41

ABSTRACT

This paper studies optimal taxation of earnings when the degree of tax progressivity is allowed to vary with age. The setting is an overlapping-generations model that incorporates irreversible skill investment, flexible labor supply, ex-ante heterogeneity in the disutility of work and the cost of skill acquisition, partially insurable wage risk, and a life cycle productivity profile. An analytically tractable version of the model without intertemporal trade is used to characterize and quantify the salient trade-offs in tax design. The key results are that progressivity should be U-shaped in age and that the average marginal tax rate should be increasing and concave in age. These findings are confirmed in a version of the model with borrowing and saving that we solve numerically.

Jonathan Heathcote
Federal Reserve Bank of Minneapolis
Research Department
90 Hennepin Ave.
Minneapolis, MN 55401
heathcote@minneapolisfed.org

Kjetil Storesletten
Department of Economics
University of Oslo
P.O. Box 1095 Blindern
0317 Oslo
Norway
kjetil.storesletten@econ.uio.no

Giovanni L. Violante
Department of Economics
Princeton University
Julis Romo Rabinowitz Building
Princeton, NJ 08540
and NBER
glv2@princeton.edu

1 Introduction

A central problem in public finance is to design a tax and transfer system to pay for public goods and provide insurance to unfortunate individuals while minimally distorting labor supply and investments in physical and human capital. One potentially important tool for mitigating tax distortions is "tagging": letting tax rates depend on observable, immutable, or hard-to-modify personal characteristics. This idea was proposed first by Akerlof (1978) and has recently gained new attention in the policy debate (see, for example, Banks and Diamond, 2010). Age is one such characteristic.

The purpose of this paper is to study optimal labor income taxation in a setting in which the parameters of the tax system are allowed to vary with age. We do not study fully optimal Mirrleesian tax system design. Rather, we restrict attention to the parametric class of income tax and transfer systems given by

$$T\left(y\right) = y - \lambda y^{1-\tau}, \tag{1}$$

where y is gross income and $T(y)$ is taxes net of transfers. The parameter τ controls the progressivity of the tax system, with $\tau = 0$ corresponding to a flat tax rate and $\tau > 0$ ($\tau < 0$) implying a progressive (regressive) tax and transfer system. Conditional on τ, the parameter λ controls the level of taxation. This class of tax systems has a long tradition in public finance; see, for example, Musgrave (1959), Jakobsson (1976), Kakwani (1977) and, more recently, Bénabou (2000, 2002) and Heathcote, Storesletten, and Violante (2017).

The key innovation in the present paper is to let the parameters λ and τ in eq. (1) be conditioned on age, so that both the level and the progressivity of the tax schedule can be made age dependent.

In Heathcote et al. (2017), we document that the parametric class in eq. (1) provides a remarkably good approximation to the actual tax and transfer scheme in the US for households aged 25-60. In particular, eq. (1) implies that after-tax earnings $y - T(y)$ should be a log-linear function of pre-tax earnings y. Using data from the Panel Study of Income Dynamics (PSID), Heathcote et al. (2017) show that a linear regression of the logarithm of post-government earnings on the logarithm of pre-government earnings yields a very good fit, with an R^2 of 0.93: when plotting average post-government against pre-government earnings for each percentile of the sample, the relationship is virtually log-linear.

In that paper, we did not investigate whether the current tax/transfer system *de facto* features elements of age dependence in progressivity. For example, one may think that certain transfers (e.g., UI benefits, child benefits) and certain provisions (e.g., mortgage interest and medical
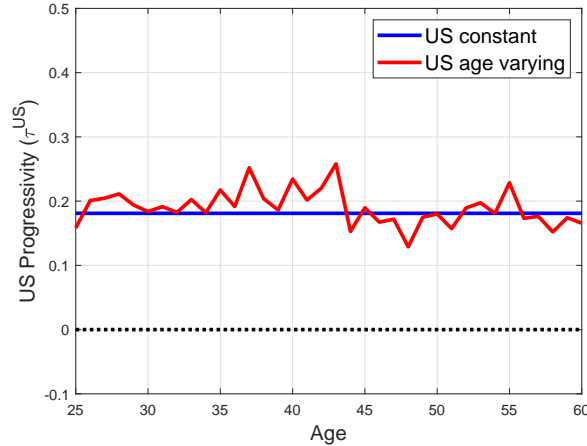
Figure 1: The coefficient τ estimated from a regression of log disposable income $y - T(y)$ on log gross income y, where intercept and slope are both allowed to vary with age. The straight line is the estimated $\tau^{US} = 0.181$ when age dependence is not allowed in the regression. See Heathcote et al. (2017) for details on the 2000-2006 PSID data used in this estimation, and on the construction of y and $T(y)$ at the household level.

expenditure deductibility) would effectively induce some age dependence. We have therefore repeated our previous estimation, allowing the intercept and slope parameters to both depend on age. Figure 1 plots the estimated τ for each age group together with the estimated age-invariant $\tau^{US} = 0.181$. The main finding is that there is no significant age dependence in progressivity embedded in the current US system.

The aim of this paper is to explore whether there is scope for improving the current US tax and transfer system by introducing explicit age dependence. Our environment, which closely follows Heathcote et al. (2017), is an overlapping-generations model in which individuals care about consumption, leisure, and a publicly provided good. Individuals make an irreversible skill investment when young and make a labor-leisure choice in each period of working life. People differ *ex ante* in their learning ability and in their willingness to work. Those with higher learning ability invest in higher skills, and those with a lower utility cost of effort work more hours. Skills are imperfect substitutes, and the price of skills is an equilibrium outcome. Deterministic life-cycle profiles for labor productivity and for the disutility of work generate systematic age variation in average wages, hours, and consumption. During working life, individuals face idiosyncratic shocks to their productivity that can only be partially insured privately. The uninsurable (and permanent) component of these wage shocks passes through to consumption, generating a rising age profile for within-cohort consumption inequality, as in the data.

Tax progressivity compresses *ex post* dispersion in consumption. Thus, the social insurance embedded in the tax and transfer system partially offsets inequality in initial conditions and

also provides a substitute for missing private insurance against life-cycle shocks. However, tax progressivity discourages labor supply and skill investment. Because the skill choice is determined by the after-tax return, the tax system affects the equilibrium skill distribution and, therefore, pre-tax skill prices.

Most of our analysis focuses on a version of the model in which there are no markets for intertemporal borrowing and lending. In this environment, we are able to derive a closed-form solution for an equally-weighted steady-state social welfare function, which we use to build intuition about the drivers of optimal age variation in tax progressivity. Toward the end of the paper, we extend the analysis to allow for life-cycle borrowing and lending. In this case, we must solve for equilibrium allocations numerically, but the optimal policy turns out to be quite similar.

A first result is that, for any age profile for τ, the optimal age profile for the tax level parameter λ (which controls the average level of taxation) equates average consumption by age. This convenient separation between the roles of λ and τ arises because under our balanced-growth-consistent utility specification, λ has no impact on either skill investment or labor supply.

The shape of the optimal age profile for the tax progressivity parameter τ trades off two key forces.

First, age is informative about the *dispersion* of productivity. Dispersion in productivity is increasing with age because individuals face permanent idiosyncratic shocks that cumulate over the life cycle. To the extent that these shocks are privately uninsurable, they will translate into increasing consumption dispersion with age. The planner has an incentive to target redistribution where inequality is concentrated, namely among the old. This *uninsurable risk channel* is a force for having *progressivity increase with age*.

Second, age is informative about average earnings, since wages net of the disutility of work are increasing during the first decades of working life. Given a generally progressive tax system, an age-increasing earnings profile will imply increasing marginal tax rates. In order to smooth marginal tax rates over the life cycle, the planner has an incentive to have *progressivity decline with age*. This force for declining progressivity is amplified by the result that average tax rates optimally rise with age (via a declining profile for λ) in order to smooth average consumption over the life cycle. Having progressivity decline with age allows the planner to smooth marginal tax rates even as average tax rates rise. We call this mechanism the *life cycle channel*.

Our quantitative analysis, with the model calibrated to the US economy, implies that, on their own, life-cycle accumulation of uninsurable risk and life-cycle variation in productivity each call for significant variation in tax progressivity over the life cycle. When both factors are combined, the two effects roughly offset, implying an optimal profile for progressivity τ that is

3

U-shaped in age.

Because skill investment is irreversible, a tax reform induces a transition. In the economy without borrowing and lending we are able to compute the full transitional dynamics for the Ramsey planner who can vary tax parameters freely by both time and age. Here, the planner has an incentive to set a high value for progressivity for existing cohorts who have already made their skill investment decisions, while keeping progressivity low for new skill-investment-elastic generations. Throughout the transition, the average level of progressivity changes, but the age profile within each cohort remains U-shaped.

In this benchmark economy without intertemporal trade, part of the gains from age-dependent taxation accrue because the planner lets the average tax rate increase with age in order to redistribute from the (more productive) old to the (less productive) young. If households could smooth consumption independently via borrowing and lending, this rationale for an age-varying tax system would presumably be weakened. How would this change the optimal age profile for progressivity?

To answer this question, we extend our model to allow households to trade a bond in zero net supply. We then solve numerically for allocations and for the optimal age-dependent tax system under various borrowing limits. Our simulations confirm the intuition that, with very loose liquidity constraints, the life cycle channel vanishes. However, when we calibrate the value for the borrowing limit based on data from the Survey of Consumer Finances, the optimal age profile for τ is quite close to the one for the baseline economy without intertemporal trade. The welfare gains of moving from the current age-invariant tax system to the optimal age-varying one are now around 2% of lifetime consumption.

Finally, we note that this U shape in the age profile for optimal progressivity becomes flatter in two empirically relevant cases. First, when the labor supply elasticity rises with age as workers approach retirement (in the spirit of Ndiaye, 2017). Second, when part of the hump-shaped age profile for household consumption in the data is explained by changing demographics. In this case, a portion of consumption variation by age is efficient, weakening the motive for redistribution across age groups.

Related Literature. We are not the first to study motives for age dependence in the optimal design of tax schedules. Several antecedents of ours follow the Ramsey tradition. Erosa and Gervais (2002) analyze optimal taxation in a life-cycle economy without any sources of within-cohort heterogeneity (i.e., all inequality is between age groups). They focus on models in which the age dependence in average tax rates is driven by the fact that the Frisch elasticity of labor supply varies over the life cycle. This channel depends on preference specifications. In our

baseline model, we have abstracted from this channel by choosing a specification in which the Frisch elasticity is constant, but in an extension we allow the Frisch to vary with age. Conesa et al. (2009) study optimal taxation within a Gouveia-Strauss class of non-linear tax functions. While more flexible than ours, this class of functions is less analytically tractable. They do not explicitly model age dependence, but they point out that a positive tax on capital income can stand in for age-dependent taxes because the age profile of wealth is correlated with that of productivity. Karabarbounis (2017) explores optimal age-varying taxation numerically using the same functional form for the net tax and transfer system as we do. However, he restricts attention to optimal age variation in the λ parameter – which controls the level of taxes – while assuming an age-invariant value for the progressivity parameter τ. We find additional welfare gains from allowing both parameters to depend on age.

A more recent literature studies the role of age variation in the Mirrleesian optimal taxation framework. Three papers are especially related to our work. The first is by Weinzierl (2011), who focuses on the rising age profile of wages, and on how these profiles differ across skill groups. His key finding, namely that optimal average and marginal tax rates are both rising with age, is qualitatively similar to ours when the only operational channel is life-cycle productivity. The second related paper is Farhi and Werning (2013), who analyze taxation in a dynamic life-cycle economy. They focus on the role of persistent productivity shocks. In their numerical example, the fully optimal history-dependent tax schedule displays the same qualitative features as our model when our risk channel is the only one operative: average wedges increase with age, average labor earnings are falling with age, and average consumption is constant. These findings are mirrored in the work of Golosov et al. (2016), who focus on the additional effect of skewness of wage shocks. Ndiaye (2017) extends Farhi and Werning to allow for a discrete retirement choice, which reduces optimal marginal tax rates around the age of retirement when labor supply is relatively elastic.

More recently, the Mirrleesian strand of the optimal tax literature has begun incorporating endogenous human capital accumulation into the optimal design problem.[1] Most closely related to ours are the papers by Best and Kleven (2013) and Stantcheva (2017). Best and Kleven (2013) extend the canonical Mirrleesian framework to incorporate endogenous on-the-job learning in a simple two-period model where working more hours increases productivity throughout one's career. This mechanism makes the labor supply elasticity lower for the young (whose return to work also accrues in the future) and offers an argument for decreasing marginal taxes with age. In our paper, we abstract from learning by working and highlight the role of skill acquisition before entry in the labor market.

---

[1]See, for example, Kapička (2015) and Findeisen and Sachs (2016).

Stantcheva (2017) studies optimal Mirrleesian taxation over the life cycle in a model in which individuals can endogenously accumulate human capital by spending on education. Her model and analysis differs from ours in several respects. First, she studies the role of human capital in increasing or reducing wage risk, while risk is independent of skill in our model. Second, she shows that to implement the constrained efficient allocation one needs policy tools that directly target the skill investment margin, such as education subsidies or income-contingent loans, while we focus exclusively on the design of the tax/transfer system.

Interestingly, recent contributions in this literature have demonstrated that indexing tax rates by age can capture most of the potential welfare gains from fully optimal, history-dependent policies (e.g., Farhi and Werning 2013; Golosov et al., 2016; Stantcheva, 2017; and Weinzierl, 2011).

With respect to this existing set of results, our contribution is threefold. First, we offer a closed-form expression for social welfare as a function of the vector $\{\tau_a\}$ and the structural parameters of the model describing preferences, technology, *ex ante* heterogeneity, and *ex post* uncertainty. Each term in our welfare expression has an economic interpretation and embodies one of the channels shaping the optimal progressivity trade-off discussed above. Second, we find that the life-cycle channel is quantitatively most important in the first half of the working life, when average wages are rising fast, while the uninsurable risk matters more later in life as permanent shocks cumulate. This distinction explains our novel result that optimal progressivity is U-shaped in age. Third, we identify a new motive for age variation in taxation that hinges on the presence of endogenous and irreversible skill investment and that operates only during the transition.

The paper proceeds as follows. Sections 2 and 3 lay out the economic environment and solve for the competitive equilibrium given a tax policy. Section 4 discusses the social welfare function and Section 5 derives analytical properties of optimal taxes in steady state and during the transition. Section 6 studies the quantitative implications of allowing for age variation in taxes and calculates the welfare gain of introducing such fiscal tools. Section 7 develops the extension of the model with intertemporal trade. Section 8 concludes.

2 Economic environment

Demographics: The model has a standard overlapping-generations structure. Agents enter the economy at age $a = 0$ and live for A periods. The total population is of mass one, and thus each age group is of mass $1/A$. There are no intergenerational links. We index agents by $i \in [0, 1]$. To simplify notation, we will abstract from time subscripts until we explore the transition from

one tax system to another in Section 5.2.

Life cycle: Upon birth, individuals have a chance to invest in skills s_i. Once the individual has chosen s_i, he or she enters the labor market. The individual provides $h_i \geq 0$ hours of labor supply, consumes a private good c_i, and enjoys a publicly provided good G.[2] Each period he or she faces stochastic fluctuations in labor productivity z_i.

Preferences: Expected lifetime utility over private consumption, hours worked, publicly provided goods, and skill investment effort for individual i is given by

$$U_i = -v_i(s_i) + \mathbb{E}_0 \left(\frac{1-\beta}{1-\beta^A} \right) \sum_{a=0}^{A-1} \beta^a u_i(c_{ia}, h_{ia}, G), \tag{2}$$

where $\beta \leq 1$ is the discount factor, common to all individuals, and the expectation is taken over future idiosyncratic productivity shocks, whose process is described below. The disutility of the initial skill investment $s_i \geq 0$ takes the form

$$v_i(s_i) = \frac{(\kappa_i)^{-1/\psi}}{1+1/\psi} (s_i)^{1+1/\psi}, \tag{3}$$

where the parameter $\psi \geq 0$ controls the elasticity of skill investment with respect to the marginal return to skill, and $\kappa_i \geq 0$ is an individual-specific parameter that determines the utility cost of acquiring skills. The larger is κ_i, the smaller is the cost, so one can think of κ_i as indexing innate learning ability. We assume that $\kappa_i \sim Exp(\eta)$, an exponential distribution with parameter η. As we demonstrate below, exponentially distributed ability yields Pareto right tails in the equilibrium wage and earnings distributions. Skill investment decisions are irreversible, and thus skills are fixed through the life cycle.[3]

The period utility function u_i is

$$u_i(c_{ia}, h_{ia}, G) = \log c_{ia} - \frac{\exp\left[(1+\sigma)(\bar{\varphi}_a + \varphi_i)\right]}{1+\sigma} (h_{ia})^{1+\sigma} + \chi \log G, \tag{4}$$

where $\exp\left[(1+\sigma)(\varphi_i + \bar{\varphi}_a)\right]$ scales the disutility of work effort. The profile $\{\bar{\varphi}_a\}_{a=0}^{A-1}$ captures the common and deterministic evolution in the disutility of work as individuals age. The parameter φ_i is a fixed individual effect that is normally distributed: $\varphi_i \sim \mathcal{N}\left(\frac{v_\varphi}{2}, v_\varphi\right)$, where v_φ

---

[2]G has two possible interpretations. The first is that it is a pure public good, such as national defense or the judicial system. The second is that it is an excludable good produced by the government, such as public education, that is distributed uniformly across households.

[3]The baseline model in Heathcote et al. (2017) assumes reversible skill investment. Given reversible investment, the skill investment decision is essentially static, whereas in the present model it will be a dynamic decision.

denotes the cross-sectional variance. We assume that κ_i and φ_i are uncorrelated. The parameter $\sigma > 0$ determines aversion to hours fluctuations. Finally, $\chi \geq 0$ measures the taste for the publicly provided good G relative to private consumption.

Technology: Output Y is a constant elasticity of substitution aggregate of effective hours supplied by the continuum of skill types $s \in [0, \infty)$,

$$Y = \left(\int_0^\infty [\bar{N}(s) \cdot m(s)]^{\frac{\theta-1}{\theta}} ds \right)^{\frac{\theta}{\theta-1}}, \tag{5}$$

where $\theta > 1$ is the elasticity of substitution across skill types, $\bar{N}(s)$ denotes average effective hours worked by individuals of skill type s, and $m(s)$ is the density of individuals with skill type s. Note that all skill levels enter symmetrically in the production technology, and thus any equilibrium differences in skill prices will reflect relative scarcity.

Labor productivity and earnings: Log individual labor efficiency z_{ia} is the sum of three orthogonal components, x_a, α_{ia}, and ε_{ia},

$$z_{ia} = x_a + \alpha_{ia} + \varepsilon_{ia}. \tag{6}$$

The first component x_a captures the deterministic age profile of labor productivity, common for all individuals. The second component α_{ia} captures idiosyncratic shocks that cannot be insured privately, and follows the unit root process $\alpha_{ia} = \alpha_{i,a-1} + \omega_{ia}$, with i.i.d. innovation $\omega_{ia} \sim \mathcal{N}\left(-\frac{v_\omega}{2}, v_\omega\right)$ and initial value $\alpha_{i0} = 0$. The third component ε_{ia} captures idiosyncratic shocks that can be insured privately. The only property of the time series process for ε_{ia} that will matter for our welfare expressions and optimal taxation results is the age profile for the cross-sectional variance, $v_{\varepsilon,a}$. For expositional simplicity we will therefore assume, without loss of generality, that shocks to ε are drawn independently over time from a Normal distribution, $\varepsilon_{ia} \sim \mathcal{N}\left(-v_{\varepsilon,a}/2, v_{\varepsilon,a}\right)$, where $v_{\varepsilon,a}$ captures the variance at age a.

A standard law of large numbers ensures that none of the individual-level shocks induce any aggregate uncertainty in the economy.

Individual earnings y_{ia} are, therefore, the product of four components:

$$y_{ia} = \underbrace{p(s_i)}_{\text{skill price}} \times \underbrace{\exp(x_a)}_{\text{age-productivity profile}} \times \underbrace{\exp(\alpha_{ia} + \varepsilon_{ia})}_{\text{labor market shocks}} \times \underbrace{h_{ia}}_{\text{hours}}. \tag{7}$$

The first component $p(s_i)$ is the equilibrium price for the type of labor supplied by an individual with skills s_i; the second component is the life-cycle profile of average labor efficiency; the third component is individual stochastic labor efficiency; and the fourth component is the number of hours worked by the individual. Thus, individual earnings are determined by (i)

8

skill investments made before labor market entry, in turn reflecting innate learning ability κ_i; (ii) productivity that grows exogenously with experience; (iii) fortune in labor market outcomes determined by the realization of idiosyncratic efficiency shocks; and (iv) work effort, reflecting, in part, innate and age-varying taste for leisure, defined by φ_i and $\bar{\varphi}_a$. Taxation affects the equilibrium pre-tax earnings distribution by changing skill investment choices, and thus skill prices, and by changing labor supply decisions.

Financial assets: We adopt a simplified version of the partial-insurance structure developed in Heathcote et al. (2014a). There is a full set of state-contingent claims for each realization of the ε shock, implying that variation in ε is fully insurable. These claims are traded within the period. Let $B_{ia}(\mathbf{E})$ and $Q(\mathbf{E})$ denote the quantity and the price, respectively, of insurance claims purchased that pay one unit of consumption if and only if $\varepsilon \in \mathbf{E} \subseteq \mathbb{R}$.[4] In Section 7 we introduce borrowing and lending, solve for the equilibrium numerically, and explore how this alternative market structure changes optimal tax policy.[5]

Labor and goods markets: The final consumption good and all types of labor services are traded in competitive markets. The final good is the numéraire of the economy.

Government: The government runs a tax and transfer scheme and provides each household with an amount of goods or services equal to G. This public good can only be provided by the government, which transforms final goods into G one for one. Let g denote government expenditures as a fraction of aggregate output (i.e., $G = gY$).

Let $T_a(y)$ be the net tax owed at income level y by age group a. We study optimal policies within the class of tax and transfer schemes defined by the function

$$T_a(y) = y - \lambda_a y^{1-\tau_a}, \tag{8}$$

where the parameters τ_a and λ_a are specific to age group a. The specification of eq. (8) with age-invariant parameters has a long tradition in public finance dating back to Feldstein, 1969. Recently, Bénabou, 2000 and 2002 and Heathcote et al., 2014 and 2017 demonstrated its tractability in the context of equilibrium macroeconomic models. Heathcote and Tsujiyama (2016) have shown that in a static environment, this functional form can closely approximate the fully optimal Mirrleesian policy.

---

[4]An alternative way to decentralize insurance with respect to ε is to assume that individuals belong to large families and that shocks to α are common across members of a given family, while shocks to ε are purely idiosyncratic and thus can be pooled within the family.

[5]In Heathcote et al. (2014), we allowed agents to trade a single non-contingent bond and showed that there is an equilibrium in which this bond is not traded, given that idiosyncratic wage shocks follow a unit root process. This result does not generalize to the present model because age variation in efficiency and disutility (x_a, $\bar{\varphi}_a$) and in the tax parameters τ_a and λ_a introduces motives for intertemporal borrowing and lending.

The parameter τ_a determines the degree of progressivity of the tax system and is the key object of interest in our analysis. There are two ways to see why τ_a is a natural index of progressivity. First, eq. (8) implies the following mapping between individual disposable (post-government) earnings $\tilde{y}$ and pre-government earnings y:

$$\tilde{y} = \lambda_a y^{1-\tau_a}. \tag{9}$$

Thus, $(1 - \tau_a)$ measures the elasticity of disposable to pre-tax income. Second, a tax scheme is commonly labeled progressive (regressive) if the ratio of marginal to average tax rates is larger (smaller) than one for every level of income y. Within our class, we have

$$\frac{1 - T_a'(y)}{1 - T_a(y)/y} = 1 - \tau_a. \tag{10}$$

When $\tau_a > 0$, marginal rates always exceed average rates, and the tax system is therefore progressive. Conversely, when $\tau_a < 0$, the tax system is regressive. The case $\tau_a = 0$ implies that marginal and average tax rates are equal: the system is a flat tax with rate $1 - \lambda_a$.

Given τ_a, the second parameter, λ_a, shifts the tax function and determines the average level of taxation in the economy. At the break-even income level $y_a^0 = (\lambda_a)^{\frac{1}{\tau_a}} > 0$, the average tax rate is zero and the marginal tax rate is τ_a for that age group. If the system is progressive (regressive), then at every income level below (above) y_a^0, the average tax rate is negative and households obtain a net transfer from the government. Thus, this function is best seen as a *tax and transfer* schedule, a property that has implications for the empirical measurement of τ_a. The income-weighted average marginal tax rate (MTR) at age a given this tax and transfer schedule is

$$MTR_a = 1 - \lambda_a(1 - \tau_a)\frac{\int (y_{ia})^{1-\tau_a} di}{\int y_{ia} di}. \tag{11}$$

The government must run a balanced budget, and the government budget constraint is therefore

$$g \sum_{a=0}^{A-1} \int y_{ia} di = \sum_{a=0}^{A-1} \int \left[y_{ia} - \lambda_a (y_{ia})^{1-\tau_a} \right] di. \tag{12}$$

The government chooses g and the sequences $\{\tau_a, \lambda_a\}_{a=0}^{A-1}$, with one instrument being determined residually by eq. (12). Since the budget constraint holds at the aggregate level (not at the level of each age group), the government can redistribute both within and between age groups.

The rate of transformation between private and public consumption is one, and thus the aggregate resource constraint for the economy (recall population has measure 1 so aggregates

10

equal averages) is

$$Y = G + \frac{1}{A} \sum_{a=0}^{A-1} \int_0^1 c_{ia} \, di. \tag{13}$$

2.1 Individual problem

At age $a = 0$, the individual chooses a skill level, given her idiosyncratic draw (κ_i, φ_i). Combining eqs. (2) and (3), the first-order necessary and sufficient condition for the skill choice is

$$\frac{\partial v_i(s_i)}{\partial s_i} = \left(\frac{s_i}{\kappa_i}\right)^{\frac{1}{\psi}} = \left(\frac{1-\beta}{1-\beta^A}\right) \mathbb{E}_0 \sum_{a=0}^{A-1} \beta^a \frac{\partial u_i(c_{ia}, h_{ia}, G)}{\partial s_i}. \tag{14}$$

Thus, the marginal disutility of skill investment for an individual with learning ability κ_i must equal the discounted present value of the corresponding expected benefits in the form of higher lifetime wages. Recall that initial skill investments are irreversible, and thus older agents cannot supplement or unwind past skill investments.

At the beginning of every period of working life a, the innovation ω_{ia} to the random walk shock α_{ia} is realized. Then, the insurance markets against the ε_{ia} shocks open and the individual buys insurance claims $B(\cdot)$. Finally, ε_{ia} is realized, insurance claims pay out, and the individual chooses hours h_{ia}, receives wage payments, and chooses consumption expenditures c_{ia}. Thus, the individual budget constraint in the middle of the period, when the insurance purchases are made, is

$$\int_{\mathbf{E}} Q(\varepsilon) B_{ia}(\varepsilon) \, d\varepsilon = 0, \tag{15}$$

and the budget constraint at the end of the period, after the realization of ε_{ia}, is

$$c_{ia} = \lambda_a \left[p(s_i) \exp(x_a + \alpha_{ia} + \varepsilon_{ia}) h_{ia} \right]^{1-\tau_a} + B(\varepsilon_{ia}). \tag{16}$$

Given an initial skill choice s_i, the problem for an agent is to choose insurance purchases, consumption, and hours worked in order to maximize lifetime utility (2) subject to sequences of budget constraints (15)-(16), taking as given the process for efficiency units described in eq. (6). In addition, agents face non-negativity constraints on consumption and hours worked.

3 Equilibrium

We now adopt a recursive formulation to define a stationary competitive equilibrium for our economy. The state vector for the skill accumulation decision at age $a = 0$ is just the pair of fixed

individual effects (κ, φ). At subsequent ages, the state vector for the beginning-of-the-period decision when insurance claims are purchased is (φ, s, a, α). The individual state vector for the end-of-period consumption and labor supply decisions is $(\varphi, s, a, \alpha, \varepsilon)$.[6] Note that age is a state variable for two reasons: (i) labor productivity and the disutility of work vary with age, and (ii) the parameters of the tax system potentially vary with age. What makes the model tractable, in spite of all the heterogeneity and risk it features, is that all the individual states are exogenous.

We now define a stationary recursive competitive equilibrium for our economy. Stationarity requires that equilibrium skill prices are constant over time, which in turn requires an invariant skill distribution $m(s)$. A stationary skill distribution is consistent with a time-invariant tax schedule, which is the focus of our steady-state welfare analysis. However, when we later consider optimal once-and-for-all tax reforms and incorporate the transition from the current tax system, the economy-wide skill distribution will vary deterministically over time. In that case, an additional assumption is required to preserve tractability. We turn to the transition case in Section 5.2.

Given a tax/transfer system $(\{\tau_a\}, \{\lambda_a\})$, a *stationary recursive competitive equilibrium* for our economy is a public good provision level g, asset prices $Q(\cdot)$, skill prices $p(s)$, decision rules $s(\kappa, \varphi)$, $c(\varphi, s, a, \alpha, \varepsilon)$, $h(\varphi, s, a, \alpha, \varepsilon)$, and $B(\cdot; \varphi, s, a, \alpha)$, effective hours by skill $\bar{N}(s)$, and a skill density $m(s)$ such that:

1. Households solve the problem described in Section 2.1, and $s(\kappa, \varphi)$, $c(\varphi, s, a, \alpha, \varepsilon)$, $h(\varphi, s, a, \alpha, \varepsilon)$, and $B(\cdot; \varphi, s, a, \alpha)$ are the associated decision rules.

2. Labor markets for each skill type clear and $p(s)$ is the value of the marginal product from an additional unit of effective hours of skill type s:

$$p(s) = \left(\frac{Y}{\bar{N}(s) \cdot m(s)} \right)^{\frac{1}{\theta}}.$$

3. The skill density $m(s)$ is consistent with individual decisions.

4. Insurance markets clear, and the prices $Q(\cdot)$ are equal to the probabilities that the realization for ε is in the corresponding set.

5. The government budget is balanced: g satisfies eq. (12).

---

[6]Since equilibrium insurance payouts $B(\varepsilon; \varphi, s, a, \alpha)$ are a known function of all the other individual states, in what follows we omit them from the state vector.

By Walras' law, the goods market clears and eq. (13) holds.

Propositions 1 and 2 below describe the equilibrium allocations and skill prices in closed form. The benefits from analytical tractability will be evident in Propositions 3 and 4, where we derive a set of results for optimal taxation based on a closed-form expression for social welfare.

Proposition 1 [hours and consumption]. *The equilibrium allocations of hours worked and consumption are given by*

$$\log h\left(\varphi, a, \varepsilon\right) = \frac{\log(1 - \tau_a)}{1 + \sigma} - \bar{\varphi}_a - \varphi + \left(\frac{1 - \tau_a}{\sigma + \tau_a}\right)\varepsilon - \frac{1}{\sigma + \tau_a}\mathcal{C}_a, \tag{17}$$

$$\log c\left(\varphi, s, a, \alpha\right) = \log \lambda_a + (1 - \tau_a)\left[\log p(s) + x_a + \alpha + \frac{\log(1 - \tau_a)}{1 + \sigma} - (\bar{\varphi}_a + \varphi)\right] + \mathcal{C}_a \tag{18}$$

where $\mathcal{C}_a = (v_{\varepsilon,a}/2) \cdot (1 - \tau_a)(1 - 2\tau_a - \sigma\tau_a)/(\sigma + \tau_a)$.

With logarithmic utility and zero individual wealth, the income and substitution effects on labor supply from differences in skill levels s, experience x_a, and uninsurable shocks α exactly offset, and hours worked are therefore independent of (s, x_a, α) and of λ_a (the level of taxation) and depend on age only through the age-dependent progressivity rate τ_a and the constant $\mathcal{C}_a$. The hours allocation is composed of four terms. The first term captures the effect of taxes on labor supply in the absence of within-age heterogeneity. This can be interpreted as the hours of a "representative agent" of age a. This term depends on age through progressivity and disutility of work, and is decreasing in both arguments. The second captures the fact that a higher disutility of work leads an agent to choose lower hours. The third term captures the response of hours worked to an insurable shock ε. Note that it has no income effect precisely because it is insurable. The response here is proportional to the tax-modified Frisch elasticity $(1 - \tau_a)/(\sigma + \tau_a)$. This elasticity collapses to the standard Frisch elasticity $1/\sigma$ when $\tau_a = 0$. Note that a progressive system ($\tau_a > 0$) dampens the response of hours to insurable shocks. The fourth term captures the welfare-improving effect of insurable wage variation. As illustrated by Heathcote et al. (2008), greater dispersion of insurable shocks allows agents to work more when they are more productive and take more leisure when they are less productive, thereby raising average productivity, average leisure, and welfare. Progressivity weakens this effect because it reduces the covariance between hours and wages.

Consumption is increasing in the skill price $p(s)$, in the predictable component of labor efficiency x_a, and in the uninsurable stochastic component of wages α. Since hours worked are decreasing in the disutility of work $(\bar{\varphi}_a + \varphi)$, so are earnings and consumption. The redistributive role of progressive taxation is evident from the fact that a larger τ_a shrinks the pass-through to consumption from heterogeneity in initial conditions s and φ and from real-

13

izations of uninsurable shocks α and efficiency units x_a. A lower level of taxation (higher λ_a) increases consumption. Insurable variation in productivity has a positive level effect on average consumption in addition to average leisure. Again, higher progressivity weakens this effect. Because of the assumed separability between consumption and leisure in preferences, consumption is independent of the insurable shock ε.

Proposition 2 [skill price and skill choice]. *In a stationary recursive equilibrium, skill prices are given by*

$$\log p(s) = \pi_0(\bar{\tau}) + \pi_1(\bar{\tau}) \cdot s, \tag{19}$$

where $\bar{\tau}$ is discounted average progressivity, $\bar{\tau} = \left(\frac{1-\beta}{1-\beta^A}\right) \sum_{a=0}^{A-1} \beta^a \tau_a$, and $\pi_1(\bar{\tau})$ and $\pi_0(\bar{\tau})$ are given by

$$\pi_1(\bar{\tau}) = \left(\frac{\eta}{\theta}\right)^{\frac{1}{1+\psi}} (1-\bar{\tau})^{-\frac{\psi}{1+\psi}} \tag{20}$$

$$\pi_0(\bar{\tau}) = \frac{1}{\theta-1} \left\{ \frac{1}{1+\psi} \left[\psi \log\left(\frac{1-\bar{\tau}}{\theta}\right) - \log(\eta) \right] + \log\left(\frac{\theta}{\theta-1}\right) \right\}. \tag{21}$$

The skill investment allocation is given by

$$s(\kappa) = [(1-\bar{\tau})\pi_1(\bar{\tau})]^{\psi} \cdot \kappa = \left[\frac{\eta}{\theta}(1-\bar{\tau})\right]^{\frac{\psi}{1+\psi}} \cdot \kappa, \tag{22}$$

and the equilibrium skill density $m(s)$ is exponential with parameter $\eta^{\frac{1}{1+\psi}} [\theta/(1-\bar{\tau})]^{\frac{\psi}{1+\psi}}$.

Note, first, that the log of the equilibrium skill price takes a "Mincerian" form in the sense that it is an affine function of s. The constant $\pi_0(\bar{\tau})$ is the base log price of the lowest skill level ($s = 0$), and $\pi_1(\bar{\tau})$ is the pre-tax marginal return to skill.

Eq. (20) indicates that higher average progressivity increases the equilibrium *pre-tax* marginal return $\pi_1(\bar{\tau})$. The reason is that increasing progressivity compresses the skill distribution toward zero, and as high skill types become more scarce, imperfect substitutability in production drives up the pre-tax return to skill. Thus, our model features a "Stiglitz effect" (Stiglitz 1985). The larger is ψ, the more sensitive is skill investment to a given increase in $\bar{\tau}$, and thus the larger is the increase in the pre-tax skill premium.

Note that the only aspect of the policy sequence $(\{\tau_a\}, \{\lambda_a\})$ that matters for the skill investment decision and the skill price function is discounted average progressivity, $\bar{\tau}$. Moreover, skill investment is also independent of initial heterogeneity v_φ, of the age profiles $(\{x_a\}, \{\bar{\varphi}_a\})$, and of risk $(v_\omega, \{v_{\varepsilon,a}\})$. The logic is that, with log utility, the welfare gain from additional skill investment is proportional to the log change in earnings such investment would induce, and

this log change is independent of all idiosyncratic states.

Corollary 2.1 [distribution of skill prices]. *In a stationary equilibrium, the distribution of log skill premia $\pi_1(\bar{\tau}) \cdot s(\kappa)$ is exponential with parameter θ. Thus, the cross-sectional variance of log skill prices is*

$$var\left(\log p(s)\right) = \frac{1}{\theta^2}.$$

The distribution of skill prices $p(s)$ in levels is Pareto with scale (lower bound) parameter $\exp(\pi_0(\bar{\tau}))$ and Pareto parameter θ.

Log skill premia are exponentially distributed because the log skill price is affine in skill s (eq. 19) and skills retain the exponential shape of the distribution of learning ability κ (eq. 22). It is interesting that inequality in skill prices is independent of the policy sequence $(\{\tau_a\}, \{\lambda_a\})$. The reason is that progressivity sets in motion two offsetting forces. On the one hand, as discussed earlier, higher progressivity increases the equilibrium skill premium $\pi_1(\bar{\tau})$, which tends to raise inequality in skill prices (the Stiglitz effect). On the other hand, higher progressivity compresses the distribution of skill quantities. These two forces exactly cancel out under our specification of preferences and technology.

Since the exponent of an exponentially distributed random variable is Pareto, the distribution of skill prices in levels is Pareto with parameter θ. More complementarity (lower *theta*) across skills in production stretches further the tail of the wage distribution as the most skilled, and scarcest workers, command a higher premium. The other stochastic components of wages (and hours worked) are lognormal, and thus the equilibrium distributions of wages, earnings, and consumption are Pareto-lognormal. In particular, because the Pareto component dominates at the top, it has a Pareto right tail, a robust feature of their empirical counterparts (see, e.g., Atkinson et al., 2011).

We now describe how taxation affects aggregate quantities in our model.

Corollary 2.2 [aggregate quantities]. *Average hours worked, average effective hours and average output are given by $H\left(\{\tau_a\}\right) = \frac{1}{A}\sum_{a=0}^{A-1} H\left(a, \tau_a\right)$, $\bar{N}\left(\{\tau_a\}\right) = \frac{1}{A}\sum_{a=0}^{A-1} N\left(a, \tau_a\right)$, and $Y\left(\{\tau_a\}\right) =$*

$\frac{1}{A}\sum_{a=0}^{A-1}Y(a,\tau_a)$, *where*

$$H(a,\tau_a) = \mathbb{E}\left[h(\varphi,a,\varepsilon)\right] \tag{23}$$

$$= (1-\tau_a)^{\frac{1}{1+\sigma}}\cdot\exp\left(-\bar{\varphi}_a\right)\cdot\exp\left[\frac{(1-\tau_a)(2\tau_a+\sigma(1+\tau_a))}{(\sigma+\tau_a)^2}\frac{v_{\varepsilon,a}}{2}\right].$$

$$N(a,\tau_a) = \mathbb{E}\left[\exp(x_a+\alpha+\varepsilon)h(\varphi,a,\varepsilon)\right] \tag{24}$$

$$= (1-\tau_a)^{\frac{1}{1+\sigma}}\cdot\exp\left[x_a-\bar{\varphi}_a+\left(\frac{1-\tau_a}{(\sigma+\tau_a)^2}(\sigma+2\tau_a+\sigma\tau_a)\right)\frac{v_{\varepsilon,a}}{2}\right].$$

$$Y(a,\tau_a) = \mathbb{E}\left[p(s)\right]\cdot N(a,\tau_a), \tag{25}$$

with $\mathbb{E}\left[p(s)\right]=\exp\left(\pi_0(\bar{\tau})\right)\cdot\theta/(\theta-1)$.

4 Social welfare function

The baseline utilitarian social welfare function we use to evaluate alternative policies puts equal weight on all agents within a cohort. In our context, where agents have different disutilities of work effort, we define equal weights to mean that the planner cares equally about the utility from consumption of all agents. Thus, the contribution to social welfare from any given cohort is the within-cohort average value of remaining expected lifetime utility, where eq. (2) defines individual expected lifetime utility at age zero. The overlapping-generations structure of the model also requires us to take a stand on how the government weights cohorts that enter the economy at different dates. We assume that the planner discounts lifetime utility of future generations at the same rate β as individuals discount utility over the life cycle.

We start by focusing on optimal steady-state policy, defined as the optimal time-invariant policy $\left(\{\tau_a,\lambda_a\}_{a=0}^{A-1},G\right)$ that maximizes welfare in the associated steady state. In a steady state, expected lifetime utility is identical for each cohort. Moreover, given the assumption that the planner discounts across generations at rate β, average social welfare $\mathcal{W}^{ss}\left(\{\tau_a,\lambda_a\},g\right)$ is simply equal to average utility in a cross section:

$$\mathcal{W}^{ss}\left(\{\tau_a,\lambda_a\},G\right)=\frac{1}{A}\sum_{a=0}^{A-1}\mathbb{E}\left[u\left(c\left(\varphi,s,a,\alpha\right),h\left(\varphi,a,\varepsilon\right),G\right)\right]-\mathbb{E}\left[v\left(s(\kappa),\kappa\right)\right], \tag{26}$$

where the first expectation is taken with respect to the equilibrium cross-sectional distribution of $(\varphi,s,\alpha,\varepsilon)$ conditional on a, and the second expectation is with respect to the cross-sectional distribution of (s,κ). The "Ramsey problem" of the government is to choose $\left(\{\tau_a,\lambda_a\}_{a=0}^{A-1},G\right)$ in order to maximize (26) subject to the government budget constraint (12), where lifetime

utilities are given by (2), equilibrium allocations are given by (17) and (18), and equilibrium skill prices are given by (19).

In Section 5.2 we will consider time-varying policies that maximize welfare incorporating transition from the current tax system. In particular, we will assume an unanticipated policy change at date $t = 0$ from a pre-existing age- and time-invariant policy to a new policy regime in which the new policy parameters can vary freely by both age and time. The irreversibility of the existing stock of skills induces transitional dynamics toward the new steady state.

There are two special cases in which policies that maximize steady-state welfare are identical – in welfare terms – to those that maximize welfare incorporating transition. The first is the case in which $\beta \to 1$. In this case, there is a transition to the new steady state, but because the planner is perfectly patient, existing cohorts receive zero weight in social welfare relative to the planner's concern for future cohorts. Thus, the planner effectively seeks to maximize steady-state welfare.[7] In particular, note that when $\beta = 1$, social welfare is simply expected lifetime utility for a cohort entering in the new steady state, U^{ss}. Then note that in the expression for lifetime utility (eq. 2), the weight $\frac{1-\beta}{1-\beta^A}\beta^a \to \frac{1}{A}$ as $\beta \to 1$.

The second special case in which incorporating transition makes no difference is the case in which skills are perfect substitutes ($\theta \to \infty$) so that there is no skill investment in equilibrium. In this case, transition in response to a change in the tax system is instantaneous, and social welfare incorporating transition is therefore equal to average period utility in the cross section – that is, equal to steady-state welfare.[8]

5 Optimal age-dependent taxes: characterization

For ease of exposition, it is convenient to begin by abstracting from transitional dynamics and to consider optimal policy design in steady state with $\beta = 1$. This approach also has the advantage that we can derive a number of analytical results for optimal taxation. Recall that given $\beta = 1$, transition is irrelevant for welfare, so the policy that is optimal in steady state can also be interpreted as a policy that maximizes welfare incorporating transition.

---

[7]Note that here we are assuming that under the optimal policy the economy does indeed converge to a steady state.

[8]An additional way to achieve an instantaneous transition to the new steady state is to assume that skill investment is fully reversible at any age and date. In our view, irreversible skill investment is the more realistic case.

5.1 Steady-state welfare

We start by characterizing the optimal choices of g and $\{\lambda_a\}$ for any given sequence of age-dependent progressivity $\{\tau_a\}$.

Proposition 3 [optimal g and $\{\lambda_a\}$]. *For any given sequence $\{\tau_a\}$: (i) The optimal share of government expenditure in output g^* is given by*

$$g^* = \frac{\chi}{1 + \chi}.$$

(ii) The optimal sequence $\{\lambda_a^*\}$ equalizes average consumption across age groups.

Part (i) re-establishes a result in Heathcote et al. (2017) in our more general environment with an age-dependent tax system. The optimal fraction of output devoted to public goods is independent of how much inequality there is in the economy and independent of the progressivity of the tax system. It depends only on households' relative taste for the public good χ. Since neither g nor λ_a appear in the equilibrium allocations for hours worked or skill investment, changing g will not affect aggregate income or its distribution across households. As a consequence, the government's only concern in choosing g is to optimally divide output between private and public consumption, exactly as in a version of our economy without any heterogeneity (i.e., the "representative agent economy" in Heathote et al., 2017). In particular, the planner chooses public spending so as to equate the marginal rate of substitution between private and public consumption to the marginal rate of transformation between the two goods, the so-called *Samuelson condition*.[9]

The result in part (ii) states that the planner modulates the level of taxation for each age group $\{\lambda_a\}$ in order to equate the marginal utility of average consumption across age groups. Due to the separability in preferences between consumption and leisure, this implies that average consumption is equated across age groups. Thus, through the choice for the sequence $\{\lambda_a\}$, the government can effectively replicate the role of life-cycle borrowing and saving, absent in the model by assumption, in smoothing predictable life-cycle income variation.

One can substitute the optimal decisions for g^* and $\{\lambda_a^*\}$ along with the closed-form expressions described above for equilibrium allocations into the expression for steady-state welfare, $\mathcal{W}^{ss}$ (eq. 26). One can then express welfare analytically as a function of model parameters and

---

[9]See, for example, Kaplow (2004) for an extended discussion of the Samuelson rule for optimal public good provision and its optimal financing.

of the vector of age-dependent progressivity $\{\tau_a\}$ as follows:

$$
\mathcal{W}^{ss}(\{\tau_a\}) = -\frac{1}{A}\sum_{a=0}^{A-1}\underbrace{\frac{1-\tau_a}{1+\sigma}}_{\text{Disutility of labor}} \tag{27}
$$

$$
+ (1+\chi)\log\underbrace{\left\{\sum_{a=0}^{A-1}(1-\tau_a)^{\frac{1}{1+\sigma}}\cdot\exp\left[x_a-\bar{\varphi}_a+\left(\frac{\tau_a(1+\hat{\sigma}_a)}{\hat{\sigma}_a^2}+\frac{1}{\hat{\sigma}_a}\right)\frac{v_{\varepsilon,a}}{2}\right]\right\}}_{\text{Effective hours worked }\bar{N}_a}
$$

$$
+(1+\chi)\underbrace{\frac{1}{(1+\psi)(\theta-1)}\left[\psi\log(1-\bar{\tau})+\log\left(\frac{1}{\eta\theta^\psi}\left(\frac{\theta}{\theta-1}\right)^{\theta(1+\psi)}\right)\right]}_{\text{Productivity from skill investment: }\log(\text{average skill price}) = \log(\mathbb{E}[p(s)])}
$$

$$
-\underbrace{\frac{\psi}{1+\psi}\frac{1-\bar{\tau}}{\theta}}_{\text{Avg. education cost}} + \frac{1}{A}\sum_{a=0}^{A-1}\underbrace{\left[\log\left(1-\left(\frac{1-\tau_a}{\theta}\right)\right)+\left(\frac{1-\tau_a}{\theta}\right)\right]}_{\text{Consumption dispersion across skills}}
$$

$$
-\frac{1}{A}\sum_{a=0}^{A-1}\underbrace{\frac{1}{2}(1-\tau_a)^2\left(v_\varphi+av_\omega\right)}_{\text{Cons. dispersion due to uninsurable risk and preference heterogeneity}} .
$$

Each term in this welfare expression can be given an intuitive economic interpretation (described in the bracket below each term), along the lines of the analysis contained in Heathcote et al. (2017). The following proposition establishes some properties of $\mathcal{W}^{ss}(\{\tau_a\})$ and of optimal age-dependent progressivity.

Proposition 4 [optimal age dependent progressivity]. *The steady state social welfare function $\mathcal{W}^{ss}(\{\tau_a\})$ is differentiable and globally concave in τ_a provided that σ is sufficiently large (a sufficient condition is that $\sigma \geq 2$). Moreover:*
(i) The necessary and sufficient first-order condition $\partial\mathcal{W}^{ss}(\{\tau_a\})/\partial\tau_a = 0$ implicitly determining the optimal τ_a^* can be stated analytically as:

$$
0 = \frac{1}{\theta-1+\tau_a^*}-\frac{1}{\theta}+(1-\tau_a^*)\left(v_\varphi+av_\omega\right)+\frac{1}{1+\sigma}+ \tag{28}
$$

$$
-\left[\left(\frac{1+\chi}{\theta-1}\right)\frac{1}{1-\bar{\tau}(\{\tau_a^*\})}-\frac{1}{\theta}\right]\frac{\psi}{1+\psi}
$$

$$
-\left(\frac{1+\chi}{1+\sigma}\right)\left[\frac{1}{1-\tau_a^*}+\left(\frac{\sigma+1}{\sigma+\tau_a^*}\right)^3\tau_a^*v_{\varepsilon,a}\right]\frac{N(a,\tau_a^*)}{\bar{N}(\{\tau_a^*\})},
$$

(ii) The optimal sequence $\{\tau_a^*\}$ is age invariant if the following three conditions simultaneously hold:

(1) uninsurable risk does not change over the life cycle ($v_\omega = 0$), (2) insurable risk does not change over the life cycle ($v_{\varepsilon,a}$ is constant), and (3) the age profile for efficiency net of work disutility $\{x_a - \bar{\varphi}_a\}$ is constant.

(iii) Relative to the parameterization described in (ii), introducing permanent uninsurable risk ($v_\omega > 0$) translates into an optimal profile $\{\tau_a^*\}$ that is increasing in age.

(iv) Relative to the parameterization described in (ii), if the variance of insurable risk increases with age ($v_{\varepsilon,a+1} > v_{\varepsilon,a}$) and $\tau_a^* > 0$ at some age a, then $\tau_{a+1}^* < \tau_a^*$.

(v) Relative to the parameterization described in (ii), introducing age variation in efficiency net of disutility $\{x_a - \bar{\varphi}_a\}$ translates into an optimal profile $\{\tau_a^*\}$ that is the mirror image of the profile for $\{x_a - \bar{\varphi}_a\}$.

The Appendix contains a formal proof of this proposition. In what follows, we offer some intuition for results *(ii)-(v)*.

(ii) In this special case, the first-order condition simplifies to an expression where age *a* does not appear, hence τ_a^* is constant.[10] Without loss of generality, to simplify the exposition, consider the case $\theta \to \infty$ and $v_{\varepsilon,a} = 0$ for which case the first-order condition simpifies to

$$0 = (1 - \tau^*) v_\varphi + \frac{1}{1+\sigma} - \left(\frac{1+\chi}{1+\sigma}\right) \frac{1}{1-\tau^*}.$$

where τ^* is the optimal age-invariant τ. It is immediate that τ^* is increasing in preference heterogeneity v_φ and is decreasing in the taste for the public good χ. Note that when $v_\varphi = 0$, $\tau^* = -\chi$. As we show in Heathcote et al. (2017), in this representative agent version of the model (without any source of ex ante or ex post heterogeneity), a regressive tax system allows for a positive average tax rate (which finances G) while ensuring that the representative agent faces a zero marginal rate in equilibrium.

(iii) Now consider the role of uninsurable risk. To isolate this force, we focus on the case in which this is the only source of heterogeneity and $\chi = 0$. The first-order condition (28) then simplifies to

$$0 = (1 - \tau_a^*) a v_\omega + \frac{1}{1+\sigma} \left[1 - \frac{(1-\tau_a^*)^{-\frac{\sigma}{1+\sigma}}}{A^{-1} \sum_{j=0}^{A-1} \left(1 - \tau_j^*\right)^{\frac{1}{1+\sigma}}} \right].$$

When $v_\omega > 0$, the first term is increasing in age *a*, and to satisfy the first-order condition, τ_a^* must therefore be rising in age (so as to reduce the first term and make the second term more negative). The intuition is that permanent uninsurable risk cumulates with age and the planner wants to provide more within-group risk sharing at ages when uninsurable risk is

---

[10] Note that as $\beta \to 1$, $\bar{\tau} \to A^{-1} \sum_{j=0}^{A-1} \tau_a$.

larger. Therefore, when $v_\omega > 0$, optimal progressivity increases with age, *ceteris paribus*. We label this force the *uninsurable risk channel*.

(iv) Now consider the role of insurable risk. Assume the other conditions of part *(ii)* of Proposition 4 are satisfied. The social welfare first-order condition (28) is then

$$0 = (1 - \tau_a^*)\, v_\varphi + \frac{1}{1+\sigma} - \left(\frac{1+\chi}{1+\sigma}\right) \left[\frac{1}{1-\tau_a^*} + \left(\frac{\sigma+1}{\sigma+\tau_a^*}\right)^3 \tau_a^* v_{\varepsilon,a}\right] \frac{N\left(a, \tau_a^*\right)}{\bar{N}\left(\{\tau_a\}\right)}.$$

First, suppose $v_{\varepsilon,a}$ is constant to isolate the role of age-invariant insurable wage variation. It is immediate that there is no motive for age variation in τ_a, (i.e., $\tau_a^* = \tau^*$). In addition, if $\tau^* > 0$ ($\tau^* < 0$), then increasing the level of insurable risk will reduce (increase) optimal progressivity. The intuition is that when dispersion in insurable risk increases, the cost of setting τ away from zero and distorting efficient labor supply allocations increases.

Now, consider the impact of insurable risk that increases with age between age a and $a + 1$, $v_{\varepsilon,a+1} > v_{\varepsilon,a}$. Suppose parameter values are such that τ_a^* is positive, and consider the optimal value for progressivity at age $a + 1$, τ_{a+1}^*. It is clear that the derivative of the social welfare function at $a + 1$ evaluated at τ_a^* is negative (since $N\left(a, \tau_a^*\right)$ and $v_{\varepsilon,a}$ are both increasing in a). We have already established that the social welfare expression is concave in τ_a for each age a. It follows that the optimal degree of progressivity at age $a + 1$ must be less than at age a, (i.e., $\tau_{a+1}^* < \tau_a^*$), so that the $\{\tau_a^*\}$ profile is downward-sloping between a and $a + 1$. The intuition is that when the dispersion of the insurable risk increases with age, the cost of setting τ_a positive and thereby distorting labor supply increases with age. We label this force the *insurable risk channel*.

(v) Now consider the role of the life-cycle profiles of efficiency units and the disutility of work. What matters is the shape of the net profile, $\{x_a - \bar{\varphi}_a\}$. To isolate the impact of this model ingredient, we eliminate all sources of within-age heterogeneity ($\theta \to \infty$, $v_\varphi = v_{\varepsilon,a} = v_\omega = 0$). The optimal value for τ at age a, τ_a^*, is then the solution to the following simplified version of the first-order condition (28), where we have substituted in the expression for effective hours (23):

$$1 - \tau_a^* = \left[\frac{(1+\chi)\exp\left(x_a - \bar{\varphi}_a\right)}{A^{-1}\sum_{j=0}^{A-1}\left(1-\tau_j^*\right)^{\frac{1}{1+\sigma}} \cdot \exp\left(x_j - \bar{\varphi}_a\right)}\right]^{\frac{1+\sigma}{\sigma}}$$

$$= (1+\chi)\frac{\exp\left(\frac{1+\sigma}{\sigma}\left(x_a - \bar{\varphi}_a\right)\right)}{A^{-1}\sum_{j=0}^{A-1}\exp\left(\frac{1+\sigma}{\sigma}\left(x_j - \bar{\varphi}_j\right)\right)}.$$

21

This optimality condition illustrates that *ceteris paribus*, the optimal τ_a^* is lower the larger is $x_a - \bar{\varphi}_a$. Moreover, this effect is stronger the higher is the Frisch elasticity (i.e., the lower is σ). The intuition is that, absent age variation in τ, hours worked will be independent of productivity given our utility function and tax system. The planner can increase aggregate labor productivity and welfare by having agents working longer hours when they are more productive and it is less costly for them to supply labor. When the profile for $x_a - \bar{\varphi}_a$ is upward sloping, this introduces a force for having progressivity decline with age. We label this force the *life-cycle channel*.

Another way to understand this result is that the planner wants to smooth both consumption and the labor wedge (and thus the effective marginal tax rate) over the life cycle. Earnings in this version of the model are given by $y_a = \exp(x_a - \bar{\varphi}_a)(1 - \tau_a)^{\frac{1}{1+\sigma}}$. When $x_a - \bar{\varphi}_a$ and thus earnings are increasing with age, the planner optimally chooses to let λ_a decrease with age in order to equate consumption across age groups. The effective marginal tax rate at age a is $1 - \lambda_a(1 - \tau_a)y_a^{-\tau_a}$. Given a positive and age-invariant τ_a, having λ_a decrease with age and y_a increase with age would imply increasing marginal tax rates. But the planner can smooth marginal tax rates by simultaneously letting τ_a decrease with age. This result is formalized in the following corollary.

Corollary 4.1 [optimal age-dependent taxation with life cycle only]. *Assume that $\theta \to \infty$, and $v_\varphi = v_{\varepsilon,a} = v_\omega = 0$ so that the only heterogeneity in the economy is between ages and is driven by the profile for $\{x_a - \bar{\varphi}_a\}$. Then the optimal profiles $\{\tau_a^*, \lambda_a^*\}$ implement the first best. In particular, they equate both the labor wedge and consumption across age groups. The labor wedge is equal to one at all ages (the marginal tax rate is zero). The average value for τ_a, $A^{-1} \sum_{a=0}^{A-1} \tau_a^*$, is equal to $-\chi$.*

In light of this last set of results on the role of the life cycle, it is clear that the life-cycle productivity channel would be weaker if we introduced opportunities for intertemporal trade. In particular, if households could borrow and lend freely, then hours would tend to naturally covary positively with productivity over the life cycle, even absent age variation in τ_a. Similarly, the more easily consumption can be smoothed intertemporally through markets, the less λ_a needs to vary across ages.[11] In Section 7 we allow individuals to access a non-state-contingent bond subject to a credit limit and explore this issue quantitatively.

---

[11]This effect also would not necessarily be operative if the age-wage profile were endogenous. Examples of endogenous age-wage profiles are models with learning by doing, as in Imai and Keane (2004), and models in which skill investments take time away from work, as in Ben-Porath (1967).

5.2 Welfare with discounting and transitional dynamics

The steady-state welfare expression is tractable, making it easy to understand the various forces driving age variation in tax parameters. However, a complete welfare analysis requires incorporating discounting and the transition because skill investment is an irreversible and a dynamic forward-looking decision. Because of this irreversibility, a standard issue inherent in models with sunk investments arises: in the short run, the government will be tempted to heavily tax high-skill individuals because such taxation is not distortionary ex post. This result is related to the temptation to tax initial physical capital in the neoclassical growth model.[12] In our context, the question is: how does this force affect optimal progressivity?

We therefore now assume $\beta < 1$ and consider an unanticipated policy change at date $t = 0$ from a pre-existing age- and time-invariant policy $\Gamma_{-1} = (\lambda_{-1}, \tau_{-1}, G_{-1})$ to a new policy regime in which the new policy parameters can vary freely by both age and time. Let $\Gamma_t = \{\lambda_{a,t+a}, \tau_{a,t+a}, G_{t+a}\}_{a=0}^{A-1}$ denote the tax and spending policy that will apply to the cohort born at date t, and let $U_t(\Gamma_t)$ denote the corresponding expected lifetime utility.

Social welfare can be written as

$$
\mathcal{W}\left(\{\Gamma_t\}_{t=-(A-1)}^{\infty}; \Gamma_{-1}\right) \equiv (1-\beta)\left[\sum_{t=-(A-1)}^{-1} \beta^t U_t^{old}(\Gamma_t; \Gamma_{-1}) + \sum_{t=0}^{\infty} \beta^t U_t(\Gamma_t)\right]. \tag{29}
$$

The superscript "old" distinguishes the existing cohorts ($t < 0$) already alive at the time of the reform – whose skill investments were made under the old age-invariant policy Γ_{-1} – from future cohorts ($t \geq 0$) whose skill investments are made under the new optimal system. Note that remaining lifetime utility U_t^{old} for the old does not include any skill investment costs. Those investments were made in the past and are sunk from the point of view of the government choosing a new policy.

To preserve tractability, we need to make one additional assumption relative to the baseline model, namely that production is segregated across islands defined by birth cohort. This assumption is required because each cohort now faces a potentially cohort-specific profile for progressivity, and thus the distribution for skill investment will be cohort-specific. The segregation assumption ensures that the distribution of skills within each age-group island is always exponential.[13] There is still a single economy-wide government budget constraint, so the plan-

---

[12] Hassler et al. (2008) explain how it is also present in a model where growth is driven by human capital accumulation.

[13] Note that the key to tractability when analyzing the market for skills is that the distribution of skills is exponential (see Proposition 2). The problem with having different cohorts working in the same labor market would be that different cohorts potentially make human capital investments, implying different skill distributions, and a combined overall distribution of skills that would no longer be exponential (the sum of exponential random

ner can use the tax and transfer system to reshuffle resources across islands.

The equilibrium hours worked and consumption allocations in this version of the economy are analogous to those described above for the steady-state version, with the only difference being that the fiscal policy parameters in eqs. (17)-(18) are now indexed by both age and time. Skill investment decisions are modified as follows. Let

$$\bar{\tau}_{a,t} = \mathbb{E}_{t-a} \left[\frac{(1-\beta)}{(1-\beta^A)} \sum_{j=0}^{A-1} \beta^j \tau_{j,t-a+j} \right] \tag{30}$$

denote the expected discounted sequence for progressivity for the cohort entering the economy at date $t - a$. Note that for $t - a < 0$, $\bar{\tau}_{a,t} = \tau_{-1}$, while for $t - a \geq 0$, $\bar{\tau}_{a,t} = \frac{(1-\beta)}{(1-\beta^A)} \sum_{j=0}^{A-1} \beta^j \tau_{j,t-a+j}$.

Skill investment choices and skill prices for any cohort are given by the same expressions as in the baseline model, except that both are now cohort-specific and depend on the expected sequence for progressivity $\bar{\tau}_{a,t}$. Because skill investment choices are irreversible, unanticipated changes to the tax system have no impact on the skill distribution or skill prices for cohorts entering before date 0.

The Ramsey problem for the planner is now to choose $\left\{ \{\lambda_{a,t+a}, \tau_{a,t+a}\}_{a=\max\{0,-t\}}^{A-1} \right\}_{t=-(A-1)}^{\infty}$ and $\{G_t\}_{t=0}^{\infty}$ to maximize (29) given the expressions for equilibrium allocations and the government budget constraint.

How does incorporating transition change the optimal policy prescription? First, our steady-state characterizations for optimal spending and for the optimal tax level parameters $\{\lambda_{a,t}\}$ extend directly to the transition case.

Proposition 5 [optimal age-dependent taxation with transition]. *Taking the transition into account, the optimal tax system has the following properties:*
(i) At every date t, the optimal sequence $\{\lambda_{a,t}^*\}$ equalizes average consumption across age groups.
(ii) The optimal output share of government expenditures g_t^* is constant and given by

$$g_t^* = \frac{\chi}{1+\chi}.$$

The logic for part (i) is that, as in the steady state, the $\lambda_{a,t}$ parameters have no effect on labor supply or skill investment. The intuition for part (ii) is related: given that the average level of taxation does not affect output, it is optimal to set the level of government spending to equate the marginal utilities of public and private consumption.

To characterize the impact of incorporating transition on the optimal age profile of progressivity, we now focus on a special case of the model in which heterogeneity in skills is the only

variables is not an exponential).

24

source of heterogeneity. This strips out other sources of age variation in optimal progressivity and allows us to focus on incentives of the planner to exploit the fact that past skill investments are sunk and therefore are insensitive to changes in the tax system. This adds a new driver shaping optimal progressivity, which we label the *sunk skill investment channel*. To obtain the sharpest characterization of this effect, we also assume inelastic labor supply.

Proposition 6 [optimal taxation with transition and inelastic labor supply].
If (i) $v_\varphi = v_\omega = v_{\varepsilon,a} = 0$, (ii) the age profiles for efficiency and disutility of work are flat, and (iii) $\sigma \to \infty$ (labor supply is inelastic), then the optimal policy has the following properties: $\tau^*_{a,t} = 1$ for all $a > t$, and $\tau^*_{0+j,t+j} = \tau^*_{0,t} < 1$ for all $j = 1, .., A - 1$ and for all $t \geq 0$.

This result states that it is optimal to impose maximally progressive taxes on all cohorts who entered the economy before the tax reform at date 0, whose past skill investments are sunk. This eliminates within-age-group consumption inequality for these cohorts, without imposing any distortions. For cohorts who enter the economy after the reform, optimal progressivity is constant over the life cycle and less than one. It is not optimal to push progressivity to the maximum because for these cohorts, progressivity reduces skill investment. Why is progressivity constant over the life cycle? Consider the trade-offs from a marginal increase in $\tau_{1,t+1}$ relative to $\tau_{0,t}$, starting from a flat profile. Skill investment at t is less sensitive by a factor β to $\tau_{1,t+1}$ relative to $\tau_{0,t}$ (see eq. 30). At the same time, the gain in terms of reduced consumption inequality from increasing $\tau_{1,t+1}$ relative to $\tau_{0,t}$ is also discounted by a factor β, since it enters social welfare at $t + 1$ rather than at t.[14]

The characterization in Proposition 6 parallels the well known result that in models with physical capital, the Ramsey planner wants a declining path for capital taxes in order to expropriate existing sunk capital without excessively discouraging new investment. In our economy, the planner effectively expropriates the returns to past skill investments, without discouraging future skill investment. However, the key to achieving this, in the context of our overlapping-generations economy, is to have progressivity vary by cohort, rather than by time, because human capital is non-tradable, and the age of the potential human capital investor perfectly delineates whether or not the investment is sunk.

The optimal policy described in Proposition 6 dictates very different optimal progressivity values for cohorts entering the economy before versus after the reform. We have also explored the optimal policy when the planner can allow progressivity to vary by age but not by time/cohort. In particular, suppose that at the time of tax reform, the planner has to choose a single age profile $\{\tau_a\}$ that will apply at every future date. We retain the assumption that the

---

[14]Note that while optimal progressivity is constant within each cohort, it potentially varies across cohorts during transition.

25

Parameter	Description	Value
A	Years of working life	35
β	Discount factor	0.97
σ	Inverse of Frisch elasticity	2
χ	Relative taste for public good	0.233
θ	Elasticity of substitution across skills	3.124
ψ	Elasticity of skill investment to return	0.65
v_φ	Heterogeneity in disutility of work	0.036
v_ω	Variance of uninsurable productivity shock	0.058
$v_{\varepsilon 0}$	Initial variance in insurable productivity	0.090
Δv_ε	Growth in variance of insurable productivity	0.044
$\{x_a\}$	Age profile for productivity	See Fig. 2
$\{\bar{\varphi}_a\}$	Age profile for disutility of work	See Fig. 2
τ^{US}	US rate of progressivity	0.181

Table 1: Model parameterization (period = 1 year)

other fiscal parameter $\lambda_{a,t}$ can vary freely by age and time. The key result in this case is that when $\beta < 1$, and given the parametric assumptions listed in Proposition 6, the optimal policy incorporating transition features an increasing profile for τ_a. Given Proposition 6, this result should come as no surprise: an increasing age profile for τ_a is a poor man's approximation to the ideal policy, which dictates high progressivity for the pre-existing old cohorts and low progressivity for new young cohorts.

In the next section, we will numerically explore optimal taxation incorporating transition in a full version of the model.

6 Quantitative analysis

In this section, we describe the model parameterization and explore the quantitative implications of the theory. We begin with the problem of the planner that maximizes steady-state welfare as in Section 5. Next, we solve for the optimal age-dependent tax system that incorporates discounting and transitional dynamics.

6.1 Parameterization

The parameterization strategy closely follows Heathcote et al. (2017). The model period is one year. Some of the parameters are set outside the model. For our steady-state analysis, we focus on the case $\beta = 1$, since in this case ignoring transition is innocuous. When we move to explore transition, we set $\beta = 0.97$, so that the path for policy and allocations during transition matters

for social welfare.

Households live for $A = 36$ years, envisioning an age range between 25 and 60. The motivation for this choice is that our focus is on the design of a tax and transfer system for the working-age population. In Section 7, we extend the analysis to a case with exogenous retirement.[15] The preference weight on the public good χ is identified directly from the size of the US government as a share of GDP, assuming that the observed level of public consumption is optimal: given $g^{US} = 0.19$, we obtain $\chi = 0.233$.[16] For calibration, we need to approximate the current US tax and transfer system. Based on the estimates of Heathcote et al. (2017), we set $\tau^{US} = 0.181$.[17] Given τ^{US} and g^{US}, we then set λ^{US} such that the budget is balanced. We set $\sigma = 2$, a value broadly consistent with the microeconomic evidence on the Frisch elasticity (see, e.g., Keane, 2011).

Other parameters are structurally estimated. In Heathcote et al. (2017), we show that one can identify and estimate the elasticity of substitution between skills θ, preference heterogeneity v_φ, and the variances of wage risk v_ω, $\{v_{\varepsilon,a}\}$, using cross-sectional within-age variances and covariances of male wages, male hours, and equivalized household consumption, which we measure from the the Panel Study of Income Dynamics (PSID) and the Consumer Expenditure Survey (CEX) for survey years 2000-2006. The identification follows from the closed-form expressions for wages, hours, and consumption derived above.

---

[15]See Ndiaye (2017) for an analysis of optimal age-dependent taxation in a model that incorporates the retirement decision.

[16]Heathcote et al. (2017) show that the fraction of output devoted to public goods is also $\frac{\chi}{1+\chi}$ when it is chosen by the median voter in the economy.

[17]For this exercise, Heathcote et al. (2017) use data from the PSID for survey years 2000-2006, in combination with the NBER's TAXSIM program. They restrict attention to households aged 25-60 with positive labor income. When measuring pre-government gross household income, Heathcote et al. (2017) include labor earnings, private transfers (alimony, child support, help from relatives, miscellaneous transfers, private retirement income, annuities, and other retirement income), plus income from interest, dividends, and rents. To construct taxable income, for each household in the data they compute the four major categories of itemized deductions in the US tax code — medical expenses, mortgage interest, state taxes paid, and charitable contributions – and subtract them from gross income. Post-government income equals pre-government income plus public cash transfers (AFDC/TANF, SSI and other welfare receipts, Social Security benefits, unemployment benefits, workers' compensation, and veterans' pensions), minus federal, payroll, and state income taxes. Transfers are measured directly from the PSID, while taxes are computed using TAXSIM.
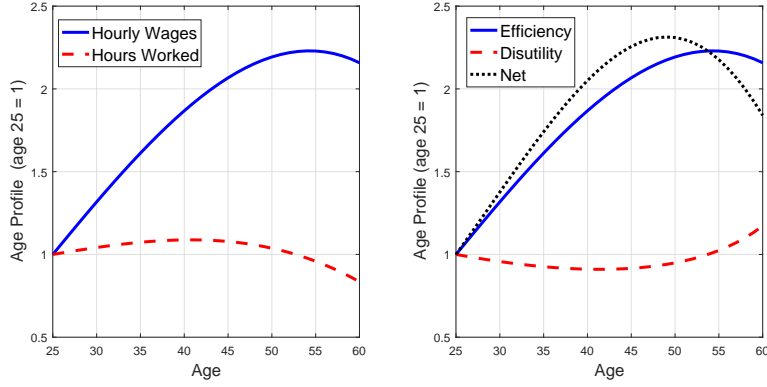
Figure 2: Left panel: life-cycle profiles of individual wages and hours. Right panel: implied profiles for x_a, $\bar{\varphi}_a$, and $x_a - \bar{\varphi}_a$.

To give a flavor of the identification, consider the following four moments:

$$var_a \left(\log w_{ia} \right) = \frac{1}{\theta^2} + v_\omega a + v_{\varepsilon,a} \tag{31}$$

$$var_a \left(\log h_{ia} \right) = v_\varphi + \left(\frac{1 - \tau^{US}}{\sigma + \tau^{US}} \right)^2 v_{\varepsilon,a}$$

$$var_a \left(\log c_{ia} \right) = \left(1 - \tau^{US} \right)^2 \left(v_\varphi + \frac{1}{\theta^2} + v_\omega a \right)$$

$$cov_a \left(\log h_{ia}, \log w_{ia} \right) = \left(\frac{1 - \tau^{US}}{\sigma + \tau^{US}} \right) v_{\varepsilon,a}$$

The moments $cov_a \left(\log h_{ia}, \log w_{ia} \right)$ observed at ages $a = 0, ..., A - 1$ identify $\{v_{\varepsilon,a}\}$. Since in the data the profile for this variance increases nearly linearly with age, we freely estimate the initial variance at age 25, $v_{\varepsilon,0}$, and then impose linearity. From $var_a \left(\log h_{ia} \right)$ we then identify v_φ. The value for $var_0 \left(\log c_{i0} \right)$ identifies θ. Then, the change in $var_a \left(\log w_{ia} \right)$ over the life cycle identifies v_ω. This is just one of the many possible combinations of moments that yield identification. Our formal estimation procedure also allows for classical measurement error in all variables and is based on an estimator that minimizes the distance between age-specific covariances in the model and the data. See Heathcote et al. (2017) for additional details.

The parameter ψ controls the elasticity of the return to skills π_1 to $\bar{\tau}$ and θ, where the return to skills is increasing in progressivity and decreasing in skill substitutability (see eq. 20). In Heathcote et al. (2017), we exploit changes in π_1, τ, and θ over time, which we can measure from PSID data between the early 1970s and the early 2000s, to estimate ψ.

The only additions relative to the parameterization in Heathcote et al. (2017) are the age profiles for productivity and the disutility of work. We estimate the life-cycle profile of individ-
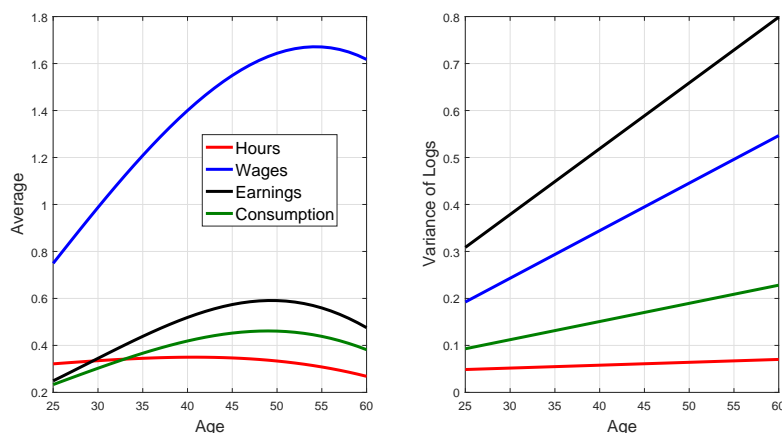
28

Figure 3: Means (left panel) and variances of logs (right panel) over the life cycle

ual hourly wages and hours from our same PSID sample for years 2000-2006. The left panel of Figure 2 plots both profiles, interpolated using a cubic function of age. The wage profile maps directly into the efficiency profile $\{x_a\}$. Given $\{x_a\}$ and the other parameter values, from the expression for average hours worked by age, eq. (24), we can residually recover the profile for disutility of work $\{\bar{\varphi}_a\}$.

The right panel of Figure 2 plots the implied profiles for $\{x_a\}$ and $\{\bar{\varphi}_a\}$ and for $\{x_a - \bar{\varphi}_a\}$, which is the one relevant for optimal age dependence in progressivity. Note that this latter age profile is strongly hump shaped, a feature that will be quantitatively important.[18].

Table 1 summarizes the parameter values. Figure 3 shows that the implied means and variances of logarithms for wages, hours, earnings, and consumption by age align well with the ones estimated from cross-sectional data (see, e.g., Heathcote et al., 2014).

6.2 Results: steady-state welfare

In line with the analytical results in Section 4, we start by analyzing optimal taxation from a steady-state welfare point of view.

Recall that Proposition 4 identified three different forces that shape the optimal age profile of tax progressivity in steady state: uninsurable risk, insurable risk, and life-cycle variation in productivity and preferences. To understand the quantitative role of each of these forces, we start from an economy where none of these channels is active, the one described in point (ii) of

---

[18]It is worth noting that our approach to identifying $\{\bar{\varphi}_a\}$ hinges on our assumption of no intertemporal borrowing and lending. If households could perfectly smooth consumption by borrowing and lending, one would naturally expect covariation between hours and wages over the life cycle, implying a smaller role for $\bar{\varphi}_a$ in generating age variation in hours worked. We will therefore also study optimal policy abstracting from age variation in preferences for work

Proposition 4.

6.2.1 Channels that do not induce age dependence

Figure 4 illustrates optimal progressivity by age $\{\tau_a^*\}$ and the implied income-weighted average marginal tax rate (left panels) together with age profiles for earnings, hours, and consumption (right panels).

The top two panels represent optimal policy in a representative-agent version of our economy, with all sources of heterogeneity shut down, that is, $\theta = \infty$, $v_\varphi = v_{\varepsilon,a} = v_\omega = 0$, $\{x_a\}$, $\{\bar{\varphi}_a\}$ constant, and $\beta = 1$. In this economy, $\tau^* = -\chi$.

Next, in the middle panels, we add heterogeneity in the disutility of work by setting v_φ to its estimated value. Since this form of initial heterogeneity translates into consumption dispersion, the planner wants to increase progressivity to redistribute from the lucky individuals born with a low disutility of work to the unlucky ones who have a higher disutility and who thus work and earn less. Since this form of heterogeneity is innate and does not vary by age, optimal progressivity remains flat.

In the bottom two panels, we activate skill investment by setting θ to its estimated value and thus introduce heterogeneity in skills. The optimal $\{\tau_a^*\}$ profile remains flat but further increases in value. Two contrasting forces emerge when we add skill investment: on the one hand, the planner can encourage skill accumulation via a less progressive tax system. On the other hand, the utilitarian planner also wants to reduce consumption inequality generated by heterogeneity in skills, and to do so, it must choose a more progressive system. Given our parameter values, this latter force dominates, pushing up optimal progressivity.

Next, we introduce the channels that induce age dependence in optimal progressivity.

6.2.2 Uninsurable risk channel

Part (iii) of Proposition 4 states that because uninsurable risk in the form of permanent shocks cumulates over the life cycle, the planner has an incentive to increase tax progressivity with age. To introduce this effect, we set the amount of uninsurable risk, v_ω, to its calibrated value. The top panels of Figure 5 illustrate that adding uninsurable risk has two effects on the profile for optimal progressivity. First, the average level of progressivity rises. Second, as expected, the progressivity profile becomes upward sloping.

Introducing uninsurable risk also leads to an upward-sloping age profile for the income-weighted average marginal tax rate. This result is reminiscent of findings in the recent literature on dynamic Mirrleesian optimal taxation, according to which, when income shocks are persistent, the optimal average effective marginal tax rate has a positive drift over the life cycle. Farhi
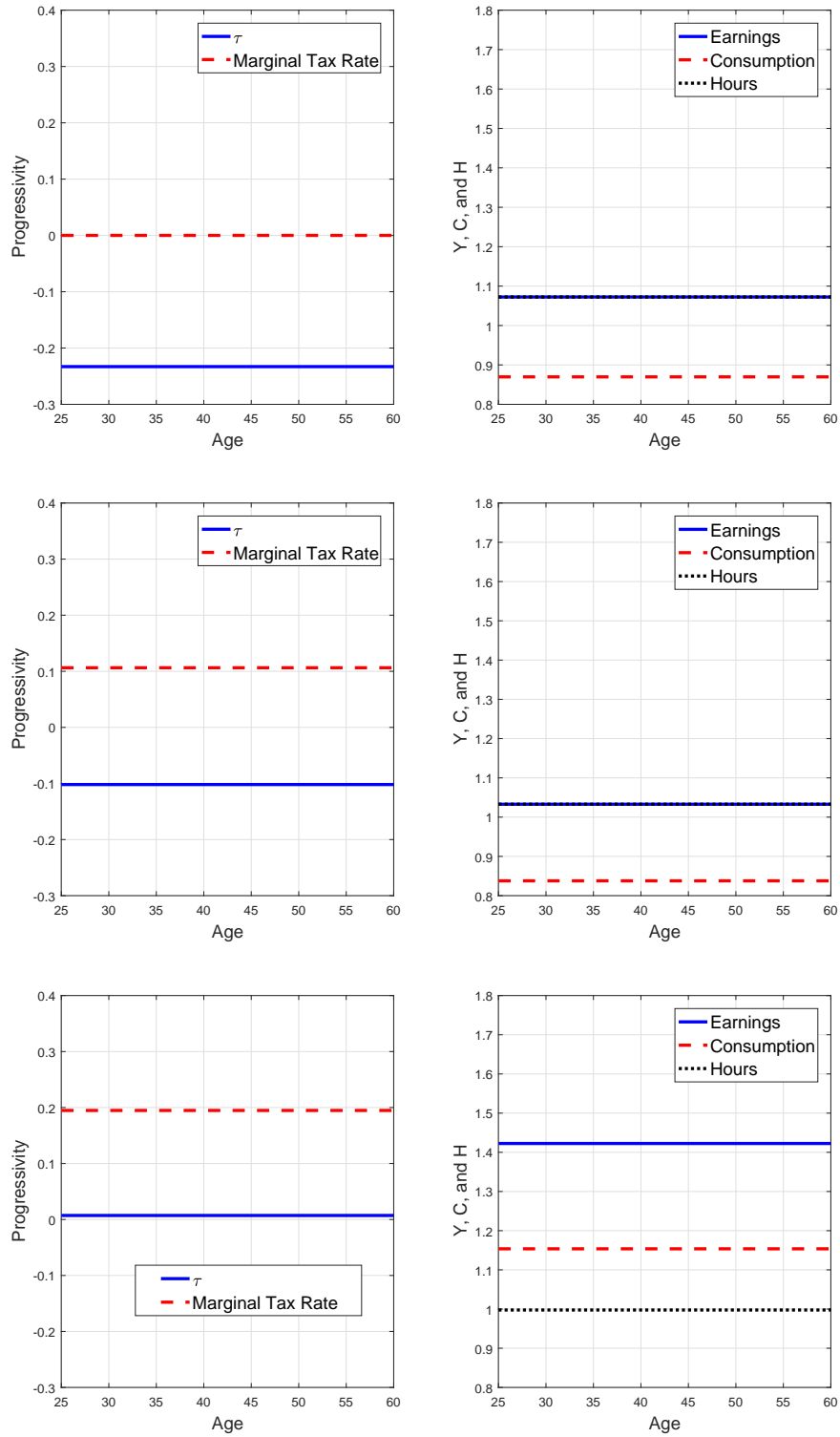
Figure 4: Left panels: optimal progressivity and income weighted average marginal tax rate. Right panels: average earnings (Y), hours (H), and consumption (C) by age. Top two panels: Representative-agent model. Middle two panels: previous case plus heterogeneity in disutility of work. Bottom two panels: previous case plus heterogeneity from skill investment.
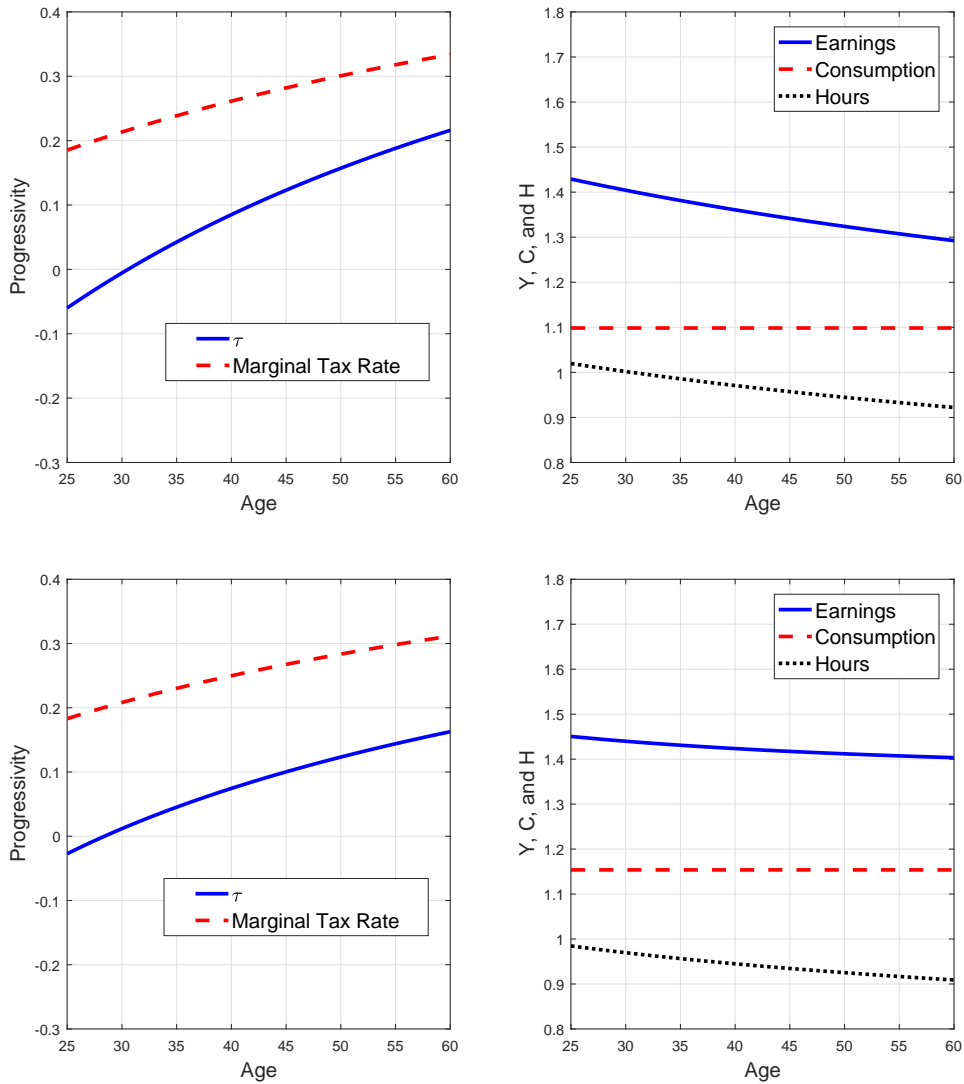
Figure 5: Left panels: optimal progressivity and income-weighted average marginal tax rate. Right panels: average earnings (Y), hours (H), and consumption (C) by age. Top two panels: previous case plus uninsurable risk. Bottom two panels: Previous case plus insurable risk.

and Werning (2013) analyze Mirrlees taxation in a dynamic life-cycle economy in which average productivity does not vary with age.[19] In their numerical example, plotted in their Figure 2, the optimal history-dependent allocation is qualitatively similar to the allocation in our Figure 5: the average effective marginal tax rate is increasing in age, average output is decreasing in age, and average consumption is invariant to age.[20]

6.2.3 Insurable risk channel

According to part (iv) of Proposition 4, if the variance of insurable wage risk $v_{\varepsilon,a}$ increases with age, the planner has an incentive to tilt the schedule for optimal progressivity downward. The bottom panels of Figure 5 illustrate that when we introduce our estimates for insurable risk, the profile of optimal progressivity does indeed tilt in a clockwise direction. As a result, the life-cycle profiles for hours and earnings become flatter.

6.2.4 Life-cycle channel

We now add the last motive for age-varying progressivity identified in Proposition 4: age-varying profiles for labor efficiency and the disutility of work. Figure 6 plots two cases. In the top panels, the productivity and disutility profiles $\{x_a\}$ and $\{\bar{\varphi}_a\}$ are both switched on. Recall that these two ingredients enter the expression for social welfare only via their net effect, $\{x_a - \bar{\varphi}_a\}$. In the bottom panels, only the labor productivity profile $\{x_a\}$ is active.

Recall that the profile for $\{x_a - \bar{\varphi}_a\}$ is generally increasing and strongly hump-shaped (see Figure 2). Thus, optimal progressivity becomes both flatter and more U-shaped when this life-cycle channel is activated, relative to the same economy without age variation in wages or preferences (see the bottom panels of Figure 5). The intuition is that life-cycle earnings have a pronounced hump shape in this calibration. To counteract earnings inequality by age and equate average consumption across age groups, the planner sets a U-shaped age profile for λ_a. Absent age variation in τ_a, this would translate into a strongly hump-shaped profile for average marginal tax rates. By simultaneously setting a U-shaped profile for τ_a, the planner can moderate the average marginal tax rate at peak productivity years. The desire to smooth taxes by age is familiar from the dynamic Mirrlees literature (Farhi and Werning, 2012). The bottom panels of Figure 6 show that the life-cycle channel is weaker when we shut down age variation in preferences.

---

[19]They also assume no endogenous skill accumulation, no preference heterogeneity and no valued government expenditures.

[20]Golosov et al. (2016) show that with negatively skewed log-income shocks, the positive drift in the labor wedge is stronger in the left tail of the income distribution.
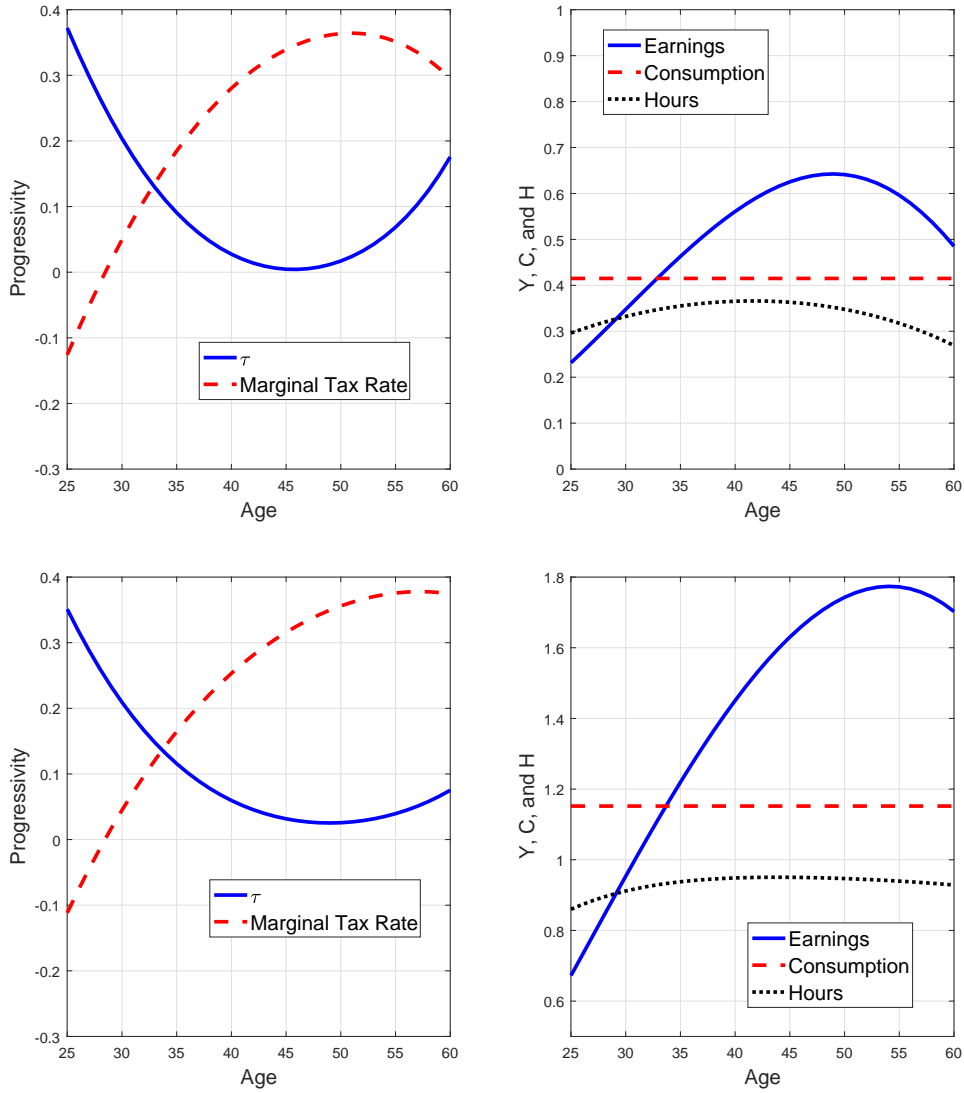
Figure 6: Left panels: optimal progressivity and income-weighted average marginal tax rate. Previous case plus life-cycle channel, that is, all channels operational with steady-state welfare. Right panels: average earnings (Y), hours (H), and consumption (C) by age. Top panels: age profile for disutility of work as estimated. Bottom panels: age profile for disutility of work constant.
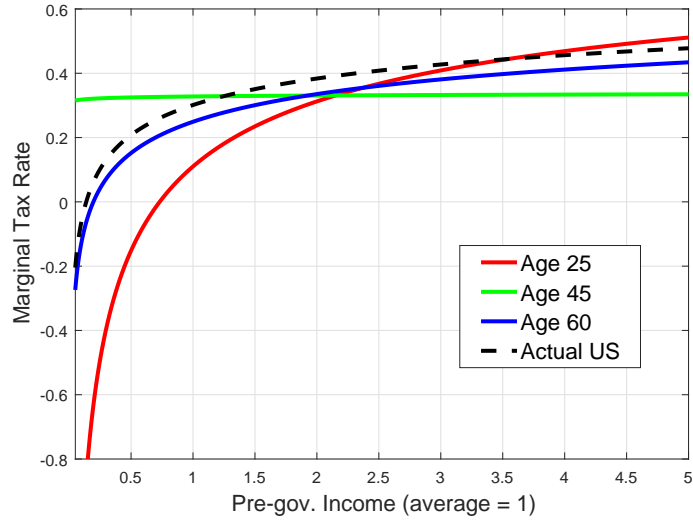
Figure 7: Marginal tax rates at different ages for the optimal age-dependent policy with $\beta = 1$ (i.e., those that maximize steady-state welfare).

All channels are now operative, so this economy (the one with age variation in $\bar{\varphi}_a$) should be viewed as our benchmark when focusing on steady-state welfare. Note, however, that the quantitative importance of the life-cycle channel is sensitive to the assumed market structure. As we will see in Section 7, allowing for borrowing and lending dampens this channel.

6.2.5 Optimal age-dependent marginal tax rates

Figure 7 plots the marginal tax rates implied by the tax system described in the top panels of Figure 6 for three age groups. The optimal age-dependent tax system dictates essentially a flat tax for middle-aged workers and a highly progressive schedule for young and old. Thus progressivity is high for young and old, but for different reasons: it provides social insurance against labor market risk to the old and it provides redistribution without distorting life-cycle labor supply incentives to the young.

Note that even though the degree of optimal progressivity is lower for the old than for the young (the curve is flatter for the old), marginal (and average) tax rates are much higher for the old for a wide income range. Mechanically, this reflects the fact that the old face smaller values for λ_a in order to redistribute income to the young and thereby equalize consumption across ages. When evaluating the optimal tax system at the endogenous distribution of earnings, the income-weighted average marginal tax rate – calculated across individuals in each age group – is increasing in age (Figure 6).

35

	Benchmark	US BL	Natural BL
(λ^*, τ^*) constant	0.10	0.16	0.18
λ^* age-varying, τ^* constant	1.69	1.07	0.67
λ^* constant, τ^* age-varying	2.10	1.63	1.36
(λ^*, τ^*) age-varying	2.42	1.76	1.38

Table 2: All numbers in the table are welfare gains expressed as additional lifetime consumption (percentage points) relative to the existing tax/transfer system. The column labeled "Benchmark" refers to the benchmark economy without intertemporal trade (autarky). The column "US BL" refers to the economy with borrowing and lending under the calibrated borrowing limit for the US economy (two times annual earnings). The column labeled "Natural BL" refers to the economy with borrowing and lending under the natural borrowing limit. These last two economies are solved in general equilibrium.

6.2.6   Welfare gains from tax reforms

We now present the welfare gains of switching from the existing tax/transfer system to the optimal age-dependent system. As with all the results presented to this point, we focus on the case $\beta = 1$, so that we can safely ignore transition when comparing different tax systems.

We begin by reporting the gains of switching to the optimal age-invariant system. Next we consider the gains of switching to a system where we allow for age variation in λ_a but not in τ_a. Then we explore the opposite configuration. Finally, we compute the gains from switching to the fully age-dependent system. All these welfare gains refer to steady-state welfare and are computed in terms of lifetime consumption-equivalent variation. The first column of Table 2 summarizes these results.

The welfare gains of moving from the existing tax system (with $\tau^{US} = 0.181$) to the optimal age-invariant tax schedule are small. Moving to the optimal age-varying system delivers welfare gains of 2.4% of consumption. A large portion of this welfare gain arises from endowing the planner with the ability to use the tax system to redistribute across age groups, so as to equate expected consumption by age. In particular, the specification in which λ can vary freely by age but τ cannot achieves around 2/3 of the maximum welfare gains from tax reform. However, our baseline parameterization likely exaggerates the potential welfare gains from redistribution across age groups, for two reasons. First, it assumes age-invariant utility from consumption: introducing age-varying utility would imply smaller potential gains from redistribution across age groups. We study this case in the Section 6.4. Second, our baseline parameterization assumes no borrowing and lending, giving the government a crucial role in

smoothing consumption over the life-cycle. We extend the model to intertemporal trade in Section 7.

6.3 Results: transitional dynamics

We now compute the age-dependent tax system that maximizes welfare taking into account transitional dynamics and the sunk investment channel, that is, the fact that cohorts born before the reform cannot adjust skills in response to a surprise change in the tax system. In particular, consider a tax reform at date 0 that implements a flexible age- and time-specific tax policy $\{\lambda_{a,t}, \tau_{a,t}, G_t\}$ for $a = 0, ..., A - 1$ and for $t = 0, ..., \infty$. We set the annual discount factor to $\beta = 0.97$.

The importance of the sunk investment channel depends on the tax system in place in the initial steady state. We assume that this system features the age-invariant value for progressivity $\tau^{US} = 0.181$ and the age-invariant λ^{US} that balances the budget given g^{US}.

We emphasize that this is an ambitious exercise because there are a large number of policy parameters to optimize. Doing so is only feasible because, conditional on the tax parameters, equilibrium allocations can be characterized in closed form. In addition, the problem is simplified because, by virtue of Proposition 5, the planner will optimally set $\{\lambda_{a,t}\}$ such that (i) consumption is equalized across age groups at each date, and (ii) the ratio of government spending to output is always equal to $\chi / (1 + \chi)$. To economize slightly on the number of policy parameters to solve for, we assume a three-year period length with each cohort active for $A = 12$ periods and adjust other parameters accordingly.[21]

We plot optimal policy for three parameterizations. The first (Figure 8) corresponds to the case described in Proposition 6 with inelastic labor supply and no sources of heterogeneity besides differences in skills. The second (Figure 9) is identical except that we introduce flexible labor supply. The third parameterization (Figure 10) is our baseline incorporating all sources of heterogeneity (but with $\beta = 0.97$ rather than $\beta = 1$). In all plots, each different colored line plots the sequence for $\{\tau_{a,t+a}\}$ for a particular cohort. The line starts at the date t that the cohort enters the economy and ends at $t + 11$. Lines for cohorts that entered the economy prior to the reform at date 0 are shorter: there is a single point for the cohort that entered at $t = -11$.

Consider first the case in which skill is the only source of heterogeneity and labor supply is inelastic. Figure 8 offers a visual illustration of Proposition 6. The planner sets $\tau_{a,t} = 1$ for cohorts entering prior to the reform, and for cohorts entering post-reform, progressivity is constant over the life cycle. In this example, there is also very little variation in progressivity

---

[21]We assume that the economy converges to a new steady state within 156 periods and thus solve for $12 \times 156 = 1,872$ values for $\tau_{a,t}$.
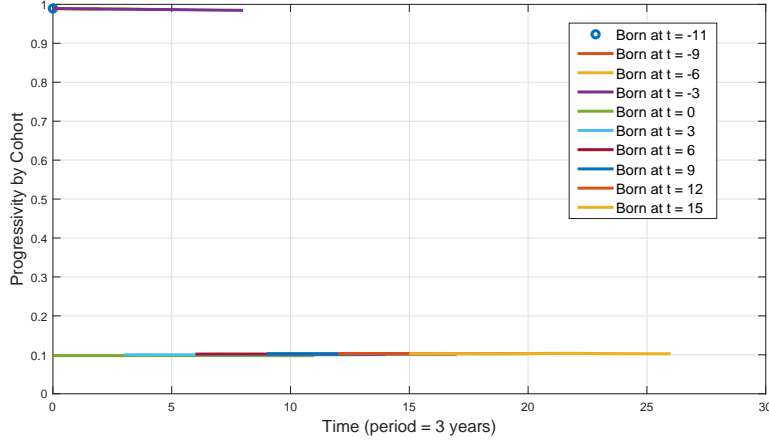
Figure 8: Optimal age and time dependent progressivity incorporating transition. Skills are the only source of heterogeneity. Labor supply is inelastic. $\beta = 0.97$.

across cohorts, but that result is a numerical accident and reflects the fact that the value for optimal progressivity in the final steady state is not far from the estimated value for the United States.

When we introduce flexible labor supply (Figure 9), the optimal policy still involves relatively high values for progressivity for cohorts entering prior to the reform and lower values for cohorts entering after the reform. However, flexible labor supply does change the picture in two ways. First, it is no longer optimal to set $\tau_{a,t} = 1$ for cohorts entering prior to the reform because even though the skill investments for these cohorts are sunk, cohort labor supply still responds negatively to progressivity. Second, the cohort-specific profiles for $\tau_{a,t}$ are generally upward sloping (rather than flat) until all the cohorts alive at the time of the reform have exited the economy. The logic for this result is that aggregate output gradually increases during the post-reform transition, as successive cohorts make skill investments given expected progressivity values $\bar{\tau}_{a,t}$ that are much lower than in the pre-reform steady state. As output increases over time, the planner gradually becomes less focused on stimulating additional output (via low values for progressivity) and more focused on reducing inequality (via high values for progressivity). Thus, during transition, optimal progressivity increases both within cohorts (the upward-sloping profiles) and between profiles (each successive cohort's profile starts at a higher level).[22]

---

[22]The planner's first-order conditions can be used to establish the result that age profiles for progressivity are upward sloping when output is rising. Contemplate a candidate optimal policy with the property $\tau_{0,t} = \tau_{1,t+1}$ such that the first-order condition for $\tau_{0,t}$ (eq. A26) is satisfied. Now consider the first-order condition for $\tau_{1,t+1}$. Substituting eq. (A28) into eq. (A27), it is clear that on the margin, it will be welfare improving to increase $\tau_{1,t+1}$ above $\tau_{0,t}$ if and only if the second term on the right-hand side of eq. (A27) is positive, which will be the case when $Y_t < Y_{t+1}$. Thus, if a cohort will live through a period of rising output, it will optimally face an increasing age
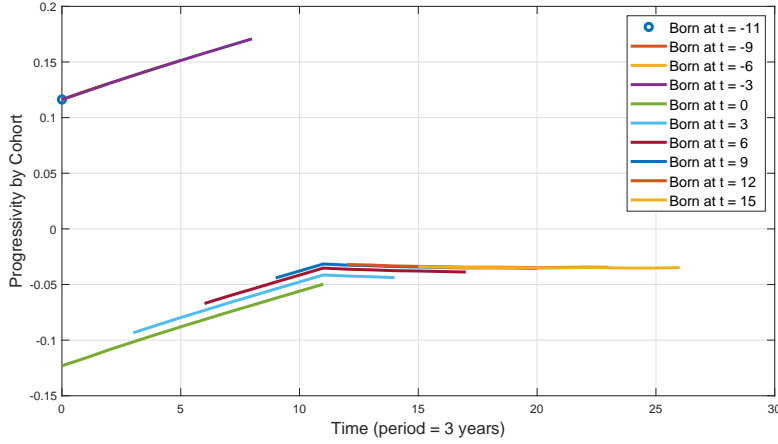
Figure 9: Optimal age and time dependent progressivity incorporating transition. Skills are the only source of heterogeneity. Labor supply is elastic. $\beta = 0.97$.

Now consider optimal policy incorporating transition for the baseline model. The optimal policy now looks like a mix of Figures 6 and 9. For any given cohort, the optimal profile $\{\tau_{a,t+a}\}$ is U-shaped, as in Figure 6. Moving across cohorts, it is clear that on average, progressivity is higher for cohorts entering prior to the reform and lower for cohorts entering later. In addition, progressivity generally increases modestly over time post-reform.

6.4 Extension I: age variation in the taste for consumption

In this section, we generalize our baseline model by introducing life-cycle variation in the taste for consumption. The most straightforward way to interpret this additional model ingredient – and the one we will use to calibrate this version of the model – is that household consumption demand changes as individuals form couples, have children, and children age and eventually leave to form households of their own.

The implication for the planner is that a portion of consumption variation over the life cycle is efficient. This reduces the desire to redistribute across ages through λ_a, which, in turn, weakens the life-cycle channel that induces a U-shaped optimal profile for τ_a.

We modify our period utility function to

$$ u_i\left(c_{ia}, h_{ia}, G\right) = \exp((1+\sigma)\gamma_a) \log c_{ia} - \frac{\exp\left[(1+\sigma)\left(\bar{\varphi}_a + \varphi_i\right)\right]}{1+\sigma} \left(h_{ia}\right)^{1+\sigma} + \chi \log G, \qquad (32) $$

where $\gamma_a > 0$ shifts the marginal utility of consumption at age a. Consumption and hours allocations for this model are:
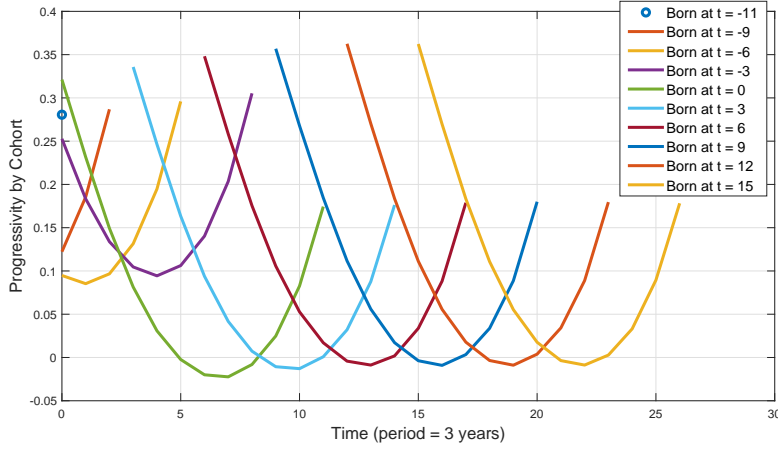
---

profile for progressivity.

39

Figure 10: Optimal age and time varying progressivity incorporating transition. Baseline calibration. $\beta = 0.97$.

$$\log c\left(\varphi, s, a, \alpha\right) = \log \lambda_a + (1 - \tau_a)\left[\log p(s, \bar{\tau}) + x_a + \alpha + \frac{\log(1 - \tau_a)}{1 + \sigma} - (\varphi + \bar{\varphi}_a - \gamma_a)\right] + \mathcal{C}_a$$

$$\log h\left(\varphi, a, \varepsilon\right) = \frac{\log(1 - \tau_a)}{1 + \sigma} - (\varphi + \bar{\varphi}_a - \gamma_a) + \left(\frac{1 - \tau_a}{\sigma + \tau_a}\right)\varepsilon - \frac{1}{\sigma + \tau_a}\mathcal{C}_a.$$

A higher value for γ_a implies a higher marginal value of a unit of consumption and thus a higher optimal consumption level. The age-dependent utility shifter γ_a also affects the individual labor supply decision: at ages when γ_a is high, individuals choose to work more in order to consume more. As a result, γ_a enters the two terms in the welfare function that capture the average disutility of work and the effective hours worked by each skill type. Moreover, γ_a shifts the relative marginal utility of consumption across ages, and as a result, it modifies the cost of consumption dispersion between and within skill types, the last two terms of the welfare function.

Following the same steps we delineated to construct the steady-state welfare expression in the baseline economy, we obtain an analogous welfare expression for the model with life-cycle variation in the taste for consumption. The modified expression for steady-state welfare for the case $\beta = 1$ is in the Appendix.

We now turn to the parameterization of this model. From the expressions for equilibrium individual allocations, it is immediate that γ_a and $\bar{\varphi}_a$ cannot be separately identified from data on hours worked or consumption. We therefore resort to the interpretation of the term $(1 + \sigma)(\gamma_a - \gamma_{a-1})$ as the log change in the consumption equivalence scale between ages $a - 1$ and a due to changes in family size. Given $\sigma = 2$, we estimate the age profile for γ_a from
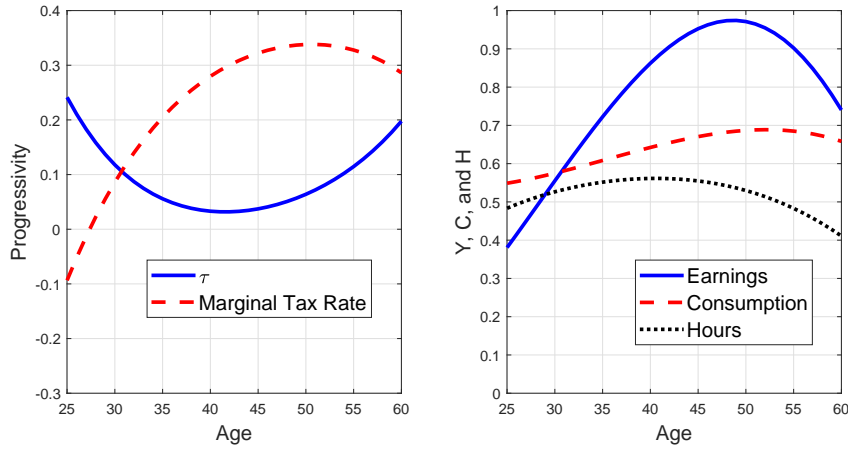
40

Figure 11: Optimal age-dependent progressivity with age variation in the taste for consumption.

Fernandez-Villaverde and Krueger (2011), who compute an average across the most commonly used equivalence scales. Then, given the vector $\{\gamma_a\}$, we estimate $\{\bar{\varphi}_a\}$ residually from the age profile for hours worked, as under the baseline calibration.

The Fernandez-Villaverde and Krueger equivalence scale increases by roughly 25% from age 25 to age 50 and declines moderately thereafter. The implied estimated path for the utility shifter γ_a follows a similar profile. From the viewpoint of the planner who has to set tax rates optimally, older households should now receive higher consumption. Relative to the benchmark case, the planner therefore redistributes less across ages and, as a consequence, chooses flatter profiles for both λ_a and τ_a. This implies smaller potential welfare gains from moving to an age-varying tax system. Figure 11 depicts optimal age-dependent progressivity in this case.

6.5   Extension II: age variation in the Frisch elasticity

In this section, we allow older workers' labor supply to become more elastic to changes in after-tax wages as they approach full retirement.

As reported by Blundell et al. (2016), direct evidence on labor supply elasticities around retirement is scarce. The few estimates cited in their survey suggest that the Frisch elasticity may be up to three times higher at ages 60-65 compared to age 45.

We now explore the impact of higher labor supply elasticities at older ages for the optimal profile of progressivity. We model age-varying elasticities in a simple mechanical fashion, assuming that the Frisch elasticity is constant at a value of 0.5 up to age 45 and then increases linearly to reach 1.5 at age 60. Figure 12 illustrates that this increase in the elasticity eliminates the rise in optimal progressivity after age 45. For example, in the baseline (Figure 6) $\tau_{60}^* = 0.17$, whereas in this extension, it is roughly zero. Thus, this *late-in-life labor supply elasticity channel* is
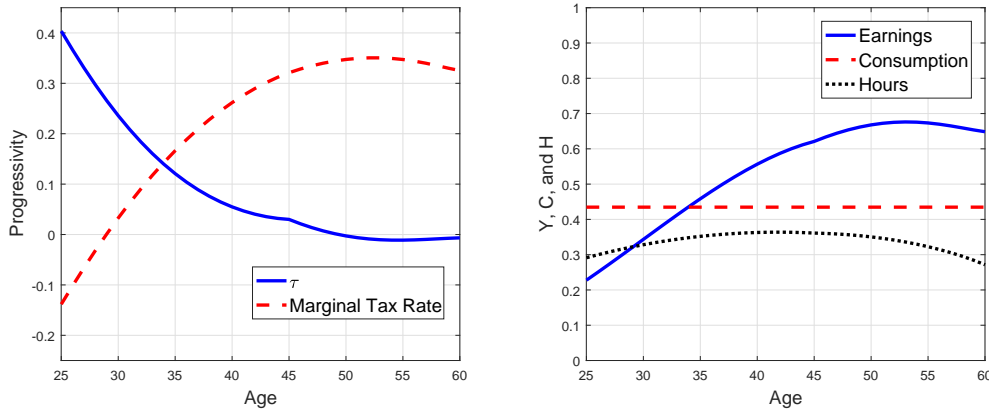
Figure 12: Optimal age-dependent progressivity when the Frisch elasticity of labor supply increases after age 45.

quite powerful in moderating the rise in optimal progressivity for the elderly.

Relative to the baseline model, old age earnings are now higher, reflecting the impact of lower progressivity. The planner therefore has a stronger incentive to reduce λ_a at older ages in order to redistribute to younger households and equate consumption across age groups. While a lower τ_a tends to lower marginal tax rates at older ages, a lower λ_a tends to raise them, with the net result that the average marginal tax rate is basically unchanged relative to the baseline model.

7 Introducing borrowing and lending

The main limitation of the benchmark model is that, to preserve analytical tractability, we shut down borrowing and lending. The risk sharing allowed in the model against insurable shocks offers some private redistribution within age groups, but only the planner can redistribute resources across age groups. Thus, one driver of age variation in optimal taxation is the planner's desire to facilitate intertemporal consumption smoothing.

The key concern is that, if private saving and borrowing were allowed, households would use financial markets to smooth consumption intertemporally, and the life-cycle channel in the design of optimal taxes would therefore be weakened. The extent to which the optimal policy will change will depend on the generosity of borrowing limits.

In this section, we extend the benchmark model by allowing households to trade a risk-free bond in zero net supply, with the interest rate r determined in the stationary equilibrium of the model. At the same time, we shut down insurable risk (i.e., we set $v_{\varepsilon,a} = 0$). In this model, wealth b is a state variable for the individual, and the steady state features a non-degenerate

wealth distribution. As a result, the equilibrium and the optimal age-dependent tax system have to be computed numerically. The optimal tax problem is rather complicated since in principle one has to choose a vector of $A = 36$ values for τ_a, one for each age, in order to maximize equilibrium welfare.

Having introduced wealth and savings, we now need to decide how to tax them. Exploring the optimal differential taxation of earnings versus savings is beyond the scope of the paper. We therefore simply assume that taxable income at age a includes capital income rb_{ia} but that savings $b_{i,a+1} - b_{ia}$ are tax deductible. Thus, the parametric tax/transfer function now applies to taxable income $p(s_i) \exp(x_a + \alpha_{ia})h_{ia} + rb_{ia} - (b_{i,a+1} - b_{ia})$, and the individual budget constraint becomes

$$c_{ia} = \lambda_a \left[p(s_i) \exp(x_a + \alpha_{ia})h_{ia} + (1+r)b_{ia} - b_{i,a+1} \right]^{1-\tau_a}.$$

This assumption is convenient because it allows us to retain a closed-form solution for the equilibrium skill price function $p(s)$. Note that this tax specification also has the property that the planner effectively taxes consumption in a progressive fashion.

We also need to specify borrowing limits. We assume that the borrowing limit for individual i at age a can be written as

$$b_{i,a+1} \geq -\bar{b}_a \cdot p(s_i) \exp(\alpha_{ia} - \varphi_i),$$

where $\{\bar{b}_a\}$ are age-varying parameters. Note that this specification implies that borrowing limits are proportional to the idiosyncratic components of individual wages and preferences. Every other element of the baseline model is unchanged.

The dynamic program of a working-age household characterizing optimal consumption/saving and labor supply decisions now has five state variables: age a, skills s, disutility of work φ, productivity α, and wealth b. Given the form of the borrowing constraint, however, it is possible to characterize optimal individual saving and labor supply decisions by solving a simpler household problem with only two states: age a and normalized wealth, defined as wealth b relative to the adjustment factor $p(s) \exp(\alpha - \varphi)$ (see Appendix B for details).

7.1   Parameterization

The parameterization is the same as the one in Table 1, with the exception that the variances of the insurable risk terms are zero at each age. The discount factor is fixed at $\beta = 1$. The only new parameters are the age-dependent borrowing limits. When the borrowing limit parameters $\{\bar{b}_a\}$ are set to zero at all ages, the wealth distribution is degenerate at zero, since assets are in zero net supply. In this case, the equilibrium coincides with the one of the benchmark model (modulo the absence of insurable risk). The loosest possible limits are natural borrowing constraints: in

43

this case, the only binding constraint over the life cycle is $b_{iA} \geq 0$, that is, a no-Ponzi condition stating that the household cannot die with negative wealth.

To set borrowing limits that realistically capture how much *consumer credit* households can access, we adopt the following strategy. We use cross-sectional data from the Survey of Consumer Finances (SCF) for the year 2012 (2013 survey) for households aged 25-60, as in the model. The SCF has information on credit limits on all credit cards and on home equity lines of credit. We begin by adding up all these limits.

The SCF also contains information on the residual value of existing installment loans for vehicles, boats, and other durables, and on residual values of other loans, such as borrowing against IRAs. We multiply the value of these loans by a factor of two to reflect the fact that, on average, households are halfway through their repayment.

We sum up these two numbers obtained from credit limits and existing loans, and express this total borrowing limit as a fraction of household labor income. This approach suggests that almost 20% of US households have a credit limit of zero. Conditional on borrowing, the median credit limit is about half annual earnings. However, it is plausible that some households could access additional credit if they wanted to do so. With this in mind, and in the interests of exploring a parameterization in which fairly extensive borrowing and saving is possible, we set the age-specific credit limit to 1.5 times annual household earnings at the corresponding age, which corresponds to roughly the 90th percentile of the distribution of our estimated credit limits.

Finally, to keep the Ramsey problem of the government manageable, instead of optimizing over the full vector of τ_a for each age, we approximate the τ_a function with a Chebyshev polynomial of order two and optimize over its three coefficients.[23] Moreover, we assume that the planner maximizes steady-state welfare, which is reasonable given our assumption that $\beta = 1$.

7.2 Results

Figure 13 plots age profiles for wages, hours worked, and consumption (left panel) and for wealth (right panel) in the economy with τ^{US} and with the US credit limit. Overall, these paths are not much different from those in our baseline economy (recall Figure 3), which is reassuring. The age profile for wealth illustrates that young households borrow extensively against future earnings. After age 45, when the productivity profile levels off, they become net savers and start lending to the young.

Figure 14 summarizes our findings on the optimal tax scheme in this extended economy with intertemporal trade. The top three panels consider the case in which the gross interest rate

---

[23]We have verified that polynomials of order three yield only negligible additional welfare gains.
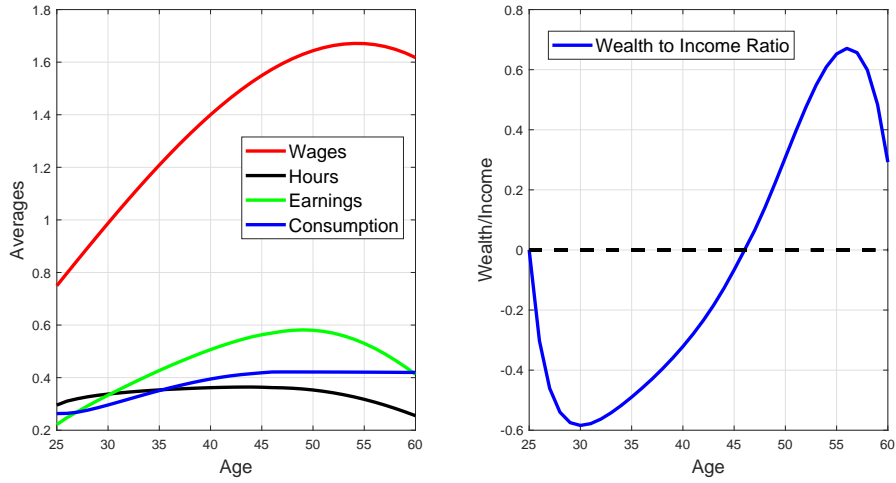
Figure 13: Left panel: means of wages, earnings, consumption, and hours worked over the life cycle. Right panel: average wealth-income ratio over the life cycle.

is fixed exogenously at $R = 1$. The bottom three panels consider the case in which the interest rate is endogenous and adjusts to clear the market for bonds. In each case, three parameterizations are plotted: one "risk only" in which the life-cycle profiles $\{x_a\}$ and $\{\bar{\varphi}_a\}$ are flat; a second "life cycle only" in which the variance of life-cycle productivity shocks is set to zero; and a third "risk + life cycle" which corresponds to the baseline parameterization. Each panel has three lines, corresponding to three alternative assumptions on the scope for borrowing: zero borrowing, the calibrated borrowing limit for the US economy, and the natural borrowing limit.[24]

Risk only. The left panels of Figure 14 illustrate that, absent systematic life-cycle earnings variation, allowing for intertemporal borrowing and lending has little impact on the optimal age profile for progressivity. In each case, the optimal profile for τ_a is driven by the uninsurable risk channel, which calls for progressivity to increase with age.

Life cycle only. Consider now the models in which there are no productivity shocks (middle panels). Recall from Proposition 4 that absent idiosyncratic risk (either uninsurable or insurable), the only motive for introducing age variation in the tax system in the autarkic version of the economy was to enable the planner to equate consumption across age groups. But when households can borrow and lend, they will use the bond to consumption-smooth predictable life-cycle variation in earnings. In fact, if R is fixed exogenously and equal to $1/\beta = 1$ (middle panel in the top row of Figure 14), then given a sufficiently loose borrowing limit, agents will use the bond to achieve constant consumption over the life cycle, absent any age variation in

---

[24]In the endogenous interest rate economy, given that bonds are in zero net supply, ruling out borrowing also implies no saving in equilibrium. In the exogenous interest rate economy, we rule out borrowing, but allow saving.

Figure 14: Economy with intertemporal trade. Left panels: uninsurable risk channel only. Middle panels: life-cycle channel only. Right panels: both channels active. In each panel, the three lines correspond to a zero borrowing limit, an ad hoc borrowing limit estimated from SCF data, and the natural borrowing limit. In the top row, the interest rate is set exogenously so that $R = \beta^{-1} = 1$. In the bottom row, the interest rate clears the bond market.

the return to saving induced by age variation in tax parameters. Because a constant consumption profile is exactly what the planner wants, the planner has no incentive to introduce any such age variation. This explains why, in the natural borrowing limit case, the profile for τ_a^* is flat.

When we endogenize the interest rate, the optimal τ_a profile in the same specification (life cycle only, natural borrowing limit –the middle panel in the bottom row) looks quite different and now declines significantly over the life cycle. This reflects a new channel mediating the optimal age profile for progressivity, which we label the *interest rate channel*. Under the natural borrowing limit, every agent is on her Euler equation throughout the life cycle. Given our calibrated parameter values – and in particular, a generally increasing age profile for average earnings – the equilibrium market-clearing interest rate is positive and thus exceeds the house-

46

hold's rate of time preference. It follows that, absent age variation in either λ_a or τ_a, households would choose positive consumption growth over the life cycle. However, as in every economy we have considered, the planner wants to equate consumption across age groups. To achieve this, it must choose profiles for λ_a and τ_a with the property that the after-tax interest rate is reduced to zero at each age. It can achieve this by choosing a declining path for λ_a, that depresses after-tax returns. Given a decreasing profile for λ_a, it is also optimal to have τ_a decrease in age in order to avoid a rising labor wedge.

When borrowing and lending are ruled out, the optimal profile for progressivity in the life-cycle-only economy is driven entirely by the life-cycle channel and is U-shaped and mirrors (inversely) the efficiency net of work disutility profile. Under the US borrowing limits, the progressivity profiles are intermediate between those under the autarky and natural borrowing constraint extremes. The optimal profile in the endogenous interest rate case is more downward sloping than when the interest rate is fixed at zero because the interest rate channel is operative in this case.

Risk + life cycle. When both the risk and life-cycle channels are present (right panels), as in the baseline model specification plotted in the right panels the optimal profile for τ_a is always a compromise, roughly averaging the profiles dictated by the uninsurable risk channel (left panels) and the life-cycle and interest rate channels (middle panels).

In autarky, the age profiles for τ_a^* essentially coincide with the one in our benchmark model (Figure 6, top left panel), modulo the absence of insurable risk in this parameterization. Under the calibrated US borrowing limit, the optimal profiles are remarkably close to those in autarky, suggesting that our previous analysis of the tractable autarkic case offers qualitatively and quantitatively relevant guidance on how taxation should vary by age. At the same time, when we impose very loose borrowing limits, the optimal policies look quite different from autarky and vary dramatically between the exogenous and endogenous interest rate cases.

7.3 Retirement Saving

One may be concerned that optimal progressivity in the economy calibrated to US credit limits is similar to that in autarky because we have abstracted from retirement. With a retirement phase in the life-cycle, the saving motive during working life might be stronger, which in turn might weaken the importance of borrowing constraints.

We therefore add a retirement phase to the model, which we solve in partial equilibrium ($\beta R = 1$). We assume exogenous retirement at age $A - 1$, corresponding to age 60, and extend the life cycle by 20 years to age 80. During this retirement period, each worker i is entitled to a pension proportional to a proxy for lifetime earnings given by $y_i^R = p(s_i) \exp(x_{A-1} + \alpha_{i,A-1} - $
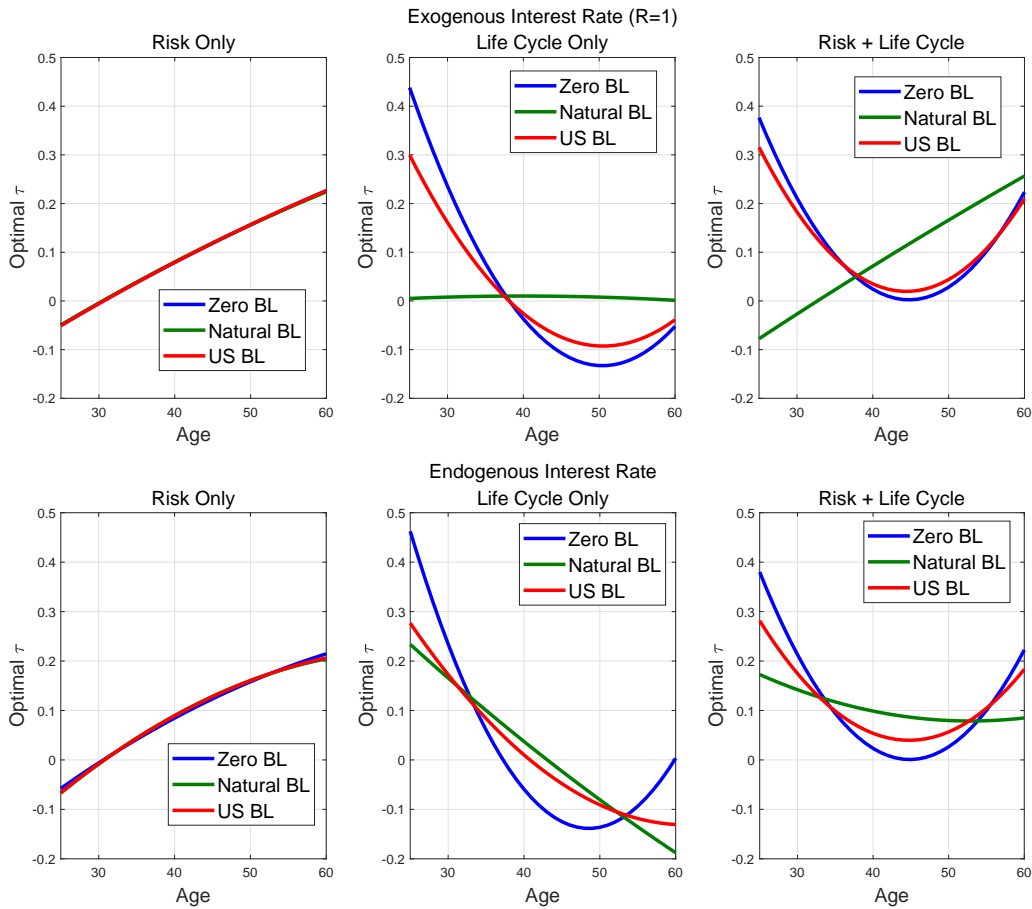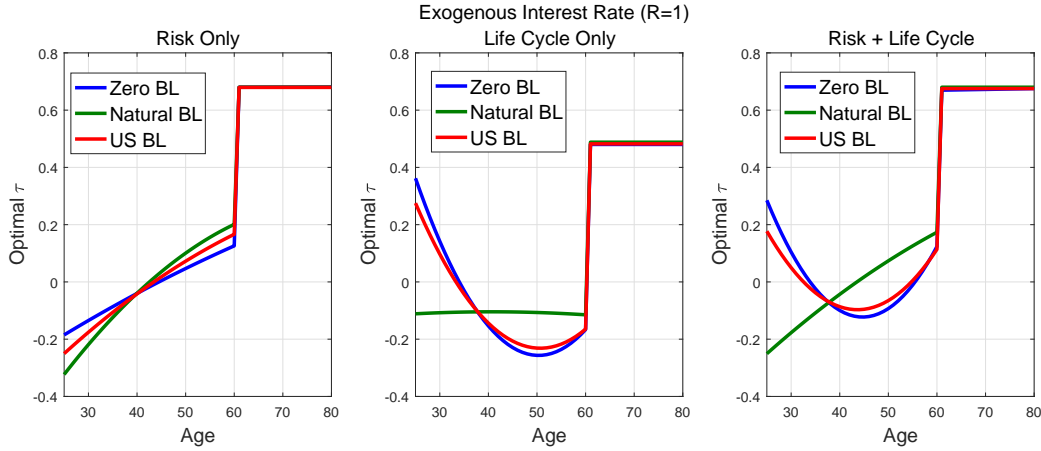
Figure 15: Economy with intertemporal trade and a retirement period. Left panel: uninsurable risk channel only. Middle panel: life-cycle channel only. Right panel: both channels active. In each panel, the three lines correspond to a zero borrowing limit, an ad hoc borrowing limit estimated from SCF data, and the natural borrowing limit. The interest rate is set exogenously so that $R = \beta^{-1} = 1$.

$\bar{\varphi}_{A-1} - \varphi_i$). This proxy is proportional to the worker's productivity at the end of her career and, through the disutility of work terms, it also contains a proxy for average hours over the working life (recall Corollary 2.2).[25]

The government takes this pension function as given and chooses how to tax pensions, subject to the same functional form for the tax schedule that applies to individuals of working age. Thus, the individual during retirement receives after-tax pension income $\lambda_a(y_i^R)^{1-\tau_a}$.[26]

Figure 15 illustrates this case. First, note that the profile for progressivity is flat during retirement because there is no active source of age dependence: retirees face no risk, and the pension baseline y_i^R is constant. Because there are no distortions to labor supply during retirement, one might be tempted to conjecture that the optimal τ_a in retirement would approach one. However, tax progressivity in retirement still disincentivizes skill investment, since progressivity diminishes the payoff from higher skills in terms of higher pension income. The higher is progressivity in retirement, the smaller are the marginal welfare gains from additional consumption compression, while the disincentive effects on skill investment are linear in τ_a. Thus, as Figure 15 illustrates, it is not optimal to push progressivity to the maximum possible value at retirement: τ_a jumps discretely as labor supply distortions disappear but stays well below one.

During working life, the optimal path for τ_a retains the U shape of the one in the economy

---

[25]While pensions depend on idiosyncratic preference and productivity components, we maintain our assumption that net taxes before retirement are a function only of labor earnings y.

[26]In this extension, we set $\chi = 0$. Given this choice, pension outlays are similar in size to government consumption in the baseline model. Recall that while the value for χ affects the level of optimal progressivity, it does not induce age variation in optimal progressivity.

without retirement (compare Figure 15 to the top panel of Figure 14). However, the overall level of progressivity is now lower. Thus the planner offsets the disincentive effects on skill investment arising from high values for τ_a in the retirement phase of life by choosing lower values for τ_a during the working phase of life.

7.4 Welfare from tax reform with intertemporal trade

The last two columns of Table 2 summarize the welfare gains from switching to the optimal tax system under the US and natural borrowing limits in the endogenous interest rate case. Under the US borrowing limit, the gains from moving to an age-dependent tax system are somewhat smaller than in the autarky case but are still substantial at around 1.8% of lifetime consumption. Under the natural borrowing limit, welfare gains shrink further to about half of the gains in autarky. As expected, the bulk of these gains can be obtained even with an age-invariant λ.

8 Conclusions

This paper has developed an equilibrium framework to study the optimal degree of progressivity in the tax and transfer system over the life cycle. Building on Heathcote et al. (2017), the main innovation in this paper is to allow for age-dependent tax progressivity. When calibrating the analytically tractable economy without intertemporal trade to the US we find that the optimal age profile for tax progressivity is U-shaped, while the average marginal tax is increasing and concave in age. This U shape survives, but is moderated, in a more realistic version of the model with borrowing and saving and calibrated credit constraints. In this economy, the welfare gains of switching from the current age-invariant tax and transfer system to the optimal age-dependent system are around 1.8% of lifetime consumption. The optimal system and the corresponding welfare gains are sensitive to the extent of age variation in the taste for consumption and leisure, and to the pattern of age variation in the elasticity of labor supply.

We conclude the paper with a short discussion of elements not included in our model that may be relevant for age dependency in the optimal tax system.

One shortcoming of our analysis is the assumption that the age-wage profile is exogenous. Various authors have argued that the age-wage profile is instead endogenous, and reflects human capital investments and learning-by-doing over the life cycle (Ben Porath, 1967; Imai and Keane, 2004). Such endogeneity could potentially have interesting implications for optimal taxation (Keane, 2011).

Our model with borrowing and lending integrates taxation of wealth within our baseline

two parameter tax and transfer system. Introducing a separate and potentially age-varying tax on savings would give the planner direct control over the return to saving. This additional instrument would weaken the interest rate channel, a force that induces progressivity to decline with age. An even more ambitious direction for future research would be to introduce capital explicitly as an input to production, and to explore a comprehensive joint analysis of optimal labor and capital taxation.

To maintain tractability we modeled the cost of skill acquisition as a utility cost. However, some of the costs of acquiring skills are in the form of time or money. A richer model of skill investment costs would better situate the model to address policy issues involving trade-offs between tax progressivity and education subsidies (see e.g. Bovenberg and Jacobs 2005; Krueger and Ludwig 2016; Stantcheva 2017).

Finally, one could enrich our model of the life cycle and to address the extent to which age variation in marriage rates and fertility choices calls for age variation in the optimal tax and transfer system. Later in the life-cycle, age variation in taxation can potentially substitute for missing private markets against longevity risk in the spirit of Hosseini and Shourideh (2017).

While all these extensions will add new trade-offs and possibly affect quantitative predictions about optimal taxation, the qualitative forces motivating age variation in tax progressivity we have described will remain salient.

References

AKERLOF, G. A. (1978): "The Economics of 'Tagging' as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning," *American Economic Review*, 68(1), 8–19.

ATKINSON, A. B., T. PIKETTY, AND E. SAEZ (2011): "Top Incomes in the Long Run of History," *Journal of Economic Literature*, 49(1), 3–71.

BANKS, J., AND P. DIAMOND (2010): "The Base for Direct Taxation," in *Dimensions of Tax Design: The Mirrlees Review*, ed. by J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles, and J. Poterba. Oxford University Press, Oxford.

BEN-PORATH, Y. (1967): "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75(4, Part 1), 352–365.

BÉNABOU, R. (2000): "Unequal Societies: Income Distribution and the Social Contract," *American Economic Review*, 90(1), 96–129.

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

BEST, M. C., AND H. J. KLEVEN (2013): "Optimal Income Taxation with Career Effects of Work Effort," Mimeo, Princeton University.

BLUNDELL, R., E. FRENCH, AND G. TETLOW (2016): "Retirement Incentives and Labor Supply," in *Handbook of the Economics of Population Aging*, vol. 1, pp. 457–566. Elsevier.

BOVENBERG, L., AND B. JACOBS (2005): "Redistribution and Education Subsidies are Siamese Twins," *Journal of Public Economics*, 89(11-12), 2005–2035.

CONESA, J. C., S. KITAO, AND D. KRUEGER (2009): "Taxing Capital? Not a Bad Idea after All!," *American Economic Review*, 99(1), 25–48.

EROSA, A., AND M. GERVAIS (2002): "Optimal Taxation in Life-Cycle Economies," *Journal of Economic Theory*, 105(2), 338–369.

FARHI, E., AND I. WERNING (2013): "Insurance and Taxation over the Life Cycle," *Review of Economic Studies*, 80(2), 596–635.

FELDSTEIN, M. S. (1969): "The Effects of Taxation on Risk Taking," *Journal of Political Economy*, 77(5), 755–764.

FERNANDEZ-VILLAVERDE, J., AND D. KRUEGER (2011): "Consumption and Saving over the Life Cycle: How Important are Consumer Durables?," *Macroeconomic Dynamics*, 15(5).

FINDEISEN, S., AND D. SACHS (2016): "Education and Optimal Dynamic Taxation: The Role of Income-Contingent Student Loans," *Journal of Public Economics*, 138(1), 1–21.

GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (2016): "Redistribution and Social Insurance," *American Economic Review*, 106(2), 359–386.

HASSLER, J., P. KRUSELL, K. STORESLETTEN, AND F. ZILIBOTTI (2008): "On the Optimal Timing of Capital Taxation," *Journal of Monetary Economics*, 55(4), 692–709.

HEATHCOTE, J., K. STORESLETTEN, AND G. L. VIOLANTE (2008): "Insurance and Opportunities: A Welfare Analysis of Labor Market Risk," *Journal of Monetary Economics*, 55(3), 501–525.

———— (2014): "Consumption and Labor Supply with Partial Insurance: An Analytical Framework," *American Economic Review*, 104(7), 2075–2126.

——— (2017): "Optimal Tax Progressivity: An Analytical Framework," *Quarterly Journal of Economics*.

HEATHCOTE, J., AND H. TSUJIYAMA (2016): "Optimal Income Taxation: Mirrlees Meets Ramsey," *Mimeo*.

HOSSEINI, R., AND A. SHOURIDEH (2017): "Retirement Financing: An Optimal Reform Approach," Mimeo, CMU.

IMAI, S., AND M. P. KEANE (2004): "Intertemporal Labor Supply and Human Capital Accumulation," *International Economic Review*, 45(2), 601–641.

JAKOBSSON, U. (1976): "On the Measurement of the Degree of Progression," *Journal of Public Economics*, 5(1-2), 161–168.

KAKWANI, N. C. (1977): "Measurement of Tax Progressivity: An International Comparison," *Economic Journal*, 87(345), 71–80.

KAPIČKA, M. (2015): "Optimal Mirrleesian Taxation in a Ben-Porath Economy," *American Economic Journal: Macroeconomics*, 7(2), 219–248.

KAPLOW, L. (2004): "On the (Ir) relevance of Distribution and Labor Supply Distortion to Government policy," *Journal of Economic Perspectives*, 18(4), 159–175.

KARABARBOUNIS, M. (2016): "A Road Map for Efficiently Taxing Heterogeneous Agents," *American Economic Journal: Macroeconomics*, 8(2), 182–214.

KEANE, M. P. (2011): "Labor Supply and Taxes: A Survey," *Journal of Economic Literature*, 49(4), 961–1075.

KRUEGER, D., AND A. LUDWIG (2016): "On the Optimal Provision of Social Insurance: Progressive Taxation versus Education Subsidies in General Equilibrium," *Journal of Monetary Economics*, 77, 72–98.

MUSGRAVE, R. A. (1959): *The Theory of Public Finance: A Study in Public Economy*. New York: McGraw-Hill.

NDIAYE, A. (2017): "Flexible Retirement and Optimal Taxation," Mimeo, Northwestern University.

STANTCHEVA, S. (forthcoming): "Optimal Taxation and Human Capital Policies over the Life Cycle," *Journal of Political Economy*.

STIGLITZ, J. E. (1985): "Equilibrium Wage Distributions," *Economic Journal*, 95(379), 595–618.

WEINZIERL, M. (2011): "The Surprising Power of Age-Dependent Taxes," *Review of Economic Studies*, 78(4), 1490–1518.

APPENDIX

A   Proofs

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

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

We only sketch this proof, since it follows the ones in Heathcote et al. (2014, 2017), which contain more comprehensive versions. We solve the model by segmenting production on "islands" indexed by age a and by the uninsurable triplet (φ, α, s). The (a, φ, α, s) island planner's problem, taking the island-specific skill prices $p(s, \bar{\tau})$ and the aggregate fiscal variables $(G, \{\lambda_a\}, \{\tau_a\})$ as given, is

$$
\max_{\{c_a, h_a\}} \int \left\{ \log c_a - \frac{\exp\left[(1+\sigma)(\varphi + \bar{\varphi}_a)\right]}{1+\sigma} h_a(\varepsilon)^{1+\sigma} + \chi \log G \right\} dF_\varepsilon
$$

subject to the island-level resource constraint (the equivalent of the no-bond-trading assumption):

$$
c_a = \lambda_a \int \exp\left[(1 - \tau_a)(p(s, \bar{\tau}) + x_a + \alpha_a + \varepsilon)\right] h_a(\varepsilon)^{1-\tau_a} dF_\varepsilon.
$$

The first-order conditions with respect to c_a and $h_a(\varepsilon)$ are, respectively,

$$
c_a^{-1} = M
$$

$$
\exp\left[(1+\sigma)(\varphi + \bar{\varphi}_a)\right] h(\varepsilon)^\sigma = M\lambda_a(1-\tau_a)\exp\left((p(s, \bar{\tau}) + x_a + \alpha)(1-\tau_a)\right)\exp\left(\varepsilon(1-\tau_a)\right) h(\varepsilon)^{-\tau_a}
$$

where M is the multiplier on the island resource constraint. Combining the two conditions gives

$$
h(\varepsilon) = c_a^{-\frac{1}{\sigma+\tau_a}} (\lambda_a(1-\tau_a))^{\frac{1}{\sigma+\tau_a}} \exp\left(-\frac{(1+\sigma)}{(\sigma+\tau_a)}(\varphi + \bar{\varphi}_a)\right) \exp\left((p(s, \bar{\tau}) + \alpha + x_a + \varepsilon)\frac{(1-\tau_a)}{(\sigma+\tau_a)}\right) \quad \text{(A1)}
$$

Note that from the first-order conditions, c_a is the same for all agents on the island, and as such it cannot depend on ε. Using this fact, and substituting (A1) into the planner's island-specific resource

constraint, yields

$$
\begin{aligned}
c_a &= \lambda_a c_a^{-\frac{1-\tau_a}{\sigma+\tau_a}} \left(\lambda_a \left(1 - \tau_a\right)\right)^{\frac{1-\tau_a}{\sigma+\tau_a}} \exp\left(-\frac{\left(1 - \tau_a\right)\left(1 + \sigma\right)}{\left(\sigma + \tau_a\right)} \left(\varphi + \bar{\varphi}_a\right)\right) \cdot \\
&\quad \int \exp\left[\left(1 - \tau_a\right)\left(p(s, \bar{\tau}) + x_a + \alpha_a + \varepsilon\right)\right] \left[\exp\left(\left(p(s, \bar{\tau}) + \alpha + x_a + \varepsilon\right)\frac{\left(1 - \tau_a\right)^2}{\left(\sigma + \tau_a\right)}\right)\right] dF_\varepsilon.
\end{aligned}
$$

After a few steps of algebra, one obtains the expression for allocations in Proposition 1.

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

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

$$
v'\left(s\right) = \left(\frac{s}{\kappa}\right)^{\frac{1}{\psi}} = \mathbb{E}_0 \left(\frac{1 - \beta}{1 - \beta^A}\right) \sum_{a=0}^{A-1} \beta^a \frac{\partial u\left(c\left(\varphi, \alpha, s; \lambda_a, \tau_a, \bar{\tau}\right), h\left(\varphi; \tau_a\right), g\right)}{\partial s}. \tag{A2}
$$

The skill level s affects only the consumption allocation (not the hours allocation) and only through the skill price p. With some abuse of notation, we denote the skill price as a function of the entire sequence $\{\tau_a\}$; $p\left(s; \{\tau_a\}\right)$ and we later show that p is a function of only s and $\bar{\tau}$. Hence, using (18), (A2) can be simplified as

$$
\left(\frac{s}{\kappa}\right)^{\frac{1}{\psi}} = \sum_{a=0}^{A-1} \beta^a \left(1 - \tau_a\right) \frac{\partial \log p\left(s; \{\tau_a\}\right)}{\partial s}.
$$

We now guess that the skill price function is log-linear in the skill choice,

$$
\log p\left(s; \{\tau_a\}\right) = \pi_0(\{\tau_a\}) + \pi_1(\{\tau_a\}) \cdot s, \tag{A3}
$$

which implies that the skill allocation has the form[27]

$$
s(\kappa; \{\tau_a\}) = \left[\pi_1(\{\tau_a\}) \cdot \left(1 - \bar{\tau}\right)\right]^\psi \cdot \kappa, \tag{A4}
$$

where $\bar{\tau}$ can be interpreted as a discounted expected progressivity rate,

$$
\bar{\tau} \equiv \left(\frac{1 - \beta}{1 - \beta^A}\right) \sum_{a=0}^\infty \beta^a \tau_a.
$$

---

[27]To see this, note that per assumption $\partial \log p\left(s; \{\tau_a\}\right) / \partial s = \pi_1(\{\tau_a\})$, so (A2) can be written as

$$
\begin{aligned}
\left(\frac{s}{\kappa}\right)^{\frac{1}{\psi}} &= \left(1 - \beta\delta\right) \sum_{a=0}^\infty \left(\beta\delta\right)^a \left(1 - \tau_a\right) \pi_1(\{\tau_a\}) \\
&= \pi_1(\{\tau_a\}) \left(1 - \left(1 - \beta\delta\right) \sum_{a=0}^\infty \left(\beta\delta\right)^a \tau_a\right).
\end{aligned}
$$

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

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

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

$$\log p(s, \bar{\tau}) = \log \left[\frac{\partial Y}{\partial [\bar{N} \cdot m(s)]} \right] = \frac{1}{\theta} \log Y - \frac{1}{\theta} \log [\bar{N} \cdot m(s)] \qquad (A5)$$

$$= \frac{1}{\theta} \log \left(\frac{Y}{\bar{N}} \right) - \frac{1}{\theta} \log \zeta + \frac{\zeta}{\theta} s.$$

Equating coefficients across equations (A3) and (A5) implies $\pi_1(\{\tau_a\}) = \frac{\zeta}{\theta} = \frac{\eta}{\theta} \left[(1 - \bar{\tau}) \pi_1(\{\tau_a\})\right]^{-\psi}$, which yields

$$\pi_1(\{\tau_a\}) = \left(\frac{\eta}{\theta} \right)^{\frac{1}{1+\psi}} (1 - \bar{\tau})^{-\frac{\psi}{1+\psi}} \qquad (A6)$$

and thus the equilibrium density of s is

$$m(s) = (\eta)^{\frac{1}{1+\psi}} \left(\frac{\theta}{1 - \bar{\tau}} \right)^{\frac{\psi}{1+\psi}} \exp \left(- (\eta)^{\frac{1}{1+\psi}} \left(\frac{\theta}{1 - \bar{\tau}} \right)^{\frac{\psi}{1+\psi}} s \right). \qquad (A7)$$

Similarly, the base skill price is

$$\pi_0(\{\tau_a\}) = \frac{1}{\theta} \log \left(\frac{Y}{\bar{N}} \right) - \frac{\log \left(\frac{\eta}{\theta} \right)}{\theta(1 + \psi)} + \frac{\psi}{\theta(1 + \psi)} \log(1 - \bar{\tau}). \qquad (A8)$$

We derive a fully structural expression for $\pi_0(\{\tau_a\})$ below in the proof of Corollary 2.2 when we solve for Y and $\bar{N}$ explicitly. From now on, we drop the vector notation $\{\tau_a\}$ and simply express the equilibrium functions as functions of $\bar{\tau}$, (i.e., $s(\kappa, \bar{\tau})$, $\pi_1(\bar{\tau})$, and $\pi_0(\bar{\tau})$).

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

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

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

where the first equality uses (A4), and the second equality follows from (A6). Thus, log skill premia are exponentially distributed with parameter θ. The variance of log skill prices is

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

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

A.4 Proof of Corollary 2.2 [aggregate quantities]

From eq. (17) and the assumption that φ and ε are independent, aggregate hours worked by individuals of age a are

$$
\begin{aligned}
H\left(a,\tau_a\right) &= \mathbb{E}\left[h\left(\varphi,\varepsilon,a;\tau_a\right)\right] = \int\int h\left(\varphi,\varepsilon,a;\tau_a\right) dF\left(\varphi\right) dF_a\left(\varepsilon\right) && \text{(A9)} \\
&= \exp\left(\frac{\log(1-\tau_a)}{(1+\hat{\sigma}_a)\left(1-\tau_a\right)} - \frac{(1-\tau_a(1+\hat{\sigma}_a))}{(\hat{\sigma}_a)^2} \cdot \frac{v_{\varepsilon,a}}{2}\right) && \text{(A10)} \\
&\quad \cdot \int \exp\left(\frac{\varepsilon}{\hat{\sigma}_a}\right) dF_a\left(\varepsilon\right) \int \exp\left(-\left(\varphi + \bar{\varphi}_a\right)\right) dF\left(\varphi\right) \\
&= \left(1-\tau_a\right)^{\frac{1}{1+\sigma}} \cdot \exp\left(-\bar{\varphi}_a\right) \cdot \exp\left[\left(\frac{\tau_a\left(1+\hat{\sigma}_a\right)}{\hat{\sigma}_a^2} - \frac{1}{\hat{\sigma}_a}\right)\frac{v_{\varepsilon,a}}{2}\right].
\end{aligned}
$$

Since α, ε, and φ are independent, it follows that $N\left(a,\tau_a\right) = \exp(x_a) \cdot \mathbb{E}\left[\exp(\alpha)\right] \cdot \mathbb{E}\left[\exp(\varepsilon)h\left(\varphi,\varepsilon,a;\tau_a\right)\right] = \exp\left(x_a\right)\left(1-\tau_a\right)^{\frac{1}{1+\sigma}}\exp\left[\left(\frac{\tau_a(1+\hat{\sigma}_a)}{\hat{\sigma}_a^2} + \frac{1}{\hat{\sigma}_a}\right)\frac{v_{\varepsilon,a}}{2}\right]$, and therefore

$$N\left(a,\tau_a\right) = \exp(x_a + \frac{v_{\varepsilon,a}}{\hat{\sigma}_a}) \cdot H\left(a,\tau_a\right). \tag{A11}$$

Finally, average output of age group a is given by

$$
\begin{aligned}
Y\left(a,\tau_a,\bar{\tau}\right) &= \mathbb{E}\left(y\left(\varphi,\alpha,\varepsilon,a;\tau_a,\bar{\tau}\right)\right) = \mathbb{E}\left[p\left(s;\bar{\tau}\right)\exp\left(x_a + \alpha\right)h\left(\varphi,\varepsilon,a;\tau_a\right)\right] \\
&= \mathbb{E}\left[p\left(s;\bar{\tau}\right)\right] \cdot N\left(a,\tau_a\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{E}\left[p\left(s;\bar{\tau}\right)\right] &= \mathbb{E}\left[\exp\left(\pi_0\left(\bar{\tau}\right) + \pi_1\left(\bar{\tau}\right)s\right)\right] \\
&= \exp\left(\pi_0\left(\bar{\tau}\right)\right) \cdot \mathbb{E}\left\{\exp\left(\left(\frac{\eta}{\theta}\right) \cdot \kappa\right)\right\} = \exp\left(\pi_0\left(\bar{\tau}\right)\right)\frac{\theta}{\theta-1}.
\end{aligned}
$$

Thus:

$$
\begin{aligned}
Y\left(a, \tau_a, \bar{\tau}\right) \quad = \quad & \left[\left(\frac{\theta}{\theta-1}\right)^{\frac{\theta}{\theta-1}}\left(\frac{1-\bar{\tau}}{\theta}\right)^{\frac{\psi}{(1+\psi)(\theta-1)}}\left(\frac{1}{\eta}\right)^{\frac{1}{(1+\psi)(\theta-1)}}\right] \\
& \cdot\left(1-\tau_a\right)^{\frac{1}{1+\sigma}} \exp\left[\left(x_a-\bar{\varphi}_a\right)+\left(\frac{\tau_a\left(1+\hat{\sigma}_a\right)}{\hat{\sigma}_a^2}+\frac{1}{\hat{\sigma}_a}\right) \frac{v_{\varepsilon, a}}{2}\right].
\end{aligned}
$$

A.5   Proof of Proposition 3 [optimal choice of g and $\{\breve{}_a\}$]

It is useful to begin by computing

$$
\begin{aligned}
\tilde{Y}\left(a, \tau_a, \bar{\tau}\right) := \quad & \int\left(y_{i, a}\right)^{1-\tau_a} di = \mathcal{K}\left(a, \tau_a, \bar{\tau}\right) \\
& \cdot \exp\left(-\tau_a\left(1-\tau_a\right) a \frac{v_\omega}{2}+\frac{\left(1-\tau_a\right)\left(1-\tau_a\left(1+\hat{\sigma}_a\right)\right)}{\hat{\sigma}_a} \frac{v_{\varepsilon, a}}{2}\right)
\end{aligned}
$$

where, after some tedious algebra, one obtains

$$
\begin{aligned}
\mathcal{K}\left(a, \tau_a, \bar{\tau}\right) \quad = \quad & \left(1-\tau_a\right)^{\frac{1-\tau_a}{1+\sigma}} \exp\left(\left(1-\tau_a\right)\left(x_a-\bar{\varphi}_a\right)-\tau_a\left(1-\tau_a\right) \frac{v_\varphi}{2}\right) \\
& \cdot(1-\bar{\tau})^{\frac{\psi}{1+\psi} \frac{1-\tau_a}{\theta-1}}\left(\frac{1}{\theta-1}\right)^{\frac{1-\tau_a}{\theta-1}} \cdot\left(\frac{\theta}{\eta}\right)^{\frac{1-\tau_a}{(1+\psi)(\theta-1)}} \cdot \frac{\theta}{\theta+\tau_a-1}.
\end{aligned}
$$

It is also useful, for what follows, to define the shorthand notation

$$
\begin{aligned}
\bar{u}\left(a, \lambda_a, \tau_a\right) \quad &: \quad = u\left(c\left(\varphi, s, a, \alpha\right), h\left(\varphi, a, \varepsilon\right)\right), \\
\bar{v}\left(\bar{\tau}\right) \quad &: \quad = \mathbb{E}\left[v\left(s(\kappa, \bar{\tau}), \kappa\right)\right].
\end{aligned}
$$

Recall that welfare in steady state is given by

$$
\mathcal{W}^{ss}\left(g,\left\{\lambda_a, \tau_a\right\}\right)=\frac{1}{A} \sum_{a=0}^{A-1} \mathbb{E}\left[u\left(c\left(\varphi, s, a, \alpha\right), h\left(\varphi, a, \varepsilon\right), G\right)\right]-\mathbb{E}\left[v\left(s(\kappa, \bar{\tau}), \kappa\right)\right].
$$

Thus, the Ramsey planner's problem can be written as

$$\max_{\{g,\lambda_a,\tau_a\}} \mathcal{W}^{ss}(g, \{\lambda_a, \tau_a\}) = \frac{1}{A} \sum_{a=0}^{A-1} \bar{u}(a, \lambda_a, \tau_a) + \chi \log \left(g \sum_{a=0}^{A-1} Y(a, \tau_a, \bar{\tau}) \right) - \bar{v}(\bar{\tau})$$

subject to $\qquad\qquad$ (A12)

$$\frac{1}{A} \sum_{a=0}^{A-1} \lambda_a \tilde{Y}(a, \tau_a, \bar{\tau}) = (1-g) \frac{1}{A} \sum_{a=0}^{A-1} Y(a, \tau_a, \bar{\tau}).$$

Letting ϑ denote the multiplier on the government budget constraint, and recognizing that $\partial \bar{u}(a, \lambda_a, \tau_a) / \partial \lambda_a = \lambda_a^{-1}$ from (18), the first-order condition with respect to λ_a yields

$$\frac{1}{\lambda_a} = \vartheta \cdot \tilde{Y}(a, \tau_a, \bar{\tau}).$$ (A13)

Since $C_a = \lambda_a \tilde{Y}(a, \tau_a, \bar{\tau})$, this first order condition implies that average consumption is equalized across ages:

$$\vartheta^{-1} = C = (1-g) \frac{1}{A} \sum_{a=0}^{A-1} Y(a, \tau_a, \bar{\tau}).$$ (A14)

Consider now the first-order condition with respect to g:

$$\frac{\chi}{g} = \vartheta \frac{1}{A} \sum_{a=0}^{A-1} Y(a, \tau_a, \bar{\tau}).$$

Using ($A14$) in the above equation yields

$$g^* = \frac{\chi}{1+\chi}.$$

A.6   Proof of Proposition 4 [optimal age-dependent progressivity]

To derive the exact analytical expression for the social welfare function in steady state, we analyze each of its components one at a time. The first term in ($A12$) can be written as

$$\bar{u}(a, \lambda_a, \tau_a, \bar{\tau}) = \int \int \int \log c(a, \varphi, \alpha, s; \lambda_a, \tau_a, \bar{\tau}) \, dF_s dF_\alpha^a dF_\varphi$$
$$- \int \int \frac{\exp((1+\sigma)(\varphi + \bar{\varphi}_a)) h(\varphi, \varepsilon, a; \tau_a)^{1+\sigma}}{1+\sigma} dF_\varphi dF_{\varepsilon a}.$$

58

Note that average log consumption for age group a is

$$\mathbb{E}\left[\log c\left(a, \varphi, \alpha_a, s; \lambda_a, \tau_a, \bar{\tau}\right) | a\right]$$
$$= \left\{\mathbb{E}\left[\log c\left(a, \varphi, \alpha_a, s; \lambda_a, \tau_a, \bar{\tau}\right) | a\right] - \log C\left(a, \lambda_a, \tau_a, \bar{\tau}\right)\right\} + \log C\left(a, \lambda_a, \tau_a, \bar{\tau}\right)$$

where

$$\mathbb{E}\left[\log c\left(a, \varphi, \alpha_a, s; \lambda_a, \tau_a, \bar{\tau}\right) | a\right]$$
$$= \log \lambda_a + (1 - \tau_a)\left(-\frac{v_\omega a}{2} - \frac{v_\varphi}{2}\right) + (1 - \tau_a)(x_a - \bar{\varphi}_a) + \frac{1 - \tau_a}{1 + \sigma}\log(1 - \tau_a)$$
$$+ (1 - \tau_a)\frac{(1 - \tau_a(1 + \widehat{\sigma}_a))}{\widehat{\sigma}_a} \cdot \frac{v_{\varepsilon,a}}{2} + (1 - \tau_a)\mathbb{E}\left[\log p\left(s; \bar{\tau}\right)\right]$$

and

$$\mathbb{E}\left[\log p\left(s; \bar{\tau}\right)\right] = \pi_0(\bar{\tau}) + \pi_1\left(\bar{\tau}\right)\mathbb{E}\left[s\right]$$
$$\pi_0(\bar{\tau}) = \frac{\psi}{(1 + \psi)(\theta - 1)}\log\left(1 - \bar{\tau}\right) + \frac{1}{(1 + \psi)(\theta - 1)}\log\left(\frac{\theta}{\eta}\right) + \frac{1}{\theta - 1}\log\left(\frac{1}{\theta - 1}\right)$$
$$\pi_1\left(\bar{\tau}\right)\mathbb{E}\left[s\right] = \left[\left(\frac{\eta}{\theta}\right)^{\frac{1}{1 + \psi}}(1 - \bar{\tau})^{-\frac{\psi}{1 + \psi}}\right]\left[\frac{\eta}{\theta}(1 - \bar{\tau})\right]^{\frac{\psi}{1 + \psi}} \cdot \eta^{-1} = \frac{1}{\theta}.$$

Thus:

$$\mathbb{E}\left[\log c\left(a, \varphi, \alpha_a, s; \lambda_a, \tau_a, \bar{\tau}\right) | a\right]$$
$$= \log \lambda_a - (1 - \tau_a)\left(\frac{v_\omega a}{2} + \frac{v_\varphi}{2}\right) + \frac{1 - \tau_a}{1 + \sigma}\log(1 - \tau_a) + (1 - \tau_a)(x_a - \bar{\varphi}_a) + (1 - \tau_a)\frac{(1 - \tau_a(1 + \widehat{\sigma}_a))}{\widehat{\sigma}_a} \cdot \frac{v_{\varepsilon,a}}{2}$$
$$+ \frac{\psi(1 - \tau_a)}{(1 + \psi)(\theta - 1)}\log\left(1 - \bar{\tau}\right) + \frac{(1 - \tau_a)}{(1 + \psi)(\theta - 1)}\log\left(\frac{\theta}{\eta}\right)$$
$$+ \frac{(1 - \tau_a)}{\theta - 1}\log\left(\frac{1}{\theta - 1}\right) + (1 - \tau_a)\left(\frac{1}{\theta}\right).$$

Moreover:

$$\log C\left(a, \lambda_a, \tau_a, \bar{\tau}\right) = \log \lambda_a - \tau_a(1 - \tau_a)a\frac{v_\omega}{2} + \left((1 - \tau_a)\frac{1 - \tau_a(1 + \widehat{\sigma}_a)}{\widehat{\sigma}_a}\frac{v_\varepsilon}{2}\right)$$
$$+ \frac{1 - \tau_a}{1 + \sigma}\log\left(1 - \tau_a\right) + (1 - \tau_a)(x_a - \bar{\varphi}_a)$$
$$- \tau_a(1 - \tau_a)\frac{v_\varphi}{2} + \frac{(1 - \tau_a)\psi}{(1 + \psi)(\theta - 1)}\log\left(1 - \bar{\tau}\right)$$
$$+ \frac{1 - \tau_a}{\theta - 1}\log\left(\frac{1}{\theta - 1}\right) + \frac{1 - \tau_a}{(1 + \psi)(\theta - 1)}\log\left(\frac{\theta}{\eta}\right) + \log\left(\frac{\theta}{\theta + \tau_a - 1}\right).$$

Therefore, the difference between these two terms is

$$
\begin{aligned}
& \mathbb{E}\left[\log c\left(a, \varphi, \alpha_a, s ; \lambda_a, \tau_a, \bar{\tau}\right) \mid a\right]-\log C\left(a, \lambda_a, \tau_a, \bar{\tau}\right) \\
= & -\left(1-\tau_a\right)^2\left(\frac{v_\omega a}{2}+\frac{v_\varphi}{2}\right)+\frac{1-\tau_a}{\theta}-\log \left(\frac{\theta}{\theta+\tau_a-1}\right) .
\end{aligned}
$$

and combining all these terms gives

$$
\frac{1}{A} \sum_{a=0}^{A-1} \bar{u}_a=\frac{1}{A} \sum_{a=0}^{A-1}\left[-\left(1-\tau_a\right)^2\left(\frac{v_\omega a}{2}+\frac{v_\varphi}{2}\right)+\frac{1-\tau_a}{\theta}-\log \left(\frac{\theta}{\theta+\tau_a-1}\right)+\log C(a)\right] .
$$

Average disutility of hours worked in age group a is

$$
\begin{aligned}
& \int \int \frac{\exp \left((1+\sigma)\left(\varphi+\bar{\varphi}_a\right)\right) h\left(\varphi, \varepsilon, a ; \tau_a\right)^{1+\sigma}}{1+\sigma} d F_\varphi d F_{\varepsilon a} \\
= & \frac{1-\tau_a}{1+\sigma} \int \exp \left((1+\sigma)\left(\varphi+\bar{\varphi}_a\right)\right) \exp \left(-(1+\sigma)\left(\varphi+\bar{\varphi}_a\right)\right) d F_\varphi \\
& \cdot \int\left[\exp \left(-\frac{1+\sigma}{\widehat{\sigma}\left(1-\tau_a\right)} \mathcal{C}_a\right) \exp \left(\frac{1+\sigma}{\widehat{\sigma}} \varepsilon\right)\right] d F_{\varepsilon a} \\
= & \frac{1-\tau_a}{1+\sigma} .
\end{aligned}
$$

The average cost of skill investment in each cohort of newborns is

$$
\bar{v}(\bar{\tau})=\int v(\kappa ; \bar{\tau}) d F_\kappa=\frac{\psi}{1+\psi}\left(\frac{1-\bar{\tau}}{\theta}\right) .
$$

Combining these components, and noting that as $\beta \rightarrow 1$ the constant $(1-\beta) /\left(1-\beta^A\right) \rightarrow 1 / A$, we can rewrite the social welfare function (up to a constant) only as a function of $\left\{\tau_a\right\}$ as

$$
\begin{aligned}
\mathcal{W}^{ss}\left(g,\left\{\tau_a\right\}\right)= & \frac{1}{A} \sum_{a=0}^{A-1} \bar{u}_a-\bar{v}+\chi \log \left(g \sum_{a=0}^{A-1} Y_a\right) \\
= & \frac{1}{A} \sum_{a=0}^{A-1}\left[-\left(1-\tau_a\right)^2\left(\frac{v_\omega a}{2}+\frac{v_\varphi}{2}\right)+\frac{1-\tau_a}{\theta}-\log \left(\frac{\theta}{\theta+\tau_a-1}\right)+\log C(a)\right] \\
& -\frac{1}{A} \sum_{a=0}^{A-1} \frac{1-\tau_a}{1+\sigma}-\left(\frac{\psi}{1+\psi}\right)\left(\frac{1-\bar{\tau}}{\theta}\right)+\chi \log g+\chi \log \sum_{a=0}^{A-1} Y_a \\
= & \frac{1}{A} \sum_{a=0}^{A-1}\left[-\left(1-\tau_a\right)^2\left(\frac{v_\omega a}{2}+\frac{v_\varphi}{2}\right)+\frac{1-\tau_a}{\theta}-\log \left(\frac{\theta}{\theta+\tau_a-1}\right)\right]+\log (1-g) \\
& -\frac{1}{A} \sum_{a=0}^{A-1} \frac{1-\tau_a}{1+\sigma}-\left(\frac{\psi}{1+\psi}\right)\left(\frac{1-\bar{\tau}}{\theta}\right)+\chi \log g+(1+\chi) \log \sum_{a=0}^{A-1} Y_a,
\end{aligned}
$$

where the last step above uses eq. ($A14$), which combines the optimality condition for λ_a (stating that consumption is equalized across ages) and the government budget constraint.

Substituting the expression for Y_a (25) into the above expression for $\mathcal{W}^{ss}(g, \{\lambda_a, \tau_a\})$, we arrive at

$$
\mathcal{W}^{ss}(g, \{\tau_a\}) = \log(1-g) + \chi \log g - \frac{1}{A} \sum_{a=0}^{A-1} \underbrace{\frac{1-\tau_a}{1+\sigma}}_{\text{Disutility of labor}} \tag{A15}
$$

$$
+ (1+\chi) \log \underbrace{\left\{ \sum_{a=0}^{A-1} (1-\tau_a)^{\frac{1}{1+\sigma}} \cdot \exp\left[x_a - \bar{\varphi}_a + \left(\frac{\tau_a(1+\hat{\sigma}_a)}{\hat{\sigma}_a^2} + \frac{1}{\hat{\sigma}_a} \right) \frac{v_{\varepsilon,a}}{2} \right] \right\}}_{\text{Effective hours } \bar{N}_a}
$$

$$
+ (1+\chi) \underbrace{\frac{1}{(1+\psi)(\theta-1)} \left[\psi \log(1-\bar{\tau}) + \log\left(\frac{1}{\eta \theta^\psi} \left(\frac{\theta}{\theta-1} \right)^{\theta(1+\psi)} \right) \right]}_{\text{Productivity: } \log(\text{average skill price}) = \log(E[p(s,\bar{\tau})])}
$$

$$
- \underbrace{\frac{\psi}{1+\psi} \frac{1-\bar{\tau}}{\theta}}_{\text{Avg. education cost}} + \frac{1}{A} \sum_{a=0}^{A-1} \underbrace{\left[\log\left(1 - \left(\frac{1-\tau_a}{\theta} \right) \right) + \left(\frac{1-\tau_a}{\theta} \right) \right]}_{\text{Cost of consumption dispersion across skills}}
$$

$$
- \underbrace{\frac{1}{A} \sum_{a=0}^{A-1} \frac{1}{2} (1-\tau_a)^2 (v_\varphi + a v_\omega)}_{\text{Cons. dispersion due to unins. shocks and preference heterogeneity}} \quad .
$$

Each term in this welfare function has the economic interpretation described under each bracket. For more details, see Heathcote et al. (2017). This welfare expression is only a function of g and $\{\tau_a\}$. The optimal choice of the public good yields $g^* = \chi / (1+\chi)$, which proves statement (i) of the proposition. Substituting this optimal choice back into ($A15$) yields an expression for welfare that is only a function of the sequence $\{\tau_a\}$. Given the sequence of optimal age-dependent progressivity obtained from maximizing ($A15$), the optimal sequence of $\{\lambda_a\}$ can be recovered residually from ($A13$).

Taking the first-order condition of ($A15$) with respect to τ_a (i.e., setting $\frac{\partial \mathcal{W}^{ss}}{\partial \tau_a} = 0$), we arrive at eq. ($28$) in the main text. Standard algebra yields the second-order condition

$$
\frac{\partial^2 \mathcal{W}^{ss}}{\partial^2 \tau_a} = -\frac{1}{(\theta-1+\tau_a)^2} - (v_\varphi + a v_\omega)
$$

$$
- \left(\frac{1+\chi}{\theta-1} \right) \left(\frac{\psi}{1+\psi} \right) \frac{(\delta\beta^2)^a}{(1-\bar{\tau})^2}
$$

$$
- \left(\frac{1+\chi}{1+\sigma} \right) \left(\frac{N_a}{\bar{N}} \right) \cdot \left[\frac{1}{(1-\tau_a)^2} + (\sigma+1)^3 \frac{\sigma-2\tau_a}{(\sigma+\tau_a)^4} v_{\varepsilon,a} \right]
$$

$$
+ \left(\frac{1+\chi}{1+\sigma} \right) \left(\frac{1}{1-\tau_a} + \left(\frac{1+\sigma}{\sigma+\tau_a} \right)^3 \tau_a v_{\varepsilon,a} \right) \frac{1}{\bar{N}} \left[1 - \left(\frac{N_a}{\bar{N}} \right) \frac{1}{A} \right] \frac{\partial N_a}{\partial \tau_a}.
$$

Clearly, the first two terms are negative. The last term is always negative since $\bar{N}A \geq N_a$ and $\frac{\partial N_a}{\partial \tau_a} < 0$, recall eq. (23). Therefore, a sufficient condition for the third term to be negative is that $\sigma \geq 2$. This establishes that the social welfare function is globally concave in $\{\tau_a\}$ when $\sigma \geq 2$, so the first-order condition (A16) is necessary and sufficient to characterize the optimal τ_a.

(i) Simple differentiation establishes that this optimality condition is

$$
\begin{aligned}
0 \;=\; & \frac{1}{\theta - 1 + \tau_a} - \frac{1}{\theta} + (1 - \tau_a)\left(v_\varphi + a v_\omega\right) + \frac{1}{1 + \sigma} + \\
& - \left[\left(\frac{1 + \chi}{\theta - 1}\right)\frac{1}{1 - \bar{\tau}} - \frac{1}{\theta}\right]\frac{\psi}{1 + \psi}\beta^a \\
& - \left(\frac{1 + \chi}{1 + \sigma}\right)\left[\frac{1}{1 - \tau_a} + \left(\frac{\sigma + 1}{\sigma + \tau_a}\right)^3 \tau_a v_{\varepsilon,a}\right]\frac{N\left(a, \tau_a\right)}{\bar{N}\left(\{\tau_a\}\right)},
\end{aligned}
\tag{A16}
$$

where the expressions for $N\left(a, \tau_a\right)$ and $\bar{N}\left(\{\tau_a\}\right)$ are given in Corollary 2.2.

(ii) By inspecting (A16), it is immediate to see that age a does not enter as an argument in the first-order condition provided that $v_\omega = 0$, the sequences $\{v_{\varepsilon a}\}$ and $\{x_a - \bar{\varphi}_a\}$ are constant, and one of the following conditions is satisfied: either $\beta \to 1$ or $\theta \to \infty$. Therefore, the sequence of optimal τ_a must be independent of age in this case. As a consequence, $\tilde{Y}\left(a\right)$ is age-invariant, and hence, from the first-order condition $\left(A13\right)$, the optimal λ_a^* must also be independent of age.

(iii) Relative to the benchmark in (ii), when $v_\omega > 0$, the optimal τ_a^* is increasing with age since a larger value for $a v_\omega$ must be balanced by a lower value for $(1 - \tau_a)$.

(iv) Relative to the benchmark in (ii), when $v_{\varepsilon,a}$ increasing in age between age a and $a + 1$, it is easy to see that $\tau_a^* > \tau_{a+1}^*$.

(v) Relative to the benchmark in (ii), the optimal τ_a^* is increasing with age also when $\beta < 1$ and $\theta < \infty$. To see this, note that the term on the second line,

$$
- \left(\frac{1 + \chi}{\theta - 1}\frac{1 - \beta\delta}{1 - \delta}\frac{1}{1 - \bar{\tau}} - \frac{1}{\theta}\right)\frac{\psi}{1 + \psi}\left(\beta\right)^a,
$$

is negative and increasing in a when $\beta < 1$ and $\bar{\tau} \geq 0$. Thus, when a increases, the other terms must fall. Note that the terms $\frac{1}{\theta - 1 + \tau_a}$, $(1 - \tau_a)\left(v_\varphi + a v_\omega\right)$, and the term in the third line are all decreasing in τ_a. It follows that τ_a must increase with age.

(vi) Relative to the benchmark in (ii), when $\{x_a - \bar{\varphi}_a\}$ is increasing with age $N\left(a\right)/\bar{N}$ is increasing in age in the last term of (A16). Thus, a lower value of $(1 - \tau_a)^{-1}$ is needed to counterbalance this force, which implies that the optimal τ_a^* is decreasing in age.

A.7 Proof of Corollary 4.1 [optimal age-dependent taxation with life cycle only]

When individuals differ only by age, the equilibrium expressions for hours and earnings simplify to

$$h(a) = \exp(-\varphi_a)(1 - \tau_a)^{\frac{1}{1+\sigma}}, \tag{A17}$$

$$w(a)h(a) = N_a(\tau_a) = \exp(x_a - \varphi_a)(1 - \tau_a)^{\frac{1}{1+\sigma}}. \tag{A18}$$

Under the assumptions stated in the corollary, the first-order condition for optimal progressivity at age a is

$$1 - \tau_a^* = (1 + \chi) \frac{N_a(\tau_a^*)}{N(\bar{\tau}(\{\tau_a^*\}))}. \tag{A19}$$

Equations (A18) and (A19) combined imply

$$1 - \tau_a^* = \left[(1 + \chi)\frac{\exp(x_a - \bar{\varphi}_a)}{N(\bar{\tau}(\{\tau_a^*\}))}\right]^{\frac{1+\sigma}{\sigma}}. \tag{A20}$$

Recall that the planner wants to choose the sequence $\{\lambda_a\}$ to equate consumption across age groups. Thus, it will set λ_a^* subject to

$$c(a) = \lambda_a^* N_a(\tau_a^*)^{1-\tau_a^*} = C,$$

which implies

$$\lambda_a^* = \frac{C}{N_a(\tau_a^*)^{1-\tau_a^*}}.$$

The intratemporal first-order condition at age a is

$$\frac{\lambda_a(1 - \tau_a)(w(a)h(a))^{-\tau_a}w(a)}{C} = \exp\left(-(1 + \sigma)\bar{\varphi}_a\right)h(a)^{\sigma},$$

and since $w(a)h(a) = N_a(\tau_a)$, the labor wedge in this intratemporal first-order condition is

$$LW_a = \lambda_a(1 - \tau_a)N_a(\tau_a)^{-\tau_a} \tag{A21}$$

$$= \frac{C}{N_a(\tau_a)}(1 - \tau_a).$$

Now plug the expression for $N_a(\tau_a)$ (A18) and the solution for $(1 - \tau_a^*)$ (A20) into $(A21)$, which gives

$$LW_a = \frac{1 + \chi}{N(\bar{\tau}(\{\tau_a^*\}))} C,$$

which demonstrates that the labor wedge is independent of age. Moreover, from the resource constraint and the optimal public good provision condition, we know that

$$C(1 + \chi) = Y = N(\bar{\tau}(\{\tau_a^*\})),$$

which implies that $LW_a = 1$ (i.e., the effective marginal tax rate is zero).

Because the optimal tax and transfer scheme leaves labor supply undistorted and equates consumption across age groups, it implements the first-best allocation.

Finally, from eq. $(A21)$, imposing $LW_a = 1$ and averaging across age groups gives

$$Y = \frac{1}{A} \sum_{a=0}^{A-1} N_a(\tau_a(\{\tau_a^*\})) = \frac{1}{A} \sum_{a=0}^{A-1} C(1 - \tau_a^*).$$

Then, using $\frac{C}{Y} = \frac{1}{1+\chi}$, we get the expression for the optimal average degree of tax progressivity,

$$\frac{1}{A} \sum_{a=0}^{A-1} \tau_a^* = -\chi.$$

A.8 Proof of Proposition 5 [optimal age dependent taxation with transition]

First, note that the derivations and expressions for equilibrium allocations, conditional on a given fiscal policy, are identical to those in the proofs of Propositions 1 and 2 and Corollaries 2.1 and 2.2. The analysis only differs once we start constructing the expression for social welfare because allocations and policies now vary by time. Define

$$Y_{a,t} = \mathbb{E}\left[p_{a,t}(s; \bar{\tau}_{a,t}) \exp(x_a + \alpha) h_{a,t}(\varphi, \varepsilon; \tau_{a,t})\right]$$

$$C_{a,t} = \lambda_{a,t} \tilde{Y}_{a,t}$$

where

$$\tilde{Y}_{a,t} = \mathbb{E}\left[\left(p_{a,t}(s; \bar{\tau}_{a,t}) \exp(x_a + \alpha) h_{a,t}(\varphi, \varepsilon; \tau_{a,t})\right)^{1 - \tau_{a,t}}\right],$$

and let

$$Y_t = \frac{1}{A} \sum_{a=0}^{A-1} Y_{a,t}$$

$$\tilde{Y}_t = \frac{1}{A} \sum_{a=0}^{A-1} \tilde{Y}_{a,t}$$

denote the corresponding population averages.

The government budget constraint can be written as

$$(1 - g_t) \, Y_t = \frac{1}{A} \sum_{a=0}^{A-1} \lambda_{a,t} \tilde{Y}_{a,t}. \tag{A22}$$

Given the allocations described, we can assemble the components of social welfare in eq. (29).

Expected utility from consumption for age group a at date t (ignoring the term $\frac{1-\beta}{1-\beta^A} \beta^a$ in eq. (2), which pre-multiplies all the utility components involving consumption, hours, and public consumption) is

$$
\begin{aligned}
\mathbb{E}\left[\log c_{a,t}(\varphi, \alpha, s; \lambda_{a,t}, \tau_{a,t}, \bar{\tau}_{a,t})\right] &= \{\mathbb{E}\left[\log c_{a,t}(\varphi, \alpha, s)\right] - \log C_{a,t}\} + \log C_{a,t} \tag{A23} \\
&= -(1 - \tau_{a,t})^2 \left(\frac{v_\omega a}{2} + \frac{v_\varphi}{2}\right) + \log\left(1 - \left(\frac{1 - \tau_{a,t}}{\theta}\right)\right) + \left(\frac{1 - \tau_{a,t}}{\theta}\right) \\
&\quad + \log \lambda_{a,t} + \log \tilde{Y}_{a,t}.
\end{aligned}
$$

Expected utility from public good provision is

$$\chi \log\left(g_t Y_t\right). \tag{A24}$$

Expected disutility from hours worked for age group a at date t is

$$-\mathbb{E}\left[\frac{\exp\left[(1 + \sigma)\left(\bar{\varphi}_a + \varphi\right)\right]}{1 + \sigma} h_{a,t}\left(\varphi, \varepsilon; \tau_{a,t}\right)^{1+\sigma}\right] = -\frac{1 - \tau_{a,t}}{1 + \sigma}.$$

The expected utility contribution from skill investment for age group a at date t is

$$-\mathbb{E}\left[\frac{\kappa^{-(1/\psi)}}{1 + 1/\psi} s(\bar{\tau}_{a,t})^{1+1/\psi}\right] = -\frac{\psi}{1 + \psi}\left(\frac{1 - \bar{\tau}_{a,t}}{\theta}\right).$$

65

We can now compute total welfare for the planner, as of date 0, using eqs. (2) and (29).

Note first that the policy parameters $\lambda_{a,t}$ only appear in the terms involving expected utility from consumption and the government budget constraint. Let ζ_t denote the multiplier on this constraint (eq. A22).

The first-order condition with respect to $\lambda_{a,t}$ is

$$\frac{1}{\lambda_{a,t}} = \zeta_t \tilde{Y}_{a,t} = \zeta_t \frac{C_{a,t}}{\lambda_{a,t}},$$

which implies

$$C_{a,t} = \frac{1}{\zeta_t}.$$

Thus, given $\tilde{Y}_{a,t}$ (which is independent of $\lambda_{a,t}$), the policy parameter $\lambda_{a,t}$ is set so that average consumption for age group a is independent of age and only varies with time. This value for $C_{a,t}$ is uniquely pinned down from the government budget constraint, given a value for g_t:

$$C_{a,t} = C_t = (1 - g_t)Y_t, \tag{A25}$$

which implies

$$\lambda_{a,t} = \frac{(1 - g_t)Y_t}{\tilde{Y}_{a,t}}.$$

With the expression for $C_{a,t}$ in eq. (A25) substituted into the first row of eq. (A23), $\lambda_{a,t}$ no longer appears in any of the terms in social welfare.

Now consider the optimality condition for g_t. Note that g_t appears in the form $\chi \log g_t$ in the contribution from publicly provided goods (eq. A24) and in the form $\log(1 - g_t)$ in the contribution from the level of average private consumption. The first-order condition with respect to g_t immediately implies the result $g_t = \frac{\chi}{1+\chi}$.

A.9 Proof of Proposition 6 [optimal taxation with transition and inelastic labor supply]

We now write out all the terms in social welfare explicitly. To start with, we allow for flexible labor supply ($\sigma < \infty$). In order to economize on space, we assume people live for only two periods: the generalization to $A > 1$ is straightforward.

The date t component of social welfare in the expression eq. (29) (ignoring the terms $\chi \log g_t$ and $\log(1 - g_t)$, which do not involve any $\tau_{a,t}$ parameters) is

$$\left(\frac{1-\beta}{1-\beta^2}\right)\left(\log\left(1-\left(\frac{1-\tau_{1,t}}{\theta}\right)\right)+\left(\frac{1-\tau_{1,t}}{\theta}\right)-\frac{1-\tau_{1,t}}{1+\sigma}+(1+\chi)\log\left\{\frac{1}{2}\left(Y_{0,t}+Y_{1,t}\right)\right\}\right)$$

$$+\left(\frac{1-\beta}{1-\beta^2}\right)\left(\log\left(1-\left(\frac{1-\tau_{0,t}}{\theta}\right)\right)+\left(\frac{1-\tau_{0,t}}{\theta}\right)-\frac{1-\tau_{0,t}}{1+\sigma}+(1+\chi)\log\left\{\frac{1}{2}\left(Y_{0,t}+Y_{1,t}\right)\right\}\right)$$

$$-\frac{\psi}{1+\psi}\left(\frac{1-\left(\frac{1-\beta}{1-\beta^2}\left(\tau_{0,t}+\beta E_t\left[\tau_{1,t+1}\right]\right)\right)}{\theta}\right),$$

where the first line reflects the contribution to welfare from the old, and the second and third lines the contribution from the young.

Output of the young and old at t is given by

$$Y_{0,t} = (1-\tau_{0,t})^{\frac{1}{1+\sigma}}\cdot \mathbb{E}_t[p_{0,t}]$$
$$Y_{1,t} = (1-\tau_{1,t})^{\frac{1}{1+\sigma}}\cdot \mathbb{E}_{t-1}[p_{1,t}],$$

where

$$\mathbb{E}_t[p_{0,t}]=\mathbb{E}_t[p_{1,t+1}]=\left[\left(\frac{\theta}{\theta-1}\right)^{\frac{\theta}{\theta-1}}\left(\frac{1-\frac{(1-\beta)}{(1-\beta^2)}\left(\tau_{0,t}+\beta\mathbb{E}_t[\tau_{1,t+1}]\right)}{\theta}\right)^{\frac{\psi}{(1+\psi)(\theta-1)}}\left(\frac{1}{\eta}\right)^{\frac{1}{(1+\psi)(\theta-1)}}\right].$$

Note that $Y_t=\frac{1}{2}\left(Y_{0,t}+Y_{1,t}\right)$ depends on $\tau_{0,t}$, $\mathbb{E}_t[\tau_{1,t+1}]$, $\tau_{1,t}$ and $\mathbb{E}_{t-1}[\tau_{1,t}]$.

At $t=0$ (the time of the reform), only three tax parameters $(\tau_{0,0},\tau_{1,1},\tau_{1,0})$ affect contemporaneous output:

$$\frac{\partial\log Y_0}{\partial\tau_{0,0}}=\frac{1}{2}\frac{-\frac{1}{1+\sigma}(1-\tau_{0,0})^{\frac{-\sigma}{1+\sigma}}\mathbb{E}_0[p_{0,0}]+(1-\tau_{0,0})^{\frac{1}{1+\sigma}}\frac{\partial\mathbb{E}_0[p_{0,0}]}{\partial\tau_{0,0}}}{Y_0}\;:\;\text{young adjust hours \& skill inv. at } t=0$$

$$\frac{\partial\log Y_0}{\partial\tau_{1,1}}=\frac{1}{2}\frac{(1-\tau_{1,1})^{\frac{1}{1+\sigma}}\frac{\partial\mathbb{E}_0[p_{1,1}]}{\partial\tau_{1,1}}}{Y_0}\;:\;\text{young adjust skill investment at } t=0 \text{ in response to } \tau_{1,1}$$

$$\frac{\partial\log Y_0}{\partial\tau_{1,0}}=\frac{1}{2}\frac{-\frac{1}{1+\sigma}(1-\tau_{1,0})^{\frac{-\sigma}{1+\sigma}}\mathbb{E}_{-1}[p_{1,0}(\tau_{-1})]}{Y_0}\;:\;\text{old adjust hours at } t=0$$

In contrast, for a generic date $t>0$, output depends on four different parameters:

$$\frac{\partial \log Y_t}{\partial \tau_{0,t}} = \frac{1}{2} \frac{-\frac{1}{1+\sigma}(1-\tau_{0,t})^{\frac{-\sigma}{1+\sigma}} \mathbb{E}[p_{0,t}] + (1-\tau_{0,t})^{\frac{1}{1+\sigma}} \frac{\partial \mathbb{E}[p_{0,t}]}{\partial \tau_{0,t}}}{Y_t} \quad : \quad \text{young adjust hours \& skill inv. at } t$$

$$\frac{\partial \log Y_t}{\partial \tau_{1,t+1}} = \frac{1}{2} \frac{(1-\tau_{0,t})^{\frac{1}{1+\sigma}} \frac{\partial \mathbb{E}[p_{0,t}]}{\partial \tau_{1,t+1}}}{Y_t} \quad : \quad \text{young adjust skill investment at } t$$

$$\frac{\partial \log Y_t}{\partial \tau_{1,t}} = \frac{1}{2} \frac{-\frac{1}{1+\sigma}(1-\tau_{1,t})^{\frac{-\sigma}{1+\sigma}} \mathbb{E}[p_{1,t}] + (1-\tau_{1,t})^{\frac{1}{1+\sigma}} \frac{\partial \mathbb{E}[p_{1,t}]}{\partial \tau_{1,t}}}{Y_t} \quad : \quad \text{old adjust hours at } t \text{ \& skill inv. at } t-1$$

$$\frac{\partial \log Y_t}{\partial \tau_{0,t-1}} = \frac{1}{2} \frac{(1-\tau_{1,t})^{\frac{1}{1+\sigma}} \frac{\partial \mathbb{E}[p_{1,t}]}{\partial \tau_{0,t-1}}}{Y_t} \quad : \quad \text{old adjust hours at } t \text{ \& skill inv. at } t-1$$

Thus, in general $\tau_{0,t}$ affects both Y_t and Y_{t+1}, while $\tau_{1,t}$ affects Y_t and Y_{t-1}.

Consider the generic first-order condition for $\tau_{0,t}$ for all $t \geq 0$. We have

$$\left(\frac{1-\beta}{1-\beta^2}\right)(1+\chi)\frac{\partial \log Y_t}{\partial \tau_{0,t}} \quad : \quad \text{effect on old at } t$$

$$+ \left(\frac{1-\beta}{1-\beta^2}\right)\left(\frac{\frac{1}{\theta}}{1-\left(\frac{1-\tau_{0,t}}{\theta}\right)} - \frac{1}{\theta} + \frac{1}{1+\sigma} + (1+\chi)\frac{\partial \log Y_t}{\partial \tau_{0,t}}\right) + \frac{\psi}{1+\psi}\frac{\left(\frac{1-\beta}{1-\beta^2}\right)}{\theta} \quad : \quad \text{effect on young at } t$$

$$+\beta\left(\frac{1-\beta}{1-\beta^2}\right)\left((1+\chi)\frac{\partial \log Y_{t+1}}{\partial \tau_{0,t}}\right) \quad : \quad \text{effect on old at } t+1$$

$$+\beta\left(\frac{1-\beta}{1-\beta^2}\right)\left((1+\chi)\frac{\partial \log Y_{t+1}}{\partial \tau_{0,t}}\right) \quad : \quad \text{effect on young at } t+1$$

$$= 0$$

The terms here are readily interpretable. Increasing $\tau_{0,t}$ reduces skill investment at t, which reduces output at $t+1$. This reduces welfare for young and old at $t+1$ (the last two lines). Increasing $\tau_{0,t}$ also reduces output at t, both through the skill investment channel, and by reducing the labor supply of the young. This accounts for the first line and the term involving output in the second line. Finally, increasing $\tau_{0,t}$ compresses consumption inequality among the young at t, reduces hours worked by the young at t, and reduces skill investment costs by the young at t. These are the remaining terms in the second line.

Now consider the generic first-order condition for $\tau_{1,t+1}$ for $t \geq 0$. This condition is

$$\left(\frac{1-\beta}{1-\beta^2}\right)(1+\chi)\frac{\partial \log Y_t}{\partial \tau_{1,t+1}} \quad : \quad \text{effect on old at } t$$

$$+\left(\frac{1-\beta}{1-\beta^2}\right)(1+\chi)\frac{\partial \log Y_t}{\partial \tau_{1,t+1}} + \beta\frac{\psi}{1+\psi}\left(\frac{\left(\frac{1-\beta}{1-\beta^2}\right)}{\theta}\right) \quad : \quad \text{effect on young at } t$$

$$+\beta\left(\frac{1-\beta}{1-\beta^2}\right)\left(\frac{\frac{1}{\theta}}{1-\left(\frac{1-\tau_{1,t+1}}{\theta}\right)} - \frac{1}{\theta} + \frac{1}{1+\sigma} + (1+\chi)\frac{\partial \log Y_{t+1}}{\partial \tau_{1,t+1}}\right) \quad : \quad \text{effect on old at } t+1$$

$$+\beta\left(\frac{1-\beta}{1-\beta^2}\right)\left((1+\chi)\frac{\partial \log Y_{t+1}}{\partial \tau_{1,t+1}}\right) \quad : \quad \text{effect on young at } t+1$$

We can write these two first-order conditions more compactly as

$$2(1+\chi)\left(\frac{\partial \log Y_t}{\partial \tau_{0,t}} + \beta\frac{\partial \log Y_{t+1}}{\partial \tau_{0,t}}\right) + \frac{\psi}{1+\psi}\frac{1}{\theta} + \left(\frac{1}{(\theta-1+\tau_{0,t})} - \frac{1}{\theta} + \frac{1}{1+\sigma}\right) = 0 \quad \text{(A26)}$$

$$2(1+\chi)\left(\frac{\partial \log Y_t}{\partial \tau_{1,t+1}} + \beta\frac{\partial \log Y_{t+1}}{\partial \tau_{1,t+1}}\right) + \beta\frac{\psi}{1+\psi}\frac{1}{\theta} + \beta\left(\frac{1}{(\theta-1+\tau_{1,t+1})} - \frac{1}{\theta} + \frac{1}{1+\sigma}\right) = 0 \quad \text{(A27)}$$

Now note that

$$\frac{\partial \log Y_t}{\partial \tau_{0,t}} + \beta\frac{\partial \log Y_{t+1}}{\partial \tau_{0,t}} = \frac{1}{2}\frac{-\frac{1}{1+\sigma}(1-\tau_{0,t})^{\frac{-\sigma}{1+\sigma}}\mathbb{E}[p_{0,t}] + (1-\tau_{0,t})^{\frac{1}{1+\sigma}}\frac{\partial \mathbb{E}[p_{0,t}]}{\partial \tau_{0,t}}}{Y_t} + \beta\frac{1}{2}\frac{(1-\tau_{1,t+1})^{\frac{1}{1+\sigma}}\frac{\partial \mathbb{E}[p_{1,t+1}]}{\partial \tau_{0,t}}}{Y_{t+1}}$$

$$\frac{\partial \log Y_t}{\partial \tau_{1,t+1}} + \beta\frac{\partial \log Y_{t+1}}{\partial \tau_{1,t+1}} = \frac{1}{2}\frac{(1-\tau_{0,t})^{\frac{1}{1+\sigma}}\frac{\partial \mathbb{E}[p_{0,t}]}{\partial \tau_{1,t+1}}}{Y_t} + \beta\frac{1}{2}\frac{-\frac{1}{1+\sigma}(1-\tau_{1,t+1})^{\frac{-\sigma}{1+\sigma}}\mathbb{E}[p_{1,t+1}] + (1-\tau_{1,t+1})^{\frac{1}{1+\sigma}}\frac{\partial \mathbb{E}[p_{1,t+1}]}{\partial \tau_{1,t+1}}}{Y_{t+1}}$$

$$\frac{\partial \mathbb{E}[p_{1,t+1}]}{\partial \tau_{1,t+1}} = \frac{\partial \mathbb{E}[p_{0,t}]}{\partial \tau_{1,t+1}} = \beta\frac{\partial \mathbb{E}[p_{1,t+1}]}{\partial \tau_{0,t}} = \beta\frac{\partial \mathbb{E}[p_{0,t}]}{\partial \tau_{0,t}}$$

which jointly imply

$$\frac{\partial \log Y_t}{\partial \tau_{1,t+1}} + \beta\frac{\partial \log Y_{t+1}}{\partial \tau_{1,t+1}} = \beta\left(\frac{\partial \log Y_t}{\partial \tau_{0,t}} + \beta\frac{\partial \log Y_{t+1}}{\partial \tau_{0,t}}\right) + \beta\frac{1}{1+\sigma}\mathbb{E}[p_{0,t}]\left(\frac{(1-\tau_{0,t})^{\frac{-\sigma}{1+\sigma}}}{Y_t} - \frac{(1-\tau_{1,t+1})^{\frac{-\sigma}{1+\sigma}}}{Y_{t+1}}\right)$$

$$\text{(A28)}$$

Now consider the case with inelastic labor supply, so that $\sigma \to \infty$. The second term on the right-hand side of the above equation drops out. Then substituting this equation into the first-order condition for

$\tau_{1,t+1}$ (eq. A27), it is clear that the first-order conditions for $\tau_{0,t}$ and $\tau_{1,t+1}$ are exactly symmetric, except that all the terms in the latter are multiplied by β.

It follows that the optimal value for $\tau_{0,t}$ is equal to the optimal value for $\tau_{1,t+1}$. Note, finally, that the optimal value for these policy parameters must be strictly less than one. The reason is that from the first-order condition eq. (A26), the marginal value of consumption compression at $\tau_{0,t} = 1$ is zero, while the marginal cost in terms of reduced skill investment and output is strictly positive.

Now consider the optimal choice for $\tau_{1,0}$, progressivity for the old at the time of the tax reform, which is the only choice we have not explored so far. The first-order condition here is

$$\left(\frac{1}{(\theta - 1 + \tau_{1,0})} - \frac{1}{\theta} + \frac{1}{1 + \sigma} + (1 + \chi)\frac{\partial \log Y_0}{\partial \tau_{1,0}} \right) + (1 + \chi)\frac{\partial \log Y_0}{\partial \tau_{1,0}} = 0,$$

where

$$\frac{\partial \log Y_0}{\partial \tau_{1,0}} = \frac{1}{2}\frac{1 - \frac{1}{1+\sigma}(1 - \tau_{1,0})^{\frac{-\sigma}{1+\sigma}}\mathbb{E}_{-1}[p_{1,0}]}{Y_t}$$

Here, an increase in $\tau_{1,0}$ reduces consumption inequality among the old and reduces the old's labor supply, which translates into reduced output and thus consumption and government spending for both the young and the old at date 0.

With inelastic labor supply, the first-order condition simplifies further to

$$\frac{1}{(\theta - 1 + \tau_{1,0})} - \frac{1}{\theta} = 0,$$

which immediately implies $\tau_{1,0} = 1$.

A.10   Extension to age variation in the taste for leisure

By following the same steps of the proof of Proposition 1, we arrive at the new allocations:

$$\log h\left(\varphi, a, \varepsilon\right) = \frac{\log(1 - \tau_a)}{1 + \sigma} - (\varphi + \bar{\varphi}_a - \gamma_a) + \left(\frac{1 - \tau_a}{\sigma + \tau_a}\right)\varepsilon - \frac{1}{\sigma + \tau_a}\mathcal{C}_a, \tag{A29}$$

$$\log c\left(\varphi, s, a, \alpha\right) = \log \lambda_a + (1 - \tau_a)\left[\log p(s, \bar{\tau}) + x_a + \alpha + \frac{\log(1 - \tau_a)}{1 + \sigma} - (\varphi + \bar{\varphi}_a - \gamma_a)\right] + \mathcal{C}_a. \tag{A30}$$

The equilibrium skill prices $p(s, \bar{\tau})$ remain unchanged. By following the same derivations needed to obtain the steady-state welfare expression (27), we arrive at:

$$
\begin{aligned}
\mathcal{W}(\{\tau_a\}) \;=\; & -\frac{1}{A}\sum_{a=0}^{A-1}\underbrace{\frac{1-\tau_a}{1+\sigma}\exp\left((1+\sigma)\,\gamma_a\right)}_{\text{Disutility of labor}} && \text{(A31)}\\[2ex]
& + (1+\chi)\log\underbrace{\left\{\sum_{a=0}^{A-1}(1-\tau_a)^{\frac{1}{1+\sigma}}\cdot\exp\left[x_a-\bar\varphi_a+\gamma_a+\left(\frac{\tau_a\,(1+\hat\sigma_a)}{\hat\sigma_a^2}+\frac{1}{\hat\sigma_a}\right)\frac{v_{\varepsilon,a}}{2}\right]\right\}}_{\text{Effective hours worked }\bar N_a}\\[2ex]
& + (1+\chi)\underbrace{\frac{1}{(1+\psi)(\theta-1)}\left[\psi\log\left(1-\bar\tau\right)+\log\left(\frac{1}{\eta\theta^{\psi}}\left(\frac{\theta}{\theta-1}\right)^{\theta(1+\psi)}\right)\right]}_{\text{Productivity of skill investment: }\log(\text{average skill price})\,=\,\log(E[p(s)])}\\[2ex]
& -\underbrace{\frac{\psi}{1+\psi}\frac{1-\bar\tau}{\theta}}_{\text{Avg. education cost}}+\frac{1}{A}\sum_{a=0}^{A-1}\exp\left((1+\sigma)\,\gamma_a\right)\underbrace{\left[\log\left(1-\left(\frac{1-\tau_a}{\theta}\right)\right)+\left(\frac{1-\tau_a}{\theta}\right)\right]}_{\text{Consumption dispersion across skills}}\\[2ex]
& -\frac{1}{A}\sum_{a=0}^{A-1}\exp\left((1+\sigma)\,\gamma_a\right)\cdot\frac{1}{2}\underbrace{\left(1-\tau_a\right)^2\left(v_\varphi+a v_\omega\right)}_{}\\
& \text{Cons. dispersion due to uninsurable risk and preference heterogeneity}
\end{aligned}
$$

B   Extension to borrowing and saving

This appendix describes the extension of the benchmark model to an economy where households are allowed to trade a risk-free bond in zero net supply.

B.1   Economic Environment

The economic environment is virtually identical to the one in the main text. We therefore only highlight what differs.

Financial assets: The main difference from the benchmark model consists of the ability of households to save in a risk-free asset b with gross return $R = 1 + r$. Short positions are allowed up to an exogenous limit $-\bar{b}_a \left[\frac{p(s,\bar{\tau}) \exp(\alpha_a)}{\exp(\varphi)} \right]$, where $\bar{b}_A = 0$, that is, a no-Ponzi condition stating that the household cannot die with negative wealth needs to hold. Assets are in zero net supply, (i.e. $B = 0$). Thus, when $\bar{b}_a = 0$ for all a, the only equilibrium is autarky, and the benchmark model becomes a special case of this general model.

Government: We assume that the tax base is total income net of saving, (i.e., expenditures), or

$$m_{i,a} = p(s_i, \bar{\tau}) \exp(x_{i,a} + \alpha_{i,a}) h_{i,a} + rb_{i,a} - (b_{i,a+1} - b_{i,a}). \tag{B1}$$

As we will see, this assumption is convenient because it allows us to simplify the model and retain the closed-form solution for the equilibrium skill price function $p(s, \bar{\tau})$. The tax and transfer scheme is defined as in the main text, that is,

$$T_a(m_{i,a}) = m_{i,a} - \lambda_a m_{i,a}^{1-\tau_a}. \tag{B2}$$

We abstract from the possibility that the government can issue debt or save. The government budget constraint therefore reads as

$$g \frac{1}{A} \sum_{a=0}^{A-1} \int y_{i,a} di_a = \frac{1}{A} \sum_{a=0}^{A-1} \int \left[m_{i,a} - \lambda_a (m_{i,a})^{1-\tau_a} \right] di_a. \tag{B3}$$

B.2 Solution to the household problem

The agent chooses skills at age $a = 0$. Abstract from this choice for now and consider an individual with skill level s. After the skill choice, the household solves

$$\max_{\{b_{a+1}\}_{a=0}^{A-1}} \mathbb{E}_0 \left(\frac{1-\beta}{1-\beta^A}\right) \sum_{a=0}^{A-1} \beta^a \left[\log c_a - \frac{\exp\left[(1+\sigma)\left(\bar{\varphi}_a + \varphi\right)\right]}{1+\sigma} h_a^{1+\sigma} + \chi \log G\right]$$

$$s.t.$$

$$c_a = \lambda_a \left[p(s,\bar{\tau}) \exp(x_a + \alpha_a)h_a + Rb_a - b_{a+1}\right]^{1-\tau_a}$$

and

$$b_{a+1} \geq -\bar{b}_a \left[\frac{p\left(s,\bar{\tau}\right)\exp\left(\alpha_a\right)}{\exp\left(\varphi\right)}\right]. \tag{B4}$$

The first-order condition for hours worked and the Euler equation are

$$\frac{(1-\tau_a)p(s,\bar{\tau})\exp(x_a+\alpha_a)}{p(s,\bar{\tau})\exp(x_a+\alpha_a)h_a + Rb_a - b_{a+1}} = \exp\left[(1+\sigma)\left(\bar{\varphi}_a + \varphi\right)\right]h_a^{\sigma}$$

$$\frac{(1-\tau_a)}{p(s,\bar{\tau})\exp(x_a+\alpha_a)h_a + Rb_a - b_{a+1}} \geq \beta R\mathbb{E}_a \left[\frac{(1-\tau_{a+1})}{p(s,\bar{\tau})\exp(x_{a+1}+\alpha_{a+1})h_{a+1} + Rb_{a+1} - b_{a+2}}\right]$$

$$= \text{ if } b_{a+1} > -\bar{b}_a \left[\frac{p\left(s,\bar{\tau}\right)\exp\left(\alpha_a\right)}{\exp\left(\varphi\right)}\right].$$

Define transformations of endogenous variables as follows:

$$\hat{h}_a = h_a \cdot \exp\left(\varphi\right) \tag{B5}$$

$$\hat{b}_a = b_a \cdot \frac{\exp\left(\varphi\right)}{\exp\left(\alpha_a\right)p\left(s,\bar{\tau}\right)} \tag{B6}$$

$$\hat{b}_{a+1}^* = b_{a+1} \cdot \frac{\exp\left(\varphi\right)}{\exp\left(\alpha_a\right)p\left(s,\bar{\tau}\right)} \tag{B7}$$

$$\hat{b}_{a+1} = \frac{\hat{b}_{a+1}^*}{\exp\left(\omega_{a+1}\right)} \tag{B8}$$

With some algebra, it can easily be shown that the two first-order conditions can be rewritten in terms

of these transformed variables as

$$\frac{(1 - \tau_a) \exp(x_a)}{\exp(x_a)\hat{h}_a + R\hat{b}_a - \hat{b}^*_{a+1}} = \exp\left[(1 + \sigma)\bar{\varphi}_a\right]\hat{h}^\sigma_a$$

$$\frac{1}{\exp(x_a)\hat{h}_a + R\hat{b}_a - \hat{b}^*_{a+1}} \geq \beta R \left(\frac{1 - \tau_{a+1}}{1 - \tau_a}\right) \mathbb{E}_a \left[\frac{1}{\exp(\omega_{a+1})\left(\exp(x_{a+1})\hat{h}_{a+1} + R\hat{b}_{a+1} - \hat{b}^*_{a+2}\right)}\right]$$

$$= \quad \text{if } \hat{b}^*_{a+1} \geq -\bar{b}_a.$$

The advantage of writing the first-order conditions in this way is that age a and transformed wealth $\hat{b}_a$ are the only idiosyncratic states. We can use a backstepping algorithm starting from the known fact that $\hat{b}_A = 0$.

In particular, from the household problem, we obtain policy functions $\hat{h}_a\left(\hat{b}_a; \tau_a\right), \hat{b}^*_{a+1}\left(\hat{b}_a; \{\tau_a\}\right), \hat{c}_a\left(\hat{b}_a; \{\tau_a\}\right)$, where $\hat{c}_a := \exp(x_a)\hat{h}_a + R\hat{b}_a - \hat{b}^*_{a+1}$. From these decisions, by rescaling back to the original states, we can obtain:

$$h_a\left(b_a; \varphi, \tau_a\right) = \hat{h}_a\left(\hat{b}_a \cdot \exp\left(\alpha_a\right) \cdot p\left(s, \bar{\tau}\right)\exp\left(-\varphi\right); \tau_a\right)\exp\left(-\varphi\right)$$

$$b_{a+1}\left(b_a; \alpha_a, s, \varphi, \{\tau_a\}\right) = \exp\left(\alpha_a\right) \cdot p\left(s, \bar{\tau}\right)\exp\left(-\varphi\right) \cdot \hat{b}^*_{a+1}\left(\hat{b}_a \cdot \exp\left(\alpha_a\right) \cdot p\left(s, \bar{\tau}\right)\exp\left(-\varphi\right); \{\tau_a\}\right)$$

$$c_a\left(b_a; \alpha_a, s, \varphi, \{\tau_a\}\right) = \lambda_a \left[p\left(s, \bar{\tau}\right)\exp\left(\alpha_a\right)\exp\left(-\varphi\right)\right]^{1-\tau_a} \hat{c}_a\left(\hat{b}_a; \{\tau_a\}\right)^{1-\tau_a},$$

where the last one is obtained residually from the budget constraint.

B.3   The wealth distribution

Suppose we have come up with an interest rate such that

$$\sum_{a=0}^{A-1} \int_{\hat{b}_a} \hat{b}^*_{a+1}\left(\hat{b}_a\right) dF_{\hat{b}_a} = 0,$$

if so, it can be shown that the true bond market clears, that is,

$$\sum_{a=0}^{A-1} \int_{b_a} \int_s \int_{\alpha_a} \int_\varphi b_{a+1}(b_a; \alpha_a, s, \varphi, \varnothing)dF_s dF_{\alpha_0} dF_\varphi dF_{b_a} = 0.$$

If we denote the distribution for an individual of age a $\mu_a(\hat{b}; \{\tau_a\})$, we know that $\mu_0(\hat{b} = 0) = 1$; that is, all the mass is at $b = 0$ (individuals are born with zero wealth). We can therefore easily initiate the recursion and move forward using the household's policy functions.

B.4 Welfare

The planner maximizes welfare:

$$\mathcal{W}(g, \lambda_a, \tau_a) = \frac{1}{A} \sum_{a=0}^{A-1} \bar{u}_a - \bar{v} + \chi \log \left(g \cdot \sum_{a=0}^{A-1} Y_a \right)$$

where $\bar{v} = \frac{\psi}{1+\psi} \left(\frac{1-\bar{\tau}}{\theta} \right)$ is the average education cost for the newborn cohort and the period utility term is

$$\bar{u}_a = \int \left[\log (c_{i,a}) - \frac{\exp \left[(1+\sigma) \left(\bar{\varphi}_a + \varphi \right) \right]}{1+\sigma} h_{i,a}^{1+\sigma} \right] di_a,$$

B.5 Computation

To find the optimal tax function, we start by setting $g = \chi/(1+\chi)$, the optimal solution, and approximate the τ_a function with a Chebyshev polynomial of order two. We then maximize the welfare function with respect to the three parameters of the Chebyshev polynomial. The formal algorithm is as follows:

1. Guess coefficients of Chebyshev polynomial $\{p_j\}_{j=0}^{2}$.

2. Evaluate the Chebyshev polynomial to get the full vector $\{\tau_a\}_{a=0}^{A-1}$:

   (a) Guess an interest rate, R .

   (b) Given R and $\{\tau_a\}_{a=0}^{A-1}$, solve the household problem using the endogenous grid method .

   (c) Compute the asset distribution $\{\mu_a(\hat{b})\}_{a=0}^{A-1}$ and total asset demand .

   (d) If asset demand is zero go to 3, otherwise update R and go back to 2b .

3. Given the solution to the household problem, compute welfare .

4. If welfare is maximized, stop; otherwise, update $\{p_j\}_{j=0}^{2}$ and go back to 2 .

To solve the household problem, we use a grid with 50 points on $[-\bar{b}_a, b_{max}]$ with $b_{max} = 10$ (further increasing the number of grid points or b_{max} has no effect on results). To compute the asset distribution, we use the histogram method on the same grid used for the household problem with 3,000 grid points. Finally, we approximate the ω distribution using Gaussian quadrature with nine Gauss-Hermite nodes.