Optimal Progressivity with Age-Dependent Taxation*

Jonathan Heathcote Kjetil Storesletten Giovanni L. Violante

First draft: August 2017 – This draft: August, 2018

Abstract

This paper studies optimal taxation of labor earnings when the degree of tax progressivity is allowed to vary with age. We analyze this question in an equilibrium overlapping-generations model that incorporates irreversible skill investment, flexible labor supply, ex-ante heterogeneity in disutility of work and cost of skill acquisition, ex-post partially uninsurable 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 result is that progressivity should be U-shaped in age. This quantitative finding is confirmed in a version of the model with borrowing and saving solved numerically. Welfare gains from making the tax system age dependent exceed two percent of lifetime consumption.

JEL Codes: D30, E20, H20, H40, J22, J24.

Keywords: Tax Progressivity, Life Cycle, Income Distribution, Skill Investment, Labor Supply, Incomplete Markets, Government Expenditures, Welfare.

*Heathcote: Federal Reserve Bank of Minneapolis and CEPR, e-mail heathcote@minneapolisfed.org; Storesletten: University of Oslo and CEPR, e-mail kjetil.storesletten@econ.uio.no; Violante: Princeton University, CEBI, CEPR, IFS, IZA, and NBER, e-mail violante@princeton.edu. The first draft was prepared for the conference "New Perspectives on Consumption Measures." We are grateful to Ricardo Cioffi for outstanding research assistance. We thank Mark Huggett, Axelle Ferriere, Guy Laroque, Magne Mogstad, Nicola Pavoni, various seminar participants, and two anonymous referees for comments. Kjetil Storesletten acknowledges support from the European Research Council (ERC Advanced Grant IPCDP-324085), as well as from Oslo Fiscal Studies. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

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, 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). 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, Troshkin, and Tsyvinski 2016; Stantcheva forthcoming; and Weinzierl, 2011).

The purpose of this paper is to study optimal taxation in a setting in which the tax system can vary with age. We do not study fully optimal tax system design, in the Mirrleesian tradition, but instead restrict attention to the parametric class of income tax and transfer systems given by

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

where y is pre-tax 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), Kakwani (1977) and, more recently, Bénabou (2000, 2002) and In Heathcote, Storesletten, and Violante (2017).

The key innovation in the present paper is to let the parameters λ and τ in (1) be conditioned on age, subject to an economy-wide government budget constraint. By allowing for age variation in λ and τ, 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 (1) provides a remarkably good approximation of the actual tax and transfer scheme in the U.S. for households aged 25-60. In particular, eq. (9) implies that after-tax earnings should be a log-linear function of pre-tax earnings. 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 average earnings yields

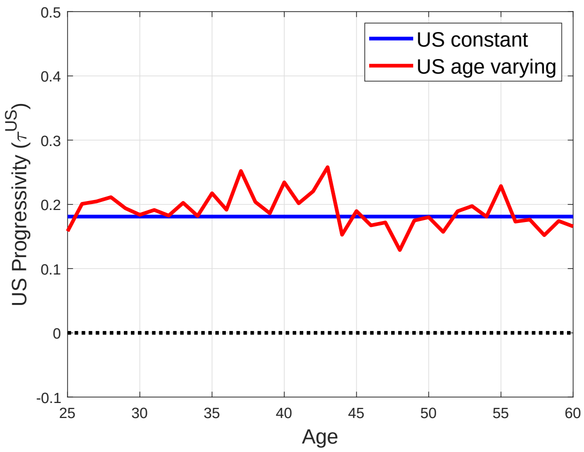

Figure 1: The slope coefficient τ of a regression of log disposable income $y - T(y)$ on log gross income y, where intercept and slope are both allowed to vary with age. The straight line is the estimated $\tau^{US} = 0.181$ when age dependence is not allowed in the regression. See Heathcote et al. (2017) for details on the 2000-2006 PSID data used this estimation, and the construction of y and $T(y)$ at the household level.

a very good fit, with an R^2 of 0.93: when plotting average pre-government against post-government earnings for each percentile of the sample, the relationship is virtually log-linear.

There, we did not investigate whether, implicitly, the current tax/transfer system features element 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 expenditure deductions) would effectively induce some age dependence. We thus repeated this estimation allowing the intercept and the slope parameters of the regression to depend on age. Figure 1 plots the estimated τ age by age together with the estimated age-invariant $\tau = 0.181$. The main finding is that there seems to be no significant age-dependence in progressivity embedded in the current U.S. system.

We aim to understand whether there is scope for improving the current U.S. 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 public good. They make an irreversible skill investments when young, and make a labor-leisure choice in each period of working life. Individuals 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 age variation in wages, hours, and consumption. During working life, individuals also face permanent shocks to their productivity that can only be partially insured privately. The uninsurable component of these wage shocks pass 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 subsitute for missing private insurance against life-cycle shocks. In addition, net tax revenue allows the government to provide the public good. However, tax progressivity discourages labor supply and skill investment. Because the tax system affects the equilibrium skill distribution, tax parameters influence pre-tax skill prices as well as after-tax returns.

Most of our analysis focuses on a version of the model in which there are no markets for inter-temporal borrowing and lending. In this environment, we are able to derive a closed-form solution for an equally-weighted 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.

The shape of the optimal age profile for the tax progressivity parameter τ trades off three 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, this will translate to increasing consumption dispersion with age. The planner has an incentive to target redistribution to where inequality is concentrated, namely among the old. This is a force for having *progressivity increase with age.*

Second, age is informative about the *average* cost of producing output, since wages net of the disutility of work are increasing during the first decades of working life. The planner has an incentive to smooth marginal tax rates by age, and given a rising life-cycle profile of wages coupled with a generally progressive tax system, this tax smoothing motive is a force for *progressivity to fall with age.*

Third, the Ramsey planner has an incentive to try to expropriate irreversible past skill investments. Because the young discount future returns and thus future taxes when choosing their skill level, the planner can expropriate from the old without excessively dis-incentivizing skill investment by future generations by having tax *progressivity increase with age*.

Given the age profile for τ that optimally balances these forces, 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 τ and λ arises because our utility specification, consistent with balanced growth, implies that λ has no impact on either skill investment or labor supply.

We parameterize the model to the U.S. economy in order to calculate the optimal age-dependent tax system. On their own, life-cycle variation in uninsurable risk, productivity and discounting each call for significant variation in tax progressivity over the life cycle, with correspondingly sizable welfare gains. However, when all factors are combined and transitional dynamics are taken into account, the three effects largely neutralize each other, so that the optimal profile for progressivity τ is mildly U-shaped in age. We compute the welfare gains of moving from the current age-invariant tax system to (i) the optimal age-invariant system, and (ii) the optimal age-varying one. We find large welfare gains from being able to smooth consumption over the life-cycle. Thus, if private borrowing and lending is ruled out, there is a strong case for introducing age variation in the tax level parameter λ. Given an optimally age-varying profile for λ, however, the additional welfare gains from introducing age-variation in τ are small.

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 setting 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. We have abstracted from this channel by choosing a specification in which the Frisch elasticity is constant. Conesa, Kitao, and Krueger (2009) study optimal taxation within a Gouveia-Strauss class of non-linear tax functions. While richer 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 a common value for the progressivity parameter τ.

A more recent literature studies the role of age variation in the Mirrlees optimal taxation framework. Three papers are especially related to our work. The first paper is by Weinzierl (2011), who focuses on the rising age profile of wages, and on how these profiles differ across skill groups. His key findings, namely that optimal average and marginal tax rates are both rising with age, are 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 and abstract from human capital investments. 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, Troshkin, and Tsyvinski (2016), who focus on the additional effect of skewness of wage shocks. Ndiaye (2017) extends Farhi and Werning to model a discrete retirement choice, which reduces optimal marginal tax rates around the age of retirement when labor supply is relatively elastic.

With respect to this existing set of results, our contribution is threefold. First, our closed-form expression for social welfare as a function of τ and the structural parameters of the model describing preferences, technology, ex ante heterogeneity, and income uncertainty leads to a transparent characterization. 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 wages are rising fast, while the uninsurable risk channel 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. This new channel induces age dependence in optimal progressivity even with a flat age-wage profile and no uninsurable risk.

Very 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 is the paper by Stantcheva (forthcoming), who studies

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

optimal Mirrleesian taxation over the life cycle in a model with endogenous human capital formation. Her analysis has a different focus from ours because she studies the role of human capital in increasing or reducing wage risk, depending on whether or not human capital is a complement to exogenous –and risky– labor productivity. Her study has novel predictions about how observable education expenses should be deducted from tax liabilities over the life cycle, a dimension of policy we abstract from, since in our model the skill investment cost is entirely in utility terms.

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 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 quantifies the welfare gain of introducing such fiscal tools. Section 8 concludes.

2 Economic Environment

Demographics: The model has a standard over-lapping 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 $i \in [0, 1]$.

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.

[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 and distributed uniformly across households, such as public education.

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

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

where $\exp[(1+\sigma)(\varphi_i + \bar{\varphi}_a)]$ measures the disutility of work effort. The profile $\{\bar{\varphi}_a\}$ 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 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 $s \in [0, \infty)$,

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

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

[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.

in the context of imperfect substitutability across different skill types.

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, $\varepsilon_{ia} \sim \mathcal{N}\left(-v_{\varepsilon a}, v_{\varepsilon a}\right)$, where $v_{\varepsilon 0}$ captures the variance at age zero.

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 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) skills accumulated 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 $B_{ia}(\varepsilon)$ and $Q(\varepsilon)$ denote the quantity and the price, respectively, of insurance claims purchased that pay one unit of consumption if and only if $\varepsilon \in \varepsilon \subseteq \mathbb{R}$. In Section 7 we introduce borrowing and lending, solve for the equilibrium numerically, and explore how this alternative market structure changes optimal tax policy.[4]

Labor and goods markets: The final consumption good and all types of labor services are traded in competitive markets. The final good is the numeraire 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 G 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 net tax revenues at income level y for 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 (8) with age-invariant parameters has a long tradition in public finance (Feldstein 1969; Persson 1983; Bénabou 2000 and 2002; Heathcote et al. 2014 and 2017). Heathcote and Tsujiyama (2016) show that in a static environment this functional form closely approximates the fully optimal Mirrleesian policy.

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 :

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

[4] 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. In the present model, age variation in efficiency and disutility $(x_a, \bar{\varphi}_a)$ and in the tax parameters τ_a and λ_a introduce motives for intertemporal borrowing and lending.

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.

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 $y_a^0 = (\lambda_a)^{\frac{1}{\tau}} > 0$, 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 at age a given this tax and transfer schedule is

$$E\left[MTR_a\right] = 1-\lambda_a(1-\tau_a)\frac{\int (y_{ia})^{1-\tau_a}\,di}{\int y_{ia}di}. \tag{11}$$

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

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

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 rate of transformation between private and public consumption is one, the aggregate resource constraint for the economy (recall population has measure 1 so aggregates 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 $(\kappa_i, \varphi_i, \varepsilon_{i0})$. Combining eqs. (2) and (3), the first-order necessary and sufficient condition for the skill choice is

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

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 agents cannot supplement or unwind past skill investments over the rest of their life cycle.

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 η_{ia} shocks open (the innovation to ε_{ia}) and the individual buys insurance claims $B(\cdot)$. Finally, η_{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

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

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

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

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 $\left(\varphi, s, a, \alpha, \varepsilon, \bar{B}\right)$, where $\bar{B} = B(\varepsilon; \varphi, s, a, \alpha)$ are state-contingent insurance payouts.[5] 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.

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, incorporating transition from the current system, the economy-wide skill distribution will vary deterministically over time, and an additional assumption is required to preserve tractability. We return 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. Insurance markets clear and the prices $Q(\cdot)$ of insurance claims equal the probabilities that the realization for ε is in the corresponding set.

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

Propositions 1 and 2 below describe the equilibrium allocations and skill prices in closed form. The payoff from analytical tractability is evident in Propositions 3 and 4,

[5] Since equilibrium $\bar{B}$ is a known function of $(\varphi, s, a, \alpha, \varepsilon)$, in what follows we omit $\bar{B}$ from the state vector.

where we derive a set of results for optimal taxation based on a closed-form expression for social welfare. In what follows, we make explicit the dependence of equilibrium allocations and prices on $(\{\tau_a\}, \{\lambda_a\})$ in preparation for our analysis of the optimal taxation problem. Moreover, from now on we express the arguments in the decision rules using the minimum set of relevant state variables.

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

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

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

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 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, that is, "hours of the representative agent of age a." This term falls with progressivity. The second captures the fact that a higher disutility of work leads an agent to choose lower hours. The third term captures that the response of hours worked to an insurable shock ε (which has no income effect precisely because it is insurable). The response here is proportional to what we label the tax-modified Frisch elasticity $(1-\tau_a)/(\sigma+\tau_a)$. This elasticity collapses to the standard Frisch elasticity $1/\sigma$ when $\tau_a = 0$, while 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 level s (because the skill price $p(s)$ is increasing in s), in the age profile of efficiency units x_a, and in the uninsurable component of wages α. Since hours worked are decreasing in the disutility of work, 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 realizations 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*

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

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 the functions π_1 and π_0 are given by

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

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

The skill investment allocation is given by

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

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

Note, first, that the log of the equilibrium skill price takes a "Mincerian" form (i.e., 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 progressivity increases the equilibrium *pre-tax* marginal return $\pi_1(\bar{\tau})$. The logic 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 in (φ, ε_0), of the age profiles $(x_a, \bar{\varphi}_a)$, and of risk $(v_\alpha, v_\varepsilon)$. 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* $\pi_1(\bar{\tau}) \cdot s(\kappa; \bar{\tau})$ *is exponential with parameter* θ. *Thus, the variance of log skill prices is*

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

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 utility specification.

Since the exponent of an exponentially distributed random variable is Pareto, the distribution of skill prices in levels is Pareto with parameter θ. 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, they have Pareto right tails, a robust feature of their empirical counterparts (see, e.g., Atkinson, Piketty, and Saez 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* $H(\{\tau_a\}) = \frac{1}{A}\sum_{a=0}^{A-1} H(a, \tau_a)$, $\bar{N}(\{\tau_a\}) =$

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

$$\begin{aligned} H(a,\tau_a) &= \mathbb{E}[h(\varphi,a,\varepsilon)] && (23)\\ &= (1-\tau_a)^{\frac{1}{1+\sigma}} \cdot \exp(-\bar{\varphi}_a) \cdot \exp\left[-\sigma\left(\frac{1-\tau_a}{\sigma+\tau_a}\right)^2 \frac{v_{\varepsilon a}}{2}\right], \\ N(a,\tau_a) &= \mathbb{E}[\exp(x_a+\alpha+\varepsilon)h(\varphi,a,\varepsilon)] && (24)\\ &= (1-\tau_a)^{\frac{1}{1+\sigma}} \cdot \exp\left[x_a - \bar{\varphi}_a + \left(\frac{1-\tau_a}{(\sigma+\tau_a)^2}(\sigma+2\tau_a+\sigma\tau_a)\right)\frac{v_{\varepsilon a}}{2}\right]. \\ Y(a,\tau_a,\bar{\tau}) &= \mathbb{E}[p(s,\bar{\tau})] \cdot N(a,\tau_a) && (25) \end{aligned}$$

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

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 weighs 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. Under these assumptions, social welfare can be written as

$$\mathcal{W}(\{\tau_a,\lambda_a\},g;\Gamma_{-1}) \equiv (1-\beta)\left[\sum_{j=-(A-1)}^{-1} \beta^j U_{j,0}^{old}(\{\tau_a,\lambda_a\},g;\Gamma_{-1}) + \sum_{j=0}^{\infty} \beta^j U_{j,0}(\{\tau_a,\lambda_a\},g)\right], \tag{26}$$

where $U_{j,0}(\cdot)$ is remaining expected lifetime utility (discounted back to date of birth) as of date 0 for the cohort that enters the economy at date j. The superscript 'old' distinguishes the existing cohorts ($j<0$) already alive at the time of the reform – whose skill investments were made under the old age-invariant government policy $\Gamma_{-1} := (\lambda_{-1},\tau_{-1},g_{-1})$ – from future cohorts ($j \geq 0$) whose skill investments are made under the new optimal system.[6] Note that remaining lifetime utility $U_{j,0}^{old}$ 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. The constant $(1-\beta)$ pre-multiplying the summation is a convenient normalization. The 'Ramsey problem' of the government is to choose $(\{\tau_a, \lambda_a\}, g)$ in order to maximize (26) subject to the government budget constraint (12) ,where lifetime utilities are given by (2) and equilibrium allocations are given by (17), (18) and (22).

We define social welfare in steady state $\mathcal{W}^{ss}(\{\tau_a, \lambda_a\}, g)$ to be equal to average utility in a cross-section:

$$\mathcal{W}^{ss}(\{\tau_a, \lambda_a\}, g) = \frac{1}{A}\sum_{a=0}^{A-1}\mathbb{E}\left[u\left(c\left(\varphi, s, a, \alpha\right), h\left(\varphi, a, \varepsilon\right), G\right)\right] - \mathbb{E}\left[v\left(s(\kappa), \kappa\right)\right], \tag{27}$$

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 irreversibility of the existing stock of skills induces transitional dynamics towards the new steady state. Moreover, 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 (see, e.g. Hassler et al. 2008 for an analysis of Ramsey taxation of human capital).[7]

It is straightforward to show that true social welfare (26) is proportional to steady state welfare (27) in two special cases. The first of these 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. In particular, note that when $\beta = 1$ social welfare is simply expected lifetime utility for a cohort entering in the new steady state, U_0^{ss}. Then note

[6]Because the planner discounts across generations at the same rate that individuals discount over time, all agents alive at the time of the reform (the 'old') receive equal weight (one) on their residual expected lifetime utility from that date (date zero) onward.

[7]We have also studied an alternative approach, which is to assume that the choice of skills is fully reversible at any point. This alternative assumption implies that transition following a tax reform is instantaneous: given a choice for the new policy, the economy immediately converges to the steady-state distribution of skills associated with this policy. In our view, irreversible skill investment is arguably the more realistic case.

that in the expression for lifetime utility $U_{j,0}$, 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 $\theta \to \infty$, so that skills are perfect substitutes and there is no skill investment. 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.

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 has also the advantage that we can derive a number of analytical results for optimal taxation. Next, we analyze the optimal policy choice allowing for discounting and transitional dynamics.

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 output share of government expenditures 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 only depends only on households' relative taste for the public good χ. Since g does not 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 representative agent version of our economy. 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.'

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 (and hence consumption, with separable utility) across age groups. This result indicates that the government, through the sequence of $\{\lambda_a\}$ can effectively replicate the role of life-cycle borrowing and saving, absent in the model by assumption, in smoothing predictable life-cycle income variation.

Exploiting these two results, one can substitute the optimal decisions for g^* and $\{\lambda_a^*\}$ into $\mathcal{W}^{ss}$ and, by plugging in the closed-form expressions described above for equilibrium allocations, one can express steady state welfare analytically as a function of model parameters and of the vector of age-dependent progressivity $\{\tau_a\}$. 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 social welfare function $W^{ss}(\{\tau_a\})$ 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:

$$
\begin{aligned}
0 \;=\; & \frac{1}{\theta-1+\tau_a^*} - \frac{1}{\theta} + (1-\tau_a^*)(v_\varphi + a v_\omega) + \frac{1}{1+\sigma} + \\
& - \left[\left(\frac{1+\chi}{\theta-1}\right)\frac{1}{1-\bar{\tau}(\{\tau_a^*\})} - \frac{1}{\theta}\right]\frac{\psi}{1+\psi} \\
& - \left(\frac{1+\chi}{1+\sigma}\right)\left[\frac{1}{1-\tau_a^*} + \left(\frac{\sigma+1}{\sigma+\tau_a^*}\right)^3 \tau_a^* v_{\varepsilon a}\right]\frac{N(a,\tau_a^*)}{\bar{N}(\{\tau_a^*\})},
\end{aligned}
\tag{28}
$$

(ii) The optimal sequence $\{\tau_a^*\}$ is age invariant if the following four 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), (3) the age profiles of efficiency units and disutility of work $\{x_a\}, \{\bar{\varphi}_a\}$ are 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), introducing age-invariant insurable risk ($v_{\varepsilon 0} > 0$) maintains a flat profile for τ_a^* but it pushes it toward zero. If the variance of insurable risk increases with age ($v_{\varepsilon,a+1} > v_{\varepsilon,a}$) and if $\tau_a^* > 0$ at age a,

then $\tau^*_{a+1} < \tau^*_a$.

(v) Relative to the parameterization described in (ii), introducing age variation in efficiency units 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 features the closed form expression for the steady-state welfare function. Each term can be given an intuitive economic interpretation, along the lines of the analysis contained in Heathcote et al. (2017). We now illustrate the other results of Proposition 4, one by one.

(ii) In this benchmark case with $\beta = 1$ (or also when $\theta \to \infty$), the FOC simplifies to an expression where age a does not appear, hence τ^*_a is constant.[8] In particular, when $\theta \to \infty$ the FOC 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 a regressive tax system induces higher labor supply and thereby corrects a public good externality.

(iii) Now consider the role of uninsurable risk. To isolate this force, we focus on the case where 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 when uninsurable risk is larger. Therefore, when $v_\omega > 0$, optimal progressivity increases with age, *ceteris paribus*. We label this force the *uninsurable risk channel*.

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

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

average effective marginal tax rate has a positive drift over the life cycle. Farhi and Werning (2013) analyze Mirrlees taxation in a dynamic life-cycle economy. Their environment is a special case of ours, with no endogenous skill accumulation.[9] In their numerical example, in which average labor productivity does not vary with age, the optimal history-dependent tax scheme has similar qualitative features to the optimal policy in our model (see Farhi and Werning 2013, Figure 2). Namely, the average effective marginal tax rate is increasing in age, average output is decreasing in age, and consumption is invariant to age.[10]

(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 (A16) 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 $v_{\varepsilon 0}$. It is immediate that there is no motive for age variation in τ_a, i.e., $\tau_a^* = \tau^*$. In addition, if $\tau^* > 0$, then increasing $v_{\varepsilon 0}$ will reduce optimal progressivity, while if $\tau^* < 0$, increasing $v_{\varepsilon 0}$ will 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. We label this force the *insurable risk channel*

(v) Now consider the role of the life-cycle profiles of efficiency units and 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 ($\theta \to 0$,

[9] They also assume no preference heterogeneity and no valued government expenditures.

[10] Golosov, Troshkin, and Tsyvinski (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.

$v_\varphi = v_{\varepsilon a} = v_\omega = 0$). The optimal value for τ at age a, τ_a^*, is then the solution to the following simplified version of the first-order condition (A16):

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

This 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 therefore 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 the labor wedge (and thus the effective marginal tax rate) over the life cycle. The effective marginal tax rate at age a in this version of the model simplifies to $1-\lambda_a(1-\tau_a)y_a^{-\tau_a}$, where earnings y_a are given by $\exp(x_a-\bar{\varphi}_a)(1-\tau_a)^{\frac{1}{1+\sigma}}$. When $x_a-\bar{\varphi}_a$ and thus earnings are increasing with age, the planner wants to have λ_a decrease with age in order to equate consumption across age groups. Absent age variation in τ_a this 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 driven by the profile of* $\{x_a - \bar{\varphi}_a\}$ *. 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] Section 7 contains an extension where we allow individuals to access a non-state-contingent bond subject to a credit limit.

5.2 Welfare with discounting and transitional dynamics

The steady state welfare expression is tractable, making it easy to understand 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 thus a dynamic forward-looking decision.

We therefore now assume $\beta < 1$ and consider an unanticipated policy change at date $t = 0$ from a pre-existing age-invariant policy $\Gamma_{-1} := (\lambda_{-1}, \tau_{-1}, g_{-1})$ to a new policy regime $\{\tau_a\}, \{\lambda_{at}\}$ and $\{g_t\}$ with new age-dependent tax rates. We impose that the reform is a once-and-for-all for progressivity, in the sense that the new age profile of progressivity $\{\tau_a\}$ is constant over time.[12] However, we do allow the age profile of the proportional factors $\{\lambda_{at}\}$ and public good provision g_t to vary over time, maintaining the assumption that the government budget must be balanced each period.

To preserve tractability, we make one new minor assumption relative to the baseline model, namely that production is segregated between two sectors: one for all the cohorts born before the tax reform (who cannot adjust skill investments in response to the new tax system) and one for all cohorts who enter the economy from date zero onward while the new tax system is in place. As time passes post tax reform, the share of the total population in the first group declines, and eventually after A years the entire population belongs to the second group. Each sector is otherwise similar to the economy laid out in Section 2, including the sector-specific production function in eq. (5). This segregation assumption ensures that the distribution of skills in each sector is always exponential.[13]

[11] This effect would also 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).

[12] This choice is made to simplify the problem. An interesting extension of our analysis would be to also allow the age profile of progressivity to vary with time.

[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 young and old working in the same market would be that the new cohorts potentially might make human capital investments that are different from those of the cohorts born before the reform, in which case the old and new skill distributions would differ and the combined overall distribution of skills would no longer be exponential. The assumption that individuals born before the reform work in a different market

The equilibrium hours worked and consumption allocations in this version of the economy are the same as in the model above, that is, given by eqs. (17)-(18), with one exception: the price of skills now differs between the cohorts who invested in skills before the reform ("the old") and the cohorts who invest after the reform ("the young"). The skill investment and skill prices for the young, $s(\kappa;\bar{\tau})$ and $p(s;\bar{\tau})$, depend on $\bar{\tau}(\{\tau_a^*\})$, which is the discounted progressivity of the new tax progressivity profile $\{\tau_a^*\}$. In contrast, the choices for the old are sunk and were determined by the tax progressivity before the reform, τ_{-1}. Thus, their skill choices $s(\kappa;\tau_{-1})$ prices $p(s;\tau_{-1})$ are exactly as in eq. (19) of Proposition 2 with $\bar{\tau} = \tau_{-1}$.

How does incorporating transition change the optimal policy prescription? Incorporating transition adds a new driver shaping the optimal age profile of progressivity, which we label the *sunk skill investment channel*.

Proposition 5 [optimal age dependent taxation with transition]. *When $\beta < 1$, taking the transition into account, the optimal tax system has the following properties:*
(i) The optimal output share of government expenditures g_t^* is constant and given by

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

(ii) At every date t, the optimal sequence $\{\lambda_{at}^*\}$ equalizes average consumption across age groups.

(iii) When $v_\varphi = v_\omega = v_{\varepsilon a} = 0$ and the age profiles for efficiency and disutility of work are flat, i.e. the only source of heterogeneity is skill investment, and $\beta < 1$ the profile for $\{\tau_a^*\}$ is increasing with age.

Part (i) establishes that optimal public good expenditures during transition are the same as in steady state – a constant fraction g^* of output Y_t. This result stems from the fact that g remains additively separable from the other policy instruments in the social welfare function.

Part (ii) restates that the optimal time-varying sequence $\{\lambda_{at}^*\}$ simply ensures that average consumption is equated across age groups at each point in time. Thus also this property of the policy that maximizes steady-state welfare again extends to the case that incorporates transition.

(iii) How does incorporating transition affect the optimal profile for τ_a? Since incorporating transition does not fundamentally change how the risk and life-cycle pro-

ensures that the distribution of skills remains exponential *in each sector*.

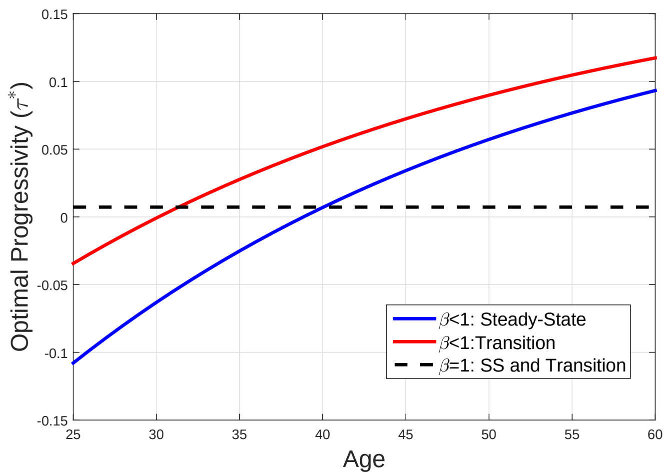

Figure 2: Optimal age dependence of taxes with no risk ($v_\omega = v_{\varepsilon a} = 0$), a flat age profile of wages ($x_a - \bar{\psi}_a = 0$), and discounting ($\beta < 1$) under a steady-state welfare objective and a welfare objective that incorporates transition. The exact parameterization for this example is the one displayed in Table 1.

ductivity channels affect age dependence in progressivity, we now assume away these channels in order to isolate the role of irreversible skill investment. Figure 2 illustrates optimal profile assuming $\beta = 1$ (in which case transition is irrelevant) and in a case with $\beta < 1$.

The key result is that the optimal policy assuming $\beta < 1$ and incorporating transition features an increasing profile for τ_a. The reason is that the planner wants to expropriate past skill investments, without excessively disincentivizing new skill investment by future generations. An age-increasing profile for τ_a achieves this. To see why, suppose we start with an age-invariant profile for τ, and consider the impact of increasing progressivity at $a = 30$ versus at age $a = 0$. A higher value for τ_{30} will compress income inequality across thirty consecutive generations before this parameter affects the taxes actually paid by any cohorts who made skill investment decisions under the new tax policy. In contrast, a higher value for τ_0 will do nothing to compress income inequality within all cohorts born prior to the tax reform. Thus, high values for τ_a at older ages offer a better equity-efficiency tradeoff to the planner. At the same time, the planner can use low values for τ_a at younger ages to offset the disincentivizing effects of high τ_a at older ages on skill investment by future generations. The result is an upward-sloping profile for τ_a. As $\beta \to 1$, the planner effectively puts a vanishingly small welfare weight on generations born prior to the tax reform, killing this mechanism. In the Appendix, we prove this result formally.

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. 3
$\{\bar{\varphi}_a\}$	Age profile for disutility of work	See Fig. 3
τ^{US}	US rate of progressivity	0.181

Table 1: Model parameterization

We call this new force the *sunk skill investment* channel. The connection with the well known result in the study of capital income taxation that the planner has an incentive to expropriate existing capital is immediate. Our new insight is that this same incentive, in the context of an OLG economy with irreversible skill investment, creates a role for age-dependent progressivity.

[GV: NEED TO DO PROOF WHICH I THINK IS DOABLE NOW. IT IS ENOUGH TO SHOW THAT THE WELFARE OF THE OLD IMPLIES TAU INCREASING BECAUSE WE KNOW HOW TO SHOW THAT THE WELFARE OF THE YOUNG IMPLIES FLAT TAU]

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 under $\beta = 1$, as in Section 4. 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. When not exogenously set to 1, the discount factor β equals 0.97. 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.[14] The preference weight on public good χ is identified directly from the size of the U.S. government as a share of GDP, assuming that the choice of public good provision in the data is optimal: given $g = 0.19$, we obtain $\chi = 0.233$.[15] For calibration, we need the actual value of τ in the US system. Based on the estimates of Heathcote et al. (2017), we set $\tau^{US} = 0.181$ and assume λ^{US} is such that the budget is balanced given g.[16] 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 from the model. 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 wages, hours, and consumption, which we measure from the the Panel Study of Income Dynamics (PSID) and the Consumer Expenditure Survey (CEX). The identification follows from the closed-form expressions for wages, hours and consumption derived above.

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

$$\begin{aligned} var_a\left(\log w_{ia}\right) &= \frac{1}{\theta^2} + v_\omega a + \left(v_{\varepsilon a}\right) \\ var_a\left(\log h_{ia}\right) &= v_\varphi + \left(\frac{1-\tau^{US}}{\sigma+\tau^{US}}\right)^2 \left(v_{\varepsilon a}\right) \\ var_a\left(\log c_{ia}\right) &= \left(1-\tau^{US}\right)^2 \left(v_\varphi + \frac{1}{\theta^2} + v_\omega a\right) \\ cov_a\left(\log h_{ia}, \log w_{ia}\right) &= \left(\frac{1-\tau^{US}}{\sigma+\tau^{US}}\right) v_{\varepsilon a} \end{aligned} \tag{29}$$

[14]See Ndiaye (2018) for a thorough analysis of optimal age-dependent taxation in a model that incorporates the retirement decision.

[15]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.

[16]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 U.S. tax code – medical expenses, mortgage interest, state taxes paid, and charitable contributions – and subtract them from gross income.

Post-government income $\tilde{y}$ 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.

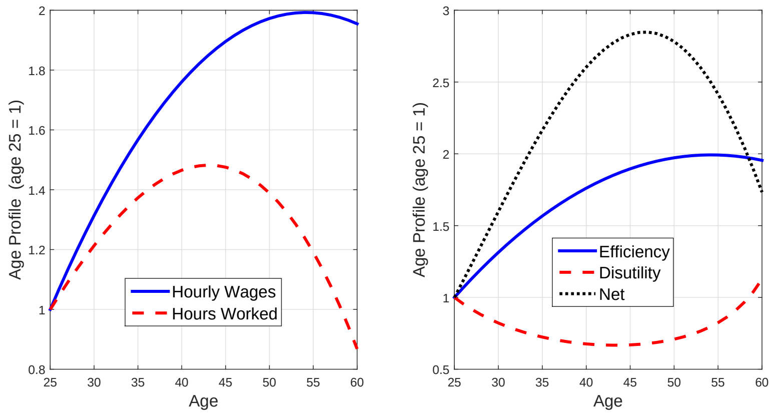

Figure 3: Left panel: life-cycle profile of individual wages and hours. Right panel: implied profiles for x_a, $\bar{\varphi}_a$, and their difference.

The moments $cov_a(\log h_{ia}, \log w_{ia})$ observed at ages $a = 0, ..., A-1$ identify $v_{\varepsilon a}$. Since in the data the profile for this variance turns out to increase with age nearly linearly, we estimate freely the initial variance at age 25, $v_{\varepsilon 0}$, and then impose linearity ex ante. From $var_a(\log h_{ia})$ we then identify v_{φ}. The value for $var_0(\log c_{i0})$ identifies θ since $v_{\alpha 0} = 0$ (a normalization). Then, the change in $var_a(\log w_{ia})$ 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 τ and θ, where the return to skills is increasing in progressivity and decreasing in skill substitutability (see equation 20). In Heathcote et al. (2017), we exploit changes in π_1, τ and θ over time, which we can measure from 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 of productivity and disutility of work. The life-cycle profile of individual hourly wages is taken from Rupert and Zanella (2015, Figure 7) and the one for hours worked is taken from McGrattan and Rogerson (2004, chart 1). The left panel of Figure 3 plots both profiles, interpolated using a cubic function of age. The wage profile maps directly into the efficiency profile for $\{x_a\}$. Given $\{x_a\}$ and the other parameter values, from the expression for average hours worked by age in (24), we can recover residually the profile for disutility of work $\{\bar{\varphi}_a\}$.

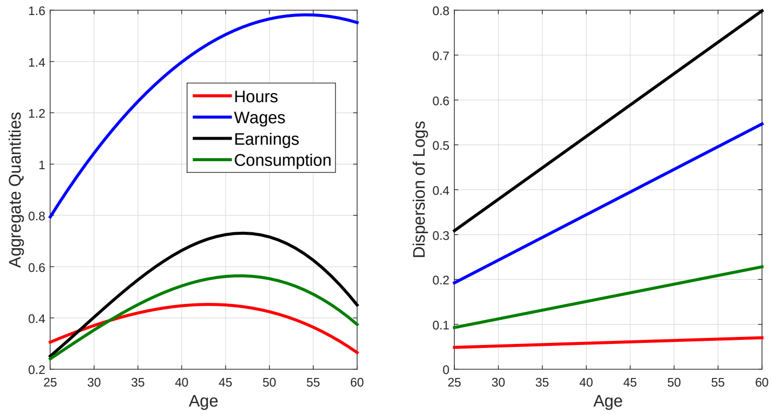

Figure 4: Means (left panel) and variances of logs (right panel) over the life cycle

The right panel of Figure 3 plots the 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. It is worth noting that our approach to identifying $\{\bar{\varphi}_a\}$ hinges on our assumption of no inter-temporal borrowing and lending. If households could perfectly smooth consumption by borrowing and lending, one would naturally expect co-variation 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 consider optimal policy assuming no age variation in preferences.[17]

Table 1 summarizes the parameter values. Figure 4 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

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

Recall that Proposition 4 identified four different forces that shape the optimal age profile of tax progressivity in steady state: discounting, uninsurable risk, insurable

[17] With free borrowing and lending, one would expect life-cycle growth in hours to be proportional to life-cycle growth in wages, absent age-variation in preferences. In addition, the Frisch elasticity of labor supply would determine the constant of proportionality. Figure 4 indicates that hours grow roughly half as much as wages between ages 25 and 45, which is consistent – given the joint assumptions of no preference variation and free borrowing and lending – with our assumed Frisch elasticity of 0.5.

risk, and life-cycle productivity. To understand the quantitative role of each of these, we start from an economy where none of these channels is active, the one described in point (ii) of Proposition 4.

6.2.1 Channels that do not induce age dependence

Figure 5 illustrates optimal progressivity $\{\tau_a^*\}$ and the implied income-weighted average marginal tax rate by age (left panels) together with earnings, hours, and consumption by age (right panels).

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

Next, in the middle panel, 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 panel, 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 progressive system. Given our parameter values, this latter force dominates and optimal progressivity rises.

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

6.2.2 Discounting channel

The first motive for age-varying taxation identified in Proposition 4 is that progressive taxation is less distortionary for skill investment at older relative to younger ages, because individuals discount future taxes when making human capital investments. To gauge the quantitative importance of this effect, we change value of the discount factor from $\beta = 1$ to $\beta = 0.97$.

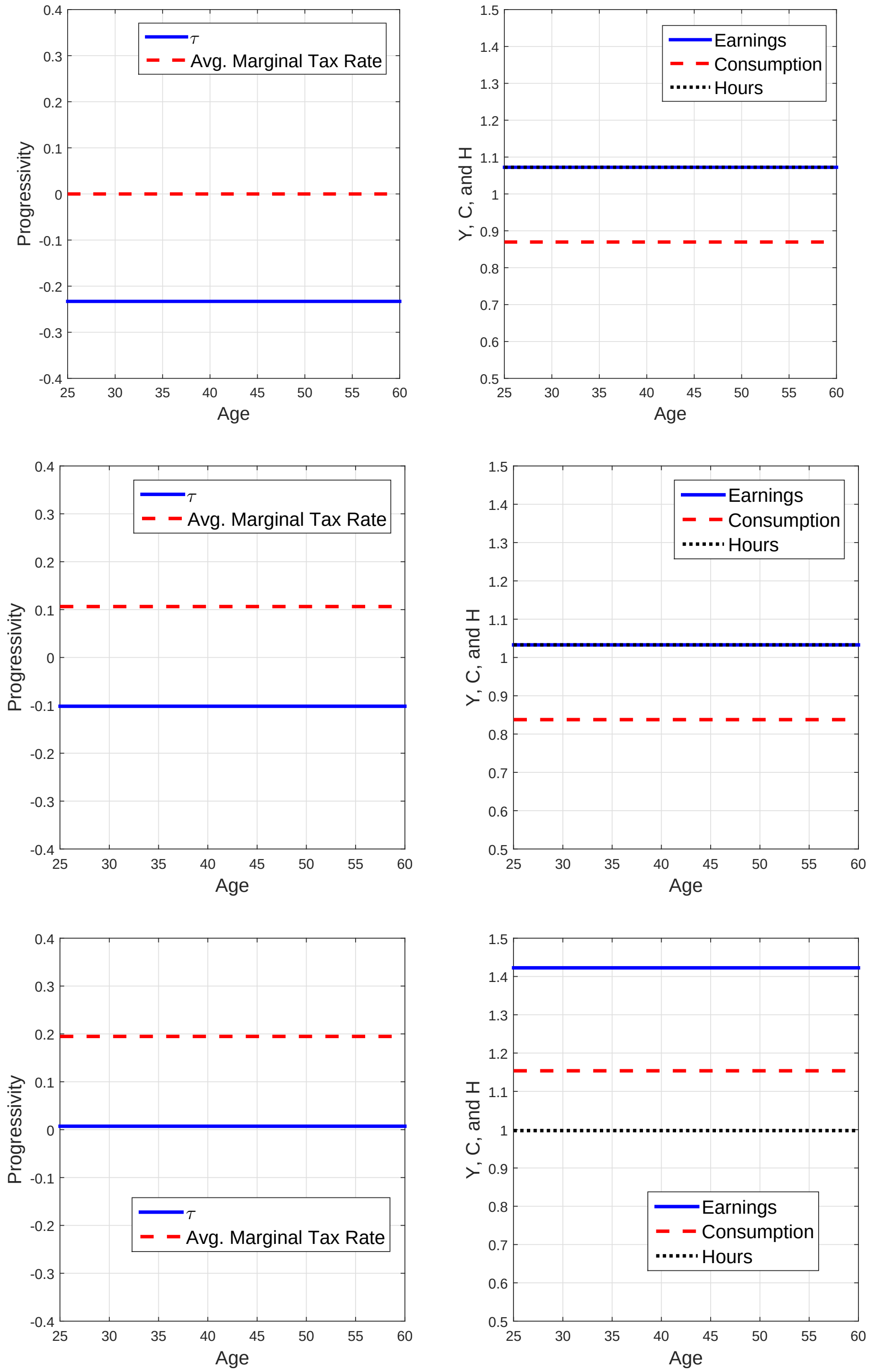

Figure 5: Left column: Optimal progressivity and income weighted average marginal tax rate. Right column: Average earnings (Y), hours (H), and consumption (C) by age. Top row: Representative agent model. Middle row: Previous case plus heterogeneity in disutility of work. Bottom row: Previous case plus heterogeneity from skill investment.

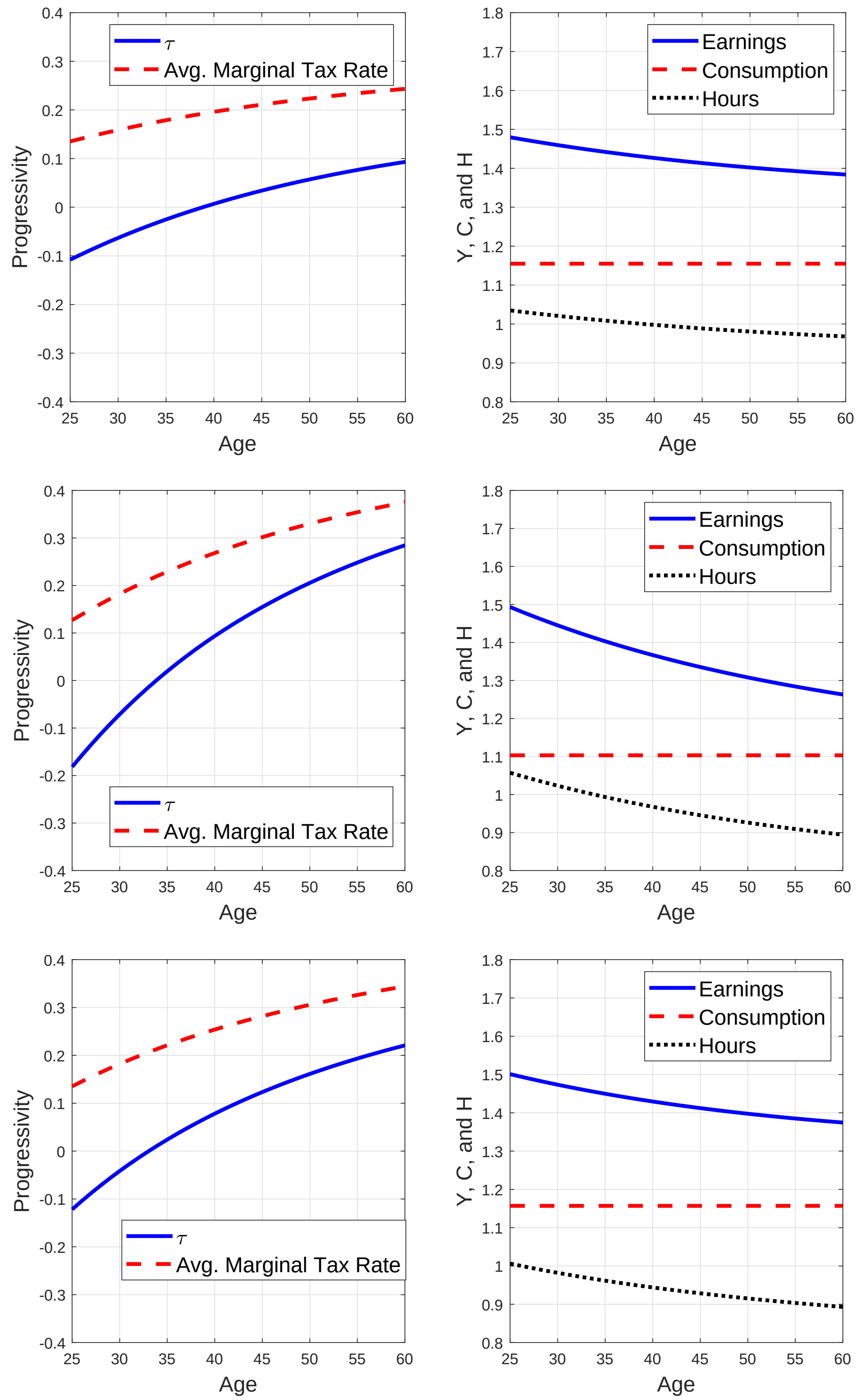

Figure 6: Left column: Optimal progressivity and income weighted average marginal tax rate. Right column: Average earnings (Y), hours (H), and consumption (C) by age. Top row: Previous case plus discounting. Middle row: Previous case plus uninsurable risk. Bottom row: Previous case plus insurable risk.

The first row of Figure 6 shows that optimal tax progressivity is now monotonically increasing in age. The effect is substantial: the optimal tax schedule is now regressive for the youngest cohorts and becomes progressive for older cohorts. As a consequence, hours worked (and therefore earnings) decline with age.

6.2.3 Uninsurable risk channel

Part (iv) of Proposition 4 states that, since uninsurable risk in the form of permanent shocks cumulates over the life cycle, the planner has an incentive to increase tax progressivity over the life cycle. To introduce this effect, we set the amount of uninsurable risk v_ω to its calibrated value.

The middle row of Figure 6 illustrates that the addition of uninsurable risk has two effects. First, the average level of optimal progressivity rises. Second, as expected, its profile becomes steeper.

6.2.4 Insurable risk channel

According to part (v) of Proposition 4, age-invariant insurable risk ($v_{\varepsilon 0} > 0$) pushes optimal progressivity toward zero, while if the variance of insurable wage risk increases with age, the planner has an incentive to tilt the schedule for optimal progressivity downward. The last row of Figure 6 illustrates 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.

6.2.5 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 7 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 (A15) 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 U-shaped profile for τ_a encourages labor supply around age 45, when labor productivity

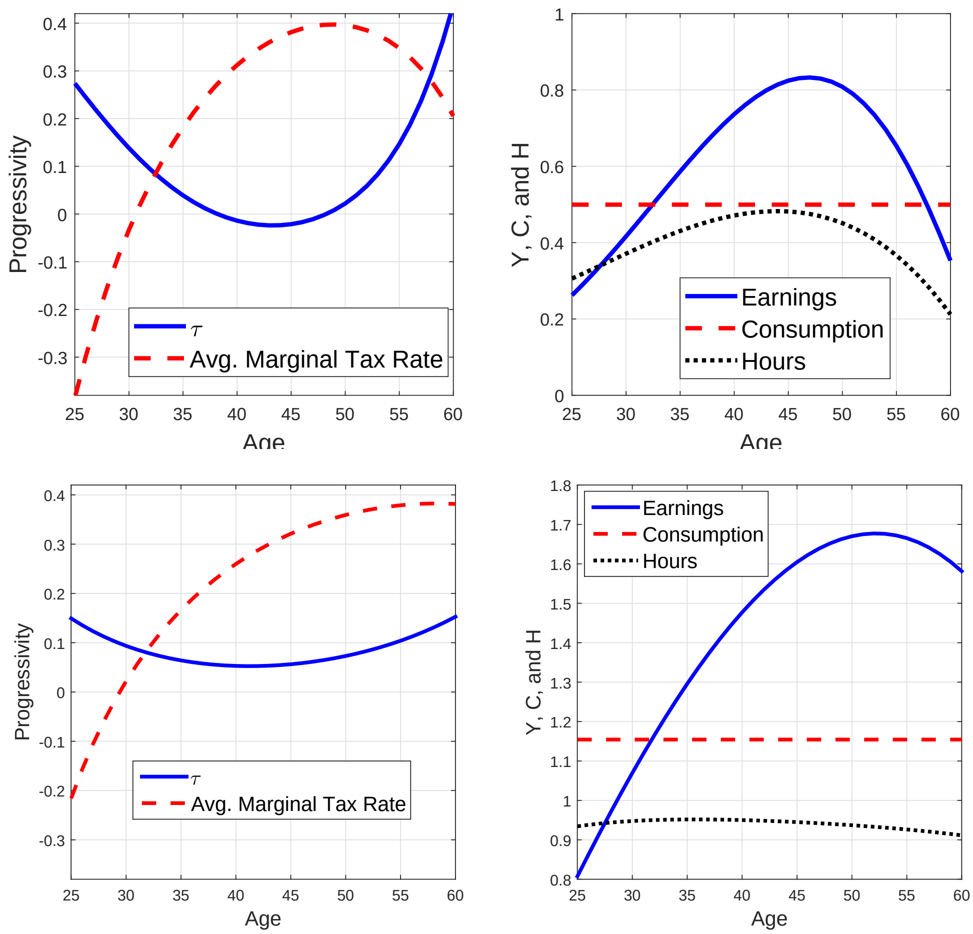

Figure 7: Left column: Optimal progressivity and income weighted average marginal tax rate. Previous case plus life cycle channel, i.e. all channels operational with steady-state welfare. Right column: Average earnings (Y), hours (H), and consumption (C) by age. Top row: age profile for disutility of work as estimated in the data. Bottom row: age profile for disutility of work constant.

net of the utility cost of work is highest. A complementary 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 7 show that the life-cycle channel is substantially weaker when we shut down age variation in preferences.

All channels are now operative, so this economy 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.

6.2.6 Transitional dynamics and the sunk skill investment channel

We now compute the age-dependent tax system that maximizes welfare taking into account transitional dynamics and the sunk investment channel. The importance of the last 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 corresponding value for λ that balances the budget with g set at the estimated value for the U.S. economy. Recall that g does not change along the transition (point (i) of Proposition 5).

The left panel of Figure 8 plots optimal age-dependent progressivity when the planner takes the transition to the new steady state into account and compares it with the optimal policy that maximizes steady state welfare. Two key differences arise when we incorporate transition in the analysis: the average level of tax progressivity is higher, and the age profile of optimal progressivity is flatter.

The right panel plots the implied optimal marginal tax rates for a range of values of income for three age groups. The age-dependent optimal tax system dictates essentially a flat tax for middle-aged workers, and a highly progressive schedule for young and old. High progressivity at young ages is optimal, because the young are relatively unproductive, so they high progressivity does not translate into high marginal tax rates. High progressivity at older ages reflects various forces: the incentive to expropriate past skill investments without disincentivizing new investment, the desire to provide social insurance against permanent labor market risk, and the high disutility of work near retirement age which reduces earnings and thus effective marginal tax rates. Finally, note that even though the degree of optimal progressivity (implied by the value of τ_a) is very similar between young and old, marginal (and average) tax rates are significantly higher for the old. 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.

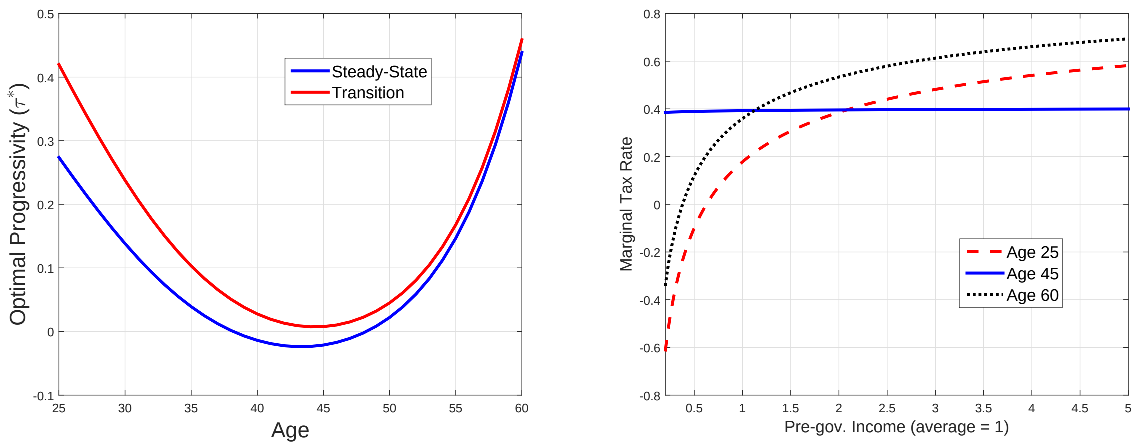

Figure 8: Left panel: Optimal age dependent τ_a^* under transition and steady-state welfare. Right panel: Marginal tax rates at different ages for the optimal age-dependent policy. These marginal tax rates are computed with the value of λ_a^* and of the income distribution in the final steady-state.

6.2.7 Welfare gains from tax reforms

We present the welfare gains of switching from the existing tax/transfer system to the optimal age dependent system in steps. First, we report 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. 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 negligible. Allowing λ_a to be age-varying increases the gains substantially – by three percentage points – because it allows the planner to equalize consumption across age groups. Finally, allowing progressivity to vary by age adds another 0.7 percentage points of welfare gains.

7 An economy with intertemporal trade

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 inter-temporal consumption smoothing.

	Benchmark	U.S. BL	Natural BL
(λ^*, τ^*) constant	0.04	0.15	0.15
λ^* age-varying, τ^* constant	3.00	1.88	1.43
(λ^*, τ^*) age-varying	3.70	2.12	1.47

Table 2: All numbers in the table are welfare gains expressed as additional lifetime consumption (pct points) relative to the existing tax/transfer system. The column 'Benchmark' refers to the benchmark economy without intertemporal trade. The column 'U.S. BL' refers to the economy with borrowing and lending under the calibrated borrowing limit for the U.S. economy (2 times annual earnings). The column 'Natural BL' refers to the economy with borrowing and lending under the natural borrowing limit.

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, both the equilibrium and, more importantly, the optimal age-dependent tax system have to be computed numerically. The latter 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 need to decide how to tax them. We assume that taxable income at age a includes capital income rb_a, but that savings $b_{a+1} - b_a$ are tax-deductible. Thus, the parametric tax / transfer function now applies to taxable income $p(s)\exp(x_a + \alpha)h_a + rb_a - (b_{a+1} - b_a)$ and the individual budget constraint therefore becomes:

$$c_a = \lambda_a \left[p(s)\exp(x_a + \alpha)h_a + (1 + r)b_a - b_{a+1}\right]^{1-\tau_a}$$

This assumption allows us to retain a closed form solution for the equilibrium skill price function $p(s)$.

The dynamic program of the working age household characterizing optimal consumption/saving and labor supply decisions has five state variables: age a, skills s, disutility of work φ, permanent productivity shocks α, and wealth b. The second assumption we make is that the borrowing limit for an individual can be written as:

$$b_{a+1}(s, \alpha, \varphi) \geq -\bar{b}_a \cdot p(s) \exp(x_a + \alpha - \varphi)$$

If the borrowing constraint has this form, we can show that optimal individual saving and labor supply decisions are the solution to a simpler household problem with two states, age and normalized wealth, defined as wealth relative to the wage net of disutility of work, $p(s) \exp(x_a + \alpha - \varphi)$.

Every other element of the baseline model is unchanged. The parameterization is the same as the one in Table 1, with the exception that the variances of the insurable risk terms are zero.

The only new parameters in this extended model are the age-dependent borrowing limits. When the borrowing limits $\bar{b}_a$ 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 this case, the only binding constraint over the life cycle is $b_{A+1} \geq 0$, i.e. a No-Ponzi condition stating that the household cannot die with negative wealth.[18]

To set borrowing limits that realistically capture how much 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 HELOCs. We begin by adding up all these limits and doubling them to allow for the fact that many households have untapped credit they could attain by asking for higher limits on existing credit cards/HELOCs, or by applying for new cards or lines of credit.

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 installment loans by 2.5 to reflect the fact that the typical length of such loans is five years. Loans against IRAs have, instead, a very short-term nature (typically 60 days).

[18] In practice, to solve the model numerically, it is useful to compute what the natural borrowing limit is at each age to make sure the lower bound for the asset grid is not too tight.

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. Our calculations imply that almost 20 percent of U.S. households have a credit limit of zero. Conditional on a positive credit limit, the median is about 0.75 times annual earnings. To demonstrate that our results from the benchmark economy are robust to introducing extensive borrowing and saving we set the credit limit to twice annual household earnings, which corresponds to the 85th percentile of the distribution of 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 three and optimize over its four coefficients.[19]. Moreover, we assume that the planner maximizes steady-state welfare.

7.1 Results

Figure 9 summarizes the optimal tax scheme in three cases: autarky, the natural borrowing limit, and the calibrated borrowing limit. The age profile for τ_a^* in the autarky case essentially coincides with that in our benchmark model modulo, again, the absence of insurable risk.

Under the natural borrowing limit, the U-shape in progressivity essentially disappears. The logic is that households can now borrow and save to almost perfectly smooth consumption over the life cycle, and thus the life cycle channel is greatly weakened. The optimal age profile of τ_a^* now resembles the one arising in the economy with ex-ante heterogeneity in disutility of work and skill investment and ex-post heterogeneity in permanent uninsurable shocks. Indeed, note that because of the permanent nature of risk, allowing for intertemporal trade does not improve smoothing of these shocks relative to autarky.

One finding that might appear surprising is that even with loose borrowing limits, the planner still wants to have λ_a decrease with age. The logic is that absent age variation in either λ_a or τ_a, households would choose an upward sloping profile for life-cycle consumption, because the equilibrium interest rate exceeds the household's rate of time preference. The planner, in contrast, wants to equate consumption by age, because he perceives he can use λ_a to costlessly reshuffle consumption across age groups. Thus the planner has λ_a decrease with age. This is a force toward having marginal tax rates increase with age, which moderates the planner's desire to implement an increasing profile for τ_a.

[19] We verified that polynomials of higher degree yields only negligible welfare gains

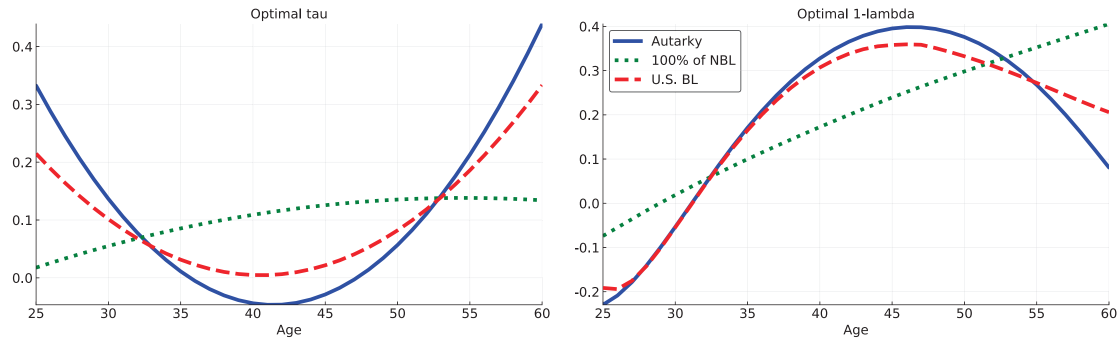

Figure 9: Left panel: Optimal age profile for τ_a^* under different values of the borrowing limits. Right panel: Optimal profile for $1 - \lambda_a^*$ which determines the level of taxation by age.

The optimal profile for progressivity under the U.S. borrowing limits sits in between these two extremes, but it is much closer to the autarky/benchmark case in that it is flatter but retains a pronounced U shape. Thus, the tractable autarky case offers quantitatively useful guidance about optimal policy even in the more realistic but less tractable model with borrowing and lending.

The last two columns of Table 2 summarize the welfare gains from switching to the optimal tax system in these economies. Under the U.S. borrowing limit, the gains from moving to an age-dependent tax system are somewhat smaller than in the autarky case, but still substantial at around 2 percent of lifetime consumption.

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. The framework, which builds on Heathcote et al. (2017), restricts the policy space to a particular functional form for the tax and transfer schedule which has been shown to provide a good representation of the U.S. tax and transfer system and which has the important advantage of making the model fully tractable. The main innovation in this paper is to allow for age-dependent tax progressivity.

We show that the optimal degree of age dependence in progressivity is driven by several forces. First, the fact that uninsurable wage dispersion is rising with age is a force toward making the tax system increasingly progressive with age. Second, the fact that average labor productivity is generally increasing over the life cycle is a force toward making optimal tax progressivity decline with age, since it is more expensive to distort labor supply for high earners. The third motive driving progressivity is that the planner can increase steady-state human capital by back-loading tax progressivity to later in the life cycle. This is a force for increasing tax progressivity with age. However, this motive is weakened when evaluating welfare incorporating transition from an initial state. Incorporating transition also pushes up average progressivity over the life cycle, reflecting the temptation to expropriate returns to existing irreversible skill investments.

When calibrating the analytically tractable economy without intertemporal trade to the United States we find that when all of these channels are operative and the optimal age profile for tax progressivity is markedly U-shaped. This U shape survives in a more realistic version of the model with borrowing and saving and plausibly calibrated credit constraints. Welfare gains of switching from the current system (not too far from the the optimal age-invariant one) to the optimal ge-dependent system exceed 2 percent of lifetime consumption.

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.

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.

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*.

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

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.

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.

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.

Persson, M. (1983): "The Distribution of Abilities and the Progressive Income Tax," *Journal of Public Economics*, 22(1), 73–88.

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

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 both 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 skill prices $p_a(s)$ 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):

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

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

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

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

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

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

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

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 price $p\left(s;\{\tau_a\}\right)$. 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[20]

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

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

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

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

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

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

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

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

$$\begin{aligned}\log p(s) &= \log\left[\frac{\partial Y}{\partial\left[\bar{N}\cdot m(s)\right]}\right] = \frac{1}{\theta}\log Y - \frac{1}{\theta}\log\left[\bar{N}\cdot m(s)\right] \\ &= \frac{1}{\theta}\log\left(\frac{Y}{\bar{N}}\right) - \frac{1}{\theta}\log\zeta + \frac{\zeta}{\theta}s.\end{aligned} \tag{A5}$$

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

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

and thus the equilibrium density of s is

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

Similarly, the base skill price is

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

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

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

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

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

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

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

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

$$N(a,\tau_a) = \exp(x_a + \frac{v_{\varepsilon a}}{\widehat{\sigma}_a}) \cdot H(a,\tau_a). \quad \text{(A11)}$$

Finally, average output of age group a is given by:

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

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:

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

A.5 Proof of Proposition 3 [optimal choice of g and λ]

It is useful to begin by computing:

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

where, after some tedious algebra, one obtains:

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

It is also useful, for what follows, define the short-hand notation

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

Recall that welfare in steady state is given by:

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

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

$$\begin{aligned}
\max_{\{g,\lambda_a,\tau_a\}} \mathcal{W}^{ss}(g,\{\lambda_a,\tau_a\}) &= \frac{1}{A}\sum_{a=0}^{A-1}\bar{u}\left(a,\lambda_a,\tau_a\right)+\chi\log\left(g\sum_{a=0}^{A-1}Y\left(a,\tau_a,\bar{\tau}\right)\right)-\bar{v}\left(\bar{\tau}\right) \\
&\quad \text{subject to} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(A12)} \\
\frac{1}{A}\sum_{a=0}^{A-1}\lambda_a\tilde{Y}\left(a,\tau_a,\bar{\tau}\right) &= (1-g)\,\frac{1}{A}\sum_{a=0}^{A-1}Y\left(a,\tau_a,\bar{\tau}\right).
\end{aligned}$$

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

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

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

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

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

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

using $(A14)$ into 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 the time. The first term in $(A12)$ can be written as:

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

Note that average log consumption for age group a is:

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

where

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

and

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

Thus:

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

Moreover:

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

Therefore, the difference between these two terms is:

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

and combining all these terms:

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

Average disutility of hours worked in age group a is:

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

The average cost of skill investment in each cohort of newborn is:

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

Combining these components, we can rewrite the social welfare function (up to a constant) only as a function of $\{\tau_a\}$ as:

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

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

Using (**??**) into the above expression for $\mathcal{W}^{ss}(g,\{\lambda_a,\tau_a\})$ we arrive at:

$$\begin{aligned}
\mathcal{W}^{ss}(g,\{\tau_a\}) &= \log(1-g)+\chi\log g-\frac{1}{A}\sum_{a=0}^{A-1}\underbrace{\frac{1-\tau_a}{1+\sigma}}_{\text{Disutility of labor}} \qquad \text{(A15)} \\
&\quad +(1+\chi)\log\underbrace{\left\{\sum_{a=0}^{A-1}(1-\tau_a)^{\frac{1}{1+\sigma}}\cdot\exp\left[x_a-\bar{\varphi}_a+\left(\frac{\tau_a(1+\hat{\sigma}_a)}{\hat{\sigma}_a^2}+\frac{1}{\hat{\sigma}_a}\right)\frac{v_{\varepsilon a}}{2}\right]\right\}}_{\text{Effective hours } \bar{N}_a} \\
&\quad +(1+\chi)\underbrace{\frac{1}{(1+\psi)(\theta-1)}\left[\psi\log(1-\bar{\tau})+\log\left(\frac{1}{\eta\theta^{\psi}}\left(\frac{\theta}{\theta-1}\right)^{\theta(1+\psi)}\right)\right]}_{\text{Productivity: log(average skill price) = log}(E[p(s)])} \\
&\quad -\underbrace{\frac{\psi}{1+\psi}\frac{1-\bar{\tau}}{\theta}}_{\text{Avg. education cost}}+\frac{1}{A}\sum_{a=0}^{A-1}\underbrace{\left[\log\left(1-\left(\frac{1-\tau_a}{\theta}\right)\right)+\left(\frac{1-\tau_a}{\theta}\right)\right]}_{\text{Cost of consumption dispersion across skills}} \\
&\quad -\frac{1}{A}\sum_{a=0}^{A-1}\underbrace{\frac{1}{2}(1-\tau_a)^2(v_\varphi+av_\omega)}_{\text{Cons. dispersion due to unins. shocks and preference heterogeneity}} .
\end{aligned}$$

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 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 equation (28) in the main text. Standard algebra yields the second-order condition

$$\begin{aligned}
\frac{\partial^2 \mathcal{W}^{ss}}{\partial^2 \tau_a} = & -\frac{1}{(\theta - 1 + \tau_a)^2} - (v_\varphi + a v_\omega) \\
& - \left(\frac{1+\chi}{\theta-1}\right)\left(\frac{\psi}{1+\psi}\right)\frac{(\delta\beta^2)^a}{(1-\bar{\tau})^2} \\
& - \left(\frac{1+\chi}{1+\sigma}\right)\left(\frac{N_a}{\bar{N}}\right)\cdot\left[\frac{1}{(1-\tau_a)^2} + (\sigma+1)^3\frac{\sigma - 2\tau_a}{(\sigma+\tau_a)^4} v_{\varepsilon a}\right] \\
& + \left(\frac{1+\chi}{1+\sigma}\right)\left(\frac{1}{1-\tau_a} + \left(\frac{1+\sigma}{\sigma+\tau_a}\right)^3 \tau_a v_{\varepsilon a}\right)\frac{1}{\bar{N}}\left[1 - \left(\frac{N_a}{\bar{N}}\right)\frac{1}{A}\right]\frac{\partial N_a}{\partial \tau_a}.
\end{aligned}$$

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 equation (25). 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:

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

where the expressions for $N(a,\tau_a)$ and $\bar{N}(\{\tau_a\})$ 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, $\tilde{Y}(a)$ is age-invariant and hence, from the FOC $(A13)$ also the optimal λ_a^* must 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 av_ω 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,

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

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 $\frac{1}{\theta-1+\tau_a}$, $(1-\tau_a)(v_\varphi + av_\omega)$, 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 $N(a)/\bar{N}$ is increasing in age in the last term of (A16). Thus a lower value of $(1-\tau_a)^{-1}$ is needed to counterbalance this foce 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

$$\begin{aligned} h(a) &= \exp(-\varphi_a)(1-\tau_a)^{\frac{1}{1+\sigma}}, && \text{(A17)} \\ w(a)h(a) &= N_a(\tau_a) = \exp(x_a-\varphi_a)(1-\tau_a)^{\frac{1}{1+\sigma}}. && \text{(A18)} \end{aligned}$$

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^* s.t.

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

which implies

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

The intra-temporal FOC at age a is

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

and since $w(a)h(a) = N_a(\tau_a)$, the labor wedge in this intra-temporal FOC is

$$\begin{aligned} LW_a &= \lambda_a(1-\tau_a)N_a(\tau_a)^{-\tau_a} \\ &= \frac{C}{N_a(\tau_a)}(1-\tau_a) \end{aligned} \tag{A21}$$

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 equation $(A21)$, 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

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