Optimal Progressivity with Age-Dependent Taxation*

Jonathan Heathcote Kjetil Storesletten Giovanni L. Violante

August, 2017

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 a tractable equilibrium overlapping-generations model that incorporates a number of salient trade-offs in tax design. Tax progressivity provides insurance against ex-ante heterogeneity and earnings uncertainty that missing markets fail to deliver. However, taxes distort labor supply and human capital investments. Uninsurable risk cumulates over the life cycle, and thus the welfare gains from income compression via progressive taxation increase with age. On the other hand, average labor productivity rises with age, and thus the welfare losses from progressive taxation's distortionary impact on labor supply also increase with age. The optimal age-varying system balances these distortions. In a calibrated version of the economy, we quantify the welfare gains of moving from the optimal age-invariant to the optimal age-dependent system and find that they are negligible.

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

Keywords: Tax Progressivity, Tagging, Income Distribution, Skill Investment, Labor Supply, Partial Insurance, 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, CEPR, IFS, IZA, and NBER, e-mail glv2@princeton.edu. Prepared for the conference "New Perspectives on Consumption Measures." We thank Guy Laroque and Magne Mogstad for comments on an earlier version of the paper. 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 $T(y) = y - \lambda y^{1-\tau}$, 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 Bénabou 2000, 2002). Moreover, in Heathcote, Storesletten, and Violante (2017), we document that this parametric class provides a remarkably good approximation to the current U.S. system.

The key innovation in the present paper is to let the parameters τ and λ 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 slope of the tax schedule can be made age-dependent.

The 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 investment 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. During working life, individuals face permanent shocks to wages that cannot be insured privately. Thus, these wage shocks pass through to consumption, increasing inequality ex post.

Tax progressivity compresses ex post dispersion in consumption. Thus, the social insurance embedded in the tax and transfer system stands in for some of the demand for insurance that markets fail to deliver because of the assumed market incompleteness. In addition, net tax revenue allows the government to provide the public good. However, tax progressivity discourages labor supply and skill investments. Skills are imperfect substitutes, and the price of skills is an equilibrium outcome. Since the tax system determines the skill distribution, it influences pre-tax skill prices as well as after-tax returns.

In this environment, we provide a closed-form solution for an equally weighted social welfare function. We then use this function to derive a number of analytical results on optimal taxation over the life cycle. The shape of the optimal age profile for the tax progressivity parameter τ trades off two key forces.

First, the model incorporates the standard argument in favor of tagging, namely that age is informative about *average productivity* since wage rates are increasing during the first decades of working life, peaking at around age 50. Because tax progressivity discourages work effort, a rising life-cycle profile of wages is a force for tax *progressivity to fall with age.*

The second motive for age variation in taxes is that 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. Therefore, the planner has an incentive to target redistribution to where inequality is concentrated, namely among the old. This is a force for *progressivity to increase with age.*

In addition, we show that, when the planner maximizes welfare in the final steady state, progressive taxation later in the life cycle is less distortionary for skill investment because individuals discount future taxes when choosing their optimal amount of human capital at young ages. This discounting channel, however, is much weakened when the full transitional dynamics are taken into account.

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 and the welfare gains of switching from the optimal age-invariant progressivity level to optimal age-dependent progressivity. On their own, life-cycle variation in productivity, uninsurable risk, and discounting call for significant variation in tax progressivity over the life cycle, with correspondingly sizable welfare gains. However, when all factors are combined and the transitional dynamics are taken into account, the effects largely neutralize each other, so that the optimal tax system is mildly U-shaped in age. The welfare gain from allowing tax progressivity to vary with age – relative to the optimal age-invariant system – amounts to only 0.08% of lifetime consumption.

We are not the first to study motives for age dependence in the optimal design of tax schedules. Two antecedents of ours follow the Ramsey tradition. Erosa and Gervais (2002) analyze optimal taxation in a setting without any source 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.

A more recent literature studies the role of age variation in the Mirrlees optimal taxation framework. Two papers are especially related to our work. The first paper is by Weinzierl (2011), who focuses on our first channel, the rising age profile of wages and on how these profiles differ across skill groups. His key findings, namely that the average and marginal tax rate are both rising with age, are qualitatively similar to ours when the only operational channel is life-cycle productivity. The second related paper is from 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 and age variation in the life-cycle profile of efficiency units. In their model, the fully optimal history-dependent tax schedule displays the same qualitative features as our model does when our second channel (uninsurable risk) 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.

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 productivity channel is quantitatively most important in the first half of the working life, 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 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 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 5 studies the quantitative implications of allowing for age variation in taxes and quantifies the welfare gain of introducing such fiscal tools. Section 6 concludes.

2 Economic Environment

Demographics: We adopt the Yaari "perpetual youth" structure. At every age a, an agent survives into the next period with constant probability $\delta < 1$. Each period a cohort of newborn agents of size $(1 - \delta)$ enters the economy. There are no intergener-

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

ational links. We index agents by $i \in [0, 1]$.

Life cycle: Upon birth, individuals have a chance to invest in skills. 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

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

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

$$v_i(s_i) = \frac{(\kappa_i)^{-1/\psi}}{1 + 1/\psi} (s_i)^{1+1/\psi}, \tag{2}$$

where the parameter $\psi \geq 0$ determines the elasticity of skill investment with respect to the 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 specified as

$$u_i\,(c_{ia}, h_{ia}, G) = \log c_{ia} - \frac{\exp\left[(1+\sigma)\,\varphi_i\right]}{1+\sigma} (h_{ia})^{1+\sigma} + \chi \log G, \tag{3}$$

where $\exp\left[(1+\sigma)\,\varphi_i\right]$ measures the disutility of work effort. The individual-specific parameter φ_i is normally distributed: $\varphi_i \sim N\left(\frac{v_\varphi}{2}, v_\varphi\right)$, where v_φ denotes the cross-

[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.

[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.

sectional variance.[4] 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 $s \in [0, \infty)$,

$$Y = \left(\int_0^\infty \left[N(s) \cdot m(s) \right]^{\frac{\theta-1}{\theta}} ds \right)^{\frac{\theta}{\theta-1}}, \tag{4}$$

where $\theta > 1$ is the elasticity of substitution across skill types, $N(s)$ 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 in the context of imperfect substitutability across different skill types.

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

$$\log z_{ia} = x_a + \alpha_{ia}. \tag{5}$$

The first component x_a captures the deterministic age profile of labor productivity. The second component α_{ia} follows the unit root process $\alpha_{ia} = \alpha_{i,a-1} + \omega_{ia}$, with i.i.d. innovation $\omega_{ia} \sim N\left(-\frac{v_\omega}{2}, v_\omega\right)$ and with initial condition $\alpha_{i0} = 0$.[5] A standard law of large numbers ensures that individual-level shocks induce no aggregate uncertainty in the economy. In previous work, we also included transitory insurable shocks to labor efficiency. We abstract from those here because they do not play an important role in shaping the optimal age profile of tax progressivity.

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

$$y_{ia} = \underbrace{p(s_i)}_{\text{skill price}} \times \underbrace{\exp(x_a)}_{\text{age-productivity profile}} \times \underbrace{\exp(\alpha_{ia})}_{\text{labor market shocks}} \times \underbrace{h_{ia}}_{\text{hours}}. \tag{6}$$

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 ef-

[4] Introducing additional weighting parameters (common across all households) on the utility terms defining the costs of skill investment and labor supply would have no impact on the shape of the welfare-maximizing policy.

[5] There is still earnings inequality among newborn agents, reflecting heterogeneous skill levels.

ficiency; 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 grow 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 taste for leisure, measured by φ_i.

Because idiosyncratic productivity shocks are exogenous, the two channels via which taxation will affect the equilibrium pre-tax earnings distribution are by changing skill investment choices, and thus skill prices, and by changing labor supply decisions.

Financial assets: No financial assets are traded. 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 x_a and in the tax parameters τ_a and λ_a introduces motives for intertemporal borrowing and lending. Exploring such an extension numerically is an interesting avenue for future research.

Markets: The final consumption good and all types of labor services are traded in competitive markets. The public good G can only be provided by the government. The final good is the numeraire of the economy.

Government: The government runs the tax and transfer scheme and provides each household with an amount of goods or services equal to G. 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

$$T_a(y) = y - \lambda_a y^{1-\tau_a}, \tag{7}$$

where the parameters τ_a and λ_a are specific to age group a. This specification, with age-invariant parameters, has a tradition in public finance (Feldstein 1969; Persson 1983; Bénabou 2000 and 2002; Heathcote et al. 2014 and 2017).

The parameter τ_a determines the degree of progressivity of the tax system and is the key object of interest in our analysis. We can see why τ_a is a natural index of progressivity in two ways. First, eq. (7) implies the following mapping between disposable (post-government) earnings $\tilde{y}_i$ and pre-government earnings y_i :

$$\tilde{y}_i = \lambda_a y_i^{1-\tau_a}. \tag{8}$$

Thus, $(1-\tau_a)$ measures the elasticity of post-tax 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_i. Within our class, we have

$$\frac{1-T_a'(y_i)}{1-T_a(y_i)/y_i} = 1-\tau_a. \tag{9}$$

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, 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. 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.

In Heathcote et al. (2017) we document that this functional form provides a remarkably good representation of the actual tax and transfer scheme in the United States.[6] In particular, eq. (8) 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 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 linear.[7]

[6]Heathcote and Tsujiyama (2016) show that this functional form closely approximates the fully optimal Mirrleesian policy.

[7]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

We abstract from the possibility that the government can issue debt or save. Moreover, all individuals receive the same amount of publicly provided goods. The government budget constraint therefore holds period by period and reads as

$$g\left(1-\delta\right)\sum_{a=0}^{\infty}\delta^{a}\int y_{ia}di=\left(1-\delta\right)\sum_{a=0}^{\infty}\delta^{a}\int\left[y_{ia}-\lambda_{a}\left(y_{ia}\right)^{1-\tau_{a}}\right]di. \qquad (10)$$

The government chooses the sequence $\{\tau_a, \lambda_a\}_{a=0}^{\infty}$ and g, with one instrument being determined residually by eq. (10). For notational convenience, we denote the vectors of τ_a and λ_a by boldface notation, $\boldsymbol{\tau} \equiv \{\tau_a\}_{a=0}^{\infty}$ and $\boldsymbol{\lambda} \equiv \{\lambda_a\}_{a=0}^{\infty}$.

The rate of transformation between private and public consumption is one, so the aggregate resource constraint for the economy is

$$Y=G+\left(1-\delta\right)\sum_{a=0}^{\infty}\delta^{a}\int_{0}^{1}c_{ia}\;di. \qquad (11)$$

2.1 Agent's problem

At age $a = 0$, the agent chooses a skill level, given her idiosyncratic draw (κ_i, φ_i). Combining eqs. (1) and (2), the first-order necessary and sufficient condition for the skill choice is

$$\frac{\partial v_{i}\left(s_{i}\right)}{\partial s_{i}}=\left(\frac{s_{i}}{\kappa_{i}}\right)^{\frac{1}{\psi}}=\left(1-\beta\delta\right)\mathbb{E}_{0}\sum_{a=0}^{\infty}\left(\beta\delta\right)^{a}\frac{\partial u_{i}\left(c_{ia},h_{ia},G\right)}{\partial s_{i}}. \qquad (12)$$

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, each individual chooses hours h_{ia}, receives wage payments, pays taxes, and devotes after-tax income to consumption expenditures c_{ia}.

---

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.

The period budget constraint is

$$c_{ia} = \lambda_a \left[p\left(s_i\right) \exp\left(x_a + \alpha_{ia}\right) h_{ia} \right]^{1-\tau_a}. \tag{13}$$

Given an initial skill choice (s_i), the problem for an agent is to choose sequences of consumption and hours worked in order to maximize lifetime utility (1) subject to sequences of budget constraints (13), taking as given the process for efficiency units described in eq. (5). 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 is (φ, α, s, a). Note that age is a state variable for two reasons: labor productivity potentially varies with age, and 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 constant tax schedule $\{\tau_a, \lambda_a\}$, which is the focus of our steady-state welfare analysis. However, when we later consider optimal once-and-for-all tax reform incorporating transition from the current system, the economy-wide skill distribution will not be constant. In this case, equilibrium objects will be time varying, and an additional assumption is required to preserve tractability. We return to the transition case in Section 4.1.

Given a government policy $(\boldsymbol{\tau}, \boldsymbol{\lambda})$, a *stationary recursive competitive equilibrium* for our economy is a public good provision parameter g, skill prices $p(s)$, decision rules $s(\kappa, \varphi)$, $c(\varphi, \alpha, s, a)$, and $h(\varphi, \alpha, s, a)$, effective hours by skill $N(s)$, and a skill density $m(s)$ such that:

1. Households solve the problem described in Section 2.1, and $s(\kappa, \varphi)$, $c(\varphi, \alpha, s, a)$, and $h(\varphi, \alpha, s, a)$ 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:

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

3. The government budget is balanced: g satisfies eq. (10).

Propositions 1 and 2 describe the equilibrium allocations and skill prices in closed form. The payoff from analytical tractability is evident in Proposition 3, where we derive a set of analytical 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 $(\boldsymbol{\tau}, \boldsymbol{\lambda})$ in preparation for our analysis of the optimal taxation problem. Moreover, from now on we express – with some abuse of notation – the arguments in the decision rules using the minimum set of relevant state variables and policies.

Proposition 1 [hours and consumption]. *The equilibrium hours-worked allocation is given by*

$$\log h\left(\varphi; \tau_a\right) = -\varphi + \frac{\log(1-\tau_a)}{\left(1+\hat{\sigma}_a\right)\left(1-\tau_a\right)}, \tag{14}$$

where $1/\widehat{\sigma}_a$ denotes the tax-modified Frisch elasticity,

$$\frac{1}{\hat{\sigma}_a} = \frac{1-\tau_a}{\sigma+\tau_a}. \tag{15}$$

The consumption allocation is given by

$$\log c\left(\varphi, \alpha, s, a; \boldsymbol{\tau}, \lambda_a\right) = \left(1-\tau_a\right)\left[\log p\left(s; \boldsymbol{\tau}\right) + \alpha - \varphi + x_a\right] + \frac{\log(1-\tau_a)}{1+\hat{\sigma}_a} + \log \lambda_a. \tag{16}$$

With logarithmic utility and zero individual wealth, the income and substitution effects on labor supply from differences in uninsurable shocks α, skill levels s, and experience x_a exactly offset, and hours worked are independent of (s, α) and depend on age only through τ_a. The hours allocation is composed of two terms. The first captures the fact that a higher idiosyncratic disutility of work leads an agent to choose lower hours. The second 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 consumption allocation is additive in six separate components. Consumption is increasing in the skill level s (because the skill price $p\left(s; \boldsymbol{\tau}\right)$ 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 φ, 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 wage shocks α and efficiency units x_a. The fifth and sixth components are what consumption would be in the absence of within-age heterogeneity.

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

$$\log p\,(s;\bar{\tau}) = \pi_0\,(\bar{\tau}) + \pi_1\,(\bar{\tau}) \cdot s\,(\kappa;\bar{\tau})\,, \tag{17}$$

where $\bar{\tau}$ *is discounted average progressivity,* $\bar{\tau} = (1-\beta\delta)\sum_{a=0}^{\infty}(\beta\delta)^a\,\tau_a$*, and the functions* π_0 *and* π_1 *are given by*

$$\pi_1\,(\bar{\tau}) = \left(\frac{\eta}{\theta}\right)^{\frac{1}{1+\psi}}(1-\bar{\tau})^{-\frac{\psi}{1+\psi}} \tag{18}$$

$$\pi_0(\bar{\tau}) = \frac{1}{\theta-1}\left\{\frac{1}{1+\psi}\left[\psi\log\left(\frac{1-\bar{\tau}}{\theta}\right)-\log(\eta)\right]+\log\left(\frac{\theta}{\theta-1}\right)\right\}. \tag{19}$$

Moreover, the skill investment allocation is given by

$$s\,(\kappa;\bar{\tau}) = \left[(1-\bar{\tau})\,\pi_1\,(\bar{\tau})\right]^{\psi}\cdot\kappa = \left[\frac{\eta}{\theta}\,(1-\bar{\tau})\right]^{\frac{\psi}{1+\psi}}\cdot\kappa, \tag{20}$$

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

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. (18) 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 $(\boldsymbol{\tau},\boldsymbol{\lambda})$ 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 φ and α_0. The logic is that, with log utility, the welfare gain from additional skill investment is proportional to the log change in wages the investment would induce, which is independent of the *level* of wages or hours.

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*

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

The distribution of skill prices $p(s; \bar{\tau})$ 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. 17) and skills retain the exponential shape of the distribution of learning ability κ (eq. 20). It is interesting that inequality in skill prices is independent of the policy sequence $\{\lambda_a, \tau_a\}$. 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 (the Stiglitz effect on prices). On the other hand, higher progressivity compresses the distribution of skills (the quantity effect). These two forces exactly cancel out under our baseline 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. Because the Pareto component dominates at the top, the equilibrium distributions of wages and earnings have Pareto right tails, a robust feature of their empirical counterparts (see, e.g., Atkinson, Piketty, and Saez 2011). We now briefly discuss how taxation affects aggregate quantities in our model.

Corollary 2.2 [aggregate quantities]. *Average hours worked and average effective hours are independent of skill type s and given by $H(\boldsymbol{\tau}) = (1-\delta)\sum_{a=0}^{\infty}\delta^a H(\tau_a)$ and $N(\boldsymbol{\tau}) = (1-\delta)\sum_{a=0}^{\infty}\delta^a N_a(\tau_a)$, where*

$$H(\tau_a) = \mathbb{E}\left[h(\varphi; \tau_a)\right] = (1-\tau_a)^{\frac{1}{1+\sigma}}, \tag{21}$$

$$N_a(\tau_a) = \mathbb{E}\left[\exp(x_a + \alpha) h(\varphi; \tau_a)\right] = (1-\tau_a)^{\frac{1}{1+\sigma}} \cdot \exp(x_a). \tag{22}$$

Output is given by

$$Y(\boldsymbol{\tau}) = \mathbb{E}\left[p(s;\bar{\tau})(1-\delta)\sum_{a=0}^{\infty}\delta^a N_a(\tau_a)\right] = N(\boldsymbol{\tau})\cdot\mathbb{E}\left[p(s;\bar{\tau})\right], \tag{23}$$

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

4 Optimal Age-Dependent Taxes: Characterization

We start by analyzing the optimal policy in steady state. This approach allows us to derive a number of analytical results for optimal taxation. Moreover, it also allows us to abstract, for the time being, from a standard issue inherent in models with sunk human capital investments: since past investment decisions are irreversible, the government would, in the short run, be tempted to heavily tax high skill individuals because such taxation is not distortionary ex post. This result is analogous to the temptation to tax initial physical capital in the growth model. In Section 4.1 we analyze the full transition and solve numerically for the optimal progressivity profile $\boldsymbol{\tau}$, assuming an initial once-and-for-all tax reform.[8]

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 for remaining expected lifetime utility, where eq. (1) defines 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 as individuals discount utility over the life cycle, β. Social welfare evaluated as of date 0 in a steady state associated with a policy $(\boldsymbol{\tau},\boldsymbol{\lambda})$ —recall that, given this policy, g is determined residually by the government budget constraint (10)— is

$$\mathcal{W}(\boldsymbol{\tau},\boldsymbol{\lambda}) \equiv (1-\beta)\Gamma\sum_{j=-\infty}^{\infty}\beta^j U_{j,0}(\boldsymbol{\tau},\boldsymbol{\lambda}), \tag{24}$$

[8] 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. However, irreversible skill investment seems to be the more realistic case.

where $U_{j,0}(\boldsymbol{\tau},\boldsymbol{\lambda})$ is remaining expected lifetime utility (discounted back to date of birth) as of date 0 for the cohort that entered the economy at date j.[9] The constant $\Gamma=(1-\delta)/(1-\beta\delta)$ pre-multiplying the summation is a convenient normalization.

It is straightforward to show that steady-state welfare $\mathcal{W}(\boldsymbol{\tau},\boldsymbol{\lambda})$ is equal (up to an additive constant) to

$$\begin{aligned}\mathcal{W}(\boldsymbol{\tau},\boldsymbol{\lambda}) = \quad & (1-\delta)\sum_{a=0}^{\infty}\delta^{a}\mathbb{E}\left[u\left(c\left(\varphi,\alpha,s,a;\boldsymbol{\tau},\lambda_a\right),h\left(\varphi;\tau_a\right),G\left(\boldsymbol{\tau},\boldsymbol{\lambda}\right)\right)\right] \\ & -\left(\frac{1-\delta}{1-\beta\delta}\right)\mathbb{E}\left[v\left(s(\kappa;\bar{\tau}),\kappa\right)\right], \end{aligned} \tag{25}$$

where the first expectation is taken with respect to the equilibrium cross-sectional distribution of (φ,α,s,a) conditional on a, the second expectation is with respect to the cross-sectional distribution of (s,κ), and $G(\boldsymbol{\tau},\boldsymbol{\lambda})$ denotes the budget-balancing value for G. The weight on the average skill investment cost differs from that on the other terms because the steady-state welfare expression does not attribute any investment costs to cohorts who made irreversible skill investments in the past. Thus, the steady-state welfare calculation factors in the benefits of past skill investments but not their cost. The policy that attains the highest steady-state welfare $(\boldsymbol{\tau}^{*},\boldsymbol{\lambda}^{*})$ is simply the policy that maximizes eq. (25).

The proof of Proposition 3 establishes that the social welfare function $\mathcal{W}$ is differentiable and globally concave in τ_a. It follows that the first-order condition $\partial\mathcal{W}/\partial\tau_a=0$ is necessary and sufficient. This optimality condition can be stated analytically as

$$\begin{aligned} 0 \quad = \quad & \frac{1}{\theta-1+\tau_a}-\frac{1}{\theta}+\left(1-\tau_a\right)\left(v_{\varphi}+av_{\omega}\right)+\frac{1}{1+\sigma}+ \\ & -\left(\frac{1+\chi}{\theta-1}\frac{1-\beta\delta}{1-\delta}\frac{1}{1-\bar{\tau}}-\frac{1}{\theta}\right)\frac{\psi}{1+\psi}\left(\beta\right)^{a} \\ & -\frac{1+\chi}{1+\sigma}\frac{1}{1-\tau_a}\frac{N_a\left(\tau_a\right)}{N\left(\boldsymbol{\tau}\right)}, \end{aligned} \tag{26}$$

where the expressions for effective age a labor supply and aggregate labor supply, $N_a(\tau_a)$ and $N(\boldsymbol{\tau})$, are given in Corollary 2.2. Using this optimality condition, we can prove a number of analytical results about the optimal age-dependent taxation, which we summarize in the following proposition.

[9] This weighting scheme implies that all agents alive at the time of the reform receive equal weight (one) on their residual expected utility from that date onward.

Proposition 3 [optimal age-dependent taxation in steady state].
(i) The optimal output share of government expenditures g^* *is given by*

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

(ii) If (i) there is no uninsurable risk ($v_\omega = 0$)*, (ii) the age profile of efficiency units* $\{x_a\}$ *is constant, and (iii)* $\beta \to 1$*, then the optimal sequences* $\{\tau_a^*\}$ *and* $\{\lambda_a^*\}$ *are age-invariant.*

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

(iv) Relative to the parameterization described in (ii), introducing an age profile for efficiency units $\{x_a\}$ *that is increasing with age translates into optimal profiles* $\{\tau_a^*\}$ *and* $\{\lambda_a^*\}$ *that are decreasing in age.*

(v) Relative to the parameterization described in (ii), setting $\beta < 1$ *translates into optimal profiles* $\{\tau_a^*\}$ *and* $\{\lambda_a^*\}$ *that are increasing in age.*

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

We now explain the results of Proposition 3, one by one.

(i) In our economy, the optimal fraction of output to devote to public expenditure is independent of how much inequality there is in the economy and independent of how progressive the tax system is. It depends only on households' relative taste for the public good χ. 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.

(ii) Consider now the forces shaping the optimal age profiles for λ_a and τ_a. First, absent age variation in both productivity and inequality, the only possible motive for age variation in progressivity is the discounting logic described previously: if $\beta < 1$, skill investment is more sensitive to higher progressivity earlier in life and less sensitive later. As $\beta \to 1$, this motive for having progressivity rise with age vanishes.

(iii) Now consider the role of uninsurable risk. To isolate this force, we focus on the case with a constant age-wage profile ($x_a = 0$) and no human capital investment ($\theta \to \infty$), thereby shutting down alternative motives for age variation in progressivity. The social welfare first-order condition (26) then simplifies to

$$0 = (1-\tau_a)(v_\varphi + a v_\omega) + \frac{1}{1+\sigma}\left(1 - \frac{(1+\chi)(1-\tau_a)^{-\frac{\sigma}{1+\sigma}}}{(1-\delta)\sum_{t=0}^{\infty}\delta^t (1-\tau_t)^{\frac{1}{1+\sigma}}}\right).$$

When $v_\omega > 0$, the first term is increasing in age a, and to satisfy the first-order condition, the optimal τ_a^* must therefore be rising in age. 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 the *uninsurable risk channel.*

Another way to interpret the uninsurable risk channel relates to the roles of exogenous risk and endogenous skill accumulation in the determination of wage inequality. Most of the cross-sectional model variation in wages at younger ages reflects skill investment choices, with respect to which tax progressivity is distortionary. As individuals age, a larger share of wage inequality reflects uninsurable labor market luck, which is exogenous and unaffected by tax progressivity.

(iv) Now consider the role of the age profile of efficiency units $\{x_a\}$. To isolate the impact of this model ingredient, we strip out risk ($v_\omega = 0$), preference heterogeneity ($v_\varphi = 0$), and human capital investment ($\theta \to \infty$). The optimal value for τ at age a, τ_a^*, is then the solution to the following simplified version of the first-order condition (26):

$$\begin{aligned} 1 - \tau_a^* &= \frac{1}{1+\chi} \frac{N_a(\tau_a)}{N(\boldsymbol{\tau})} \\ &= \left(\frac{\exp(x_a)}{(1+\chi)(1-\delta)\sum_{t=0}^{\infty} \delta^t (1-\tau_t)^{\frac{1}{1+\sigma}} \cdot \exp(x_t)} \right)^{\frac{1+\sigma}{\sigma}}. \end{aligned}$$

This illustrates that *ceteris paribus* the optimal τ_a^* is lower the larger is x_a.[10] Moreover, this effect is stronger the higher is the Frisch elasticity (i.e., the lower is σ). The intuition is that absent age variation in τ, hours worked will be independent of productivity given our utility function and tax system. The planner can increase aggregate labor productivity and thus welfare by having more productive agents work relatively longer hours. When the labor productivity profile is upward sloping, this introduces a force for having progressivity decline with age. We label this the *life-cycle productivity channel.*

Note that the life-cycle productivity channel would be weaker if we introduced opportunities for intertemporal borrowing and lending. In particular, if households

[10]This follows from the fact that an increase in x_a increases the effective labor supply of cohort a, N_a, more than it increases aggregate labor supply N, since the labor supply of cohort a accounts for less than 100% of aggregate effective labor supply. Formally, the elasticity of N_a/N is always positive with respect to x_a; $\partial \{\log N_a(\tau_a) - \log N(\boldsymbol{\tau})\} / \partial x_a = 1 - (1-\delta)\delta^a N_a(\tau_a^*)/N(\boldsymbol{\tau}^*) > 0$.

could borrow and lend freely, then hours would tend to naturally covary positively with productivity over the life cycle, shutting down this motive for age variation in tax progressivity.[11] In practice, it is hard to borrow against future income, especially for young workers for whom life-cycle productivity is rising fast and for whom this channel is therefore strongest.

(v) When maximizing steady-state welfare, the presence of discounting ($\beta < 1$) is another force toward making the optimal τ_a^* increase with age. The reason is that raising τ_a at younger ages reduces skill investment by more than raising it at older ages – because agents discount future returns to skill.[12] To illustrate this, consider the planner problem in the special case when there is no uninsurable risk and no labor supply distortion – thereby eliminating the uninsurable risk and life-cycle productivity channels (i.e., $v_\omega = 0$ and $\sigma \to \infty$). In this case, the first-order condition of the social welfare function with respect to τ_a, (26), simplifies to

$$0 = \underbrace{\frac{1}{\theta - 1 + \tau_a} - \frac{1}{\theta}}_{>0} - \underbrace{\left[\left(\frac{1+\chi}{\theta-1}\right)\left(\frac{1-\beta\delta}{1-\delta}\right)\left(\frac{1}{1-\bar{\tau}}\right) - \frac{1}{\theta}\right]}_{>0}\left(\frac{\psi}{1+\psi}\right)\beta^a. \qquad (27)$$

The first two additive terms capture the welfare gain of increasing τ_a because of the marginal reduction in after-tax skill price inequality. The last term in eq. (27) captures the distortion to skill investment from a marginal increase in τ_a and is equal to $\partial/\partial\tau_a \{\log(\mathbb{E}(p(s))\}$. It is proportional to β^a since skill investments are forward looking. Thus, when $\beta < 1$ the distortion to skill investment is declining in age, implying that the optimal τ_a^* must increase with age. We label this the *discounting channel.* The discounting channel vanishes if $\beta = 1$, since tax progressivity then becomes equally distortive at all ages, so there is no motive for back-loading progressivity.[13] The discounting channel would also vanish if skill investments were fully reversible because in that case, introducing an upward tilt in the progressivity profile would increase disinvestment by the old, offsetting additional investment by the young.

(vi) Finally, since higher progressivity reduces labor supply, an age-varying profile

[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]Note, however, that for the future newborn individuals, who will invest in skills, this motive for back-loading progressivity is not present because the planner and the individuals agree on how to discount. We return to this point when we study the transition in Section 4.1.

[13]The survival probability δ is irrelevant for the age profile of τ_a because this parameter has the same effect on individual discounting as it has on future cohort size. Therefore, the term δ^a multiplies all terms in eq. (27) and cancels out.

$\{\tau_a^*\}$ will generate an age-varying profile for average earnings. So will an age-varying productivity profile. If between age-group earnings inequality translated into consumption inequality, it would be welfare reducing for our utilitarian planner. But by having $\{\lambda_a^*\}$ vary with age, the planner can perfectly smooth average consumption across age groups. This result indicates that the government, through the tax system, can effectively replicate the role of life-cycle borrowing and saving in smoothing predictable life-cycle income variation.

4.1 Transition

Thus far we have focused on steady-state welfare because the tractability of the social welfare expression in steady state allows us to cleanly isolate several key forces that determine optimal age variation in tax progressivity. We now turn to consider optimal taxation when taking explicit account of the transition from an initial state.

For simplicity, we assume that at the time of the reform, the economy is in a steady state associated with a particular age-invariant tax system characterized by tax progressivity τ_{-1} and government spending g_{-1}.

We then consider an unanticipated policy change at date $t = 0$ to a new policy regime $(\{\tau_a\}, \{\lambda_{at}\}, \{G_t\})$ with new age-dependent tax rates. We impose that the reform is a once-and-for-all reform for progressivity, in the sense that the new age profile of progressivity $\{\tau_a\}$ is constant over time.[14] However, we do allow the profile of the proportional factors $\{\lambda_{at}\}$ and the public good consumption G_t to vary over time, maintaining the assumption that the government budget must be balanced each period.

To preserve tractability, we make one minor new assumption relative to the baseline model described previously, namely that production is segregated on two islands: one island for all the cohorts born before the tax reform at date zero (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 on the first island declines, and eventually the entire population resides on the second island. Each island is otherwise similar to the economy laid out in Section 2, including the island-specific production function in eq. (4). This segregation assumption ensures that the distribution of skills on each island is always exponential.[15]

[14] 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.

[15] Note that the key to tractability when analyzing the market for skills is that the distribution

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. (14)-(16), 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 choices for the old are sunk and were determined by the tax progressivity before the reform, τ_{-1}. Thus, their skill prices $p\left(s;\tau_{-1}\right)$ are exactly as in eq. (17) of Proposition 2. In contrast, the skill investment and skill prices for the young, $s\left(\kappa;\bar{\tau}\right)$ and $p(s;\bar{\tau})$, depend on $\bar{\tau}$, which is the discounted progressivity of the new tax progressivity profile $\{\tau_a^*\}$.

How does incorporating transition change the optimal policy prescriptions? First, in two special cases the expression for social welfare and thus optimal policies are identical to the steady-state case considered above. 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.

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.

In both of these special cases, the steady state prescriptions in Proposition 3 all carry through unchanged to the transition experiment. In particular, the uninsurable risk channel and the life-cycle productivity channel both remain intact, pushing respectively for upward- and downward-sloping age profiles for tax progressivity. However, both special cases kill the discounting channel by shutting down discounting and skill investment, respectively.

We now consider the general case in which $\beta < 1$ and $\theta < \infty$. In this case, incorporating transition adds a new driver shaping the optimal age profile of progressivity, which we label the *sunk skill investment channel.*

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 ensures that the distribution of skills remains exponential *at each location.*

Corollary 4 [taxation with transition]. *When taking the initial 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) If (i) there is no uninsurable risk ($v_\omega = 0$), (ii) the age profile of efficiency units $\{x_a\}$ is constant, and (iii) $\beta < 1$, then the optimal sequence $\{\tau_a^*\}$ is increasing in age. This optimal profile implies a higher value for $\bar{\tau}$ than the profile than maximizes steady-state welfare.

Part (i) of Corollary 4 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) of Corollary 4 shows that the optimal time-varying sequence $\{\lambda_{at}^*\}$ simply ensures that average consumption is equated across age groups at each point in time. Thus, this property of the policy that maximizes steady-state welfare again extends to the case that incorporates transition.

How does incorporating transition change the optimal age profile for progressivity? First, incorporating transition does not fundamentally change how the uninsurable risk and life-cycle productivity channels affect age dependence in progressivity. To most sharply illustrate the impact of including transition, we therefore assume no uninsurable risk and no life-cycle variation in average productivity when comparing the progressivity profiles that maximize the two alternative welfare objectives (steady-state welfare and welfare including transition). Figure 1 illustrates the result stated in part (iii) of Corollary 4.

The policies that maximize these two different objectives exhibit two differences. First, relative to the policy that maximizes steady-state welfare, incorporating transition implies higher progressivity on average, and in particular a higher value for $\bar{\tau}$. The logic is that the planner who gets to reset taxes in an environment with irreversible skill investment recognizes that he can impose high progressivity on existing cohorts without affecting their sunk skill investments, thereby reducing consumption inequality without distorting investment for existing cohorts. This temptation is declining in β and is not present in the limit $\beta \to 1$. This force is familiar from the literature that

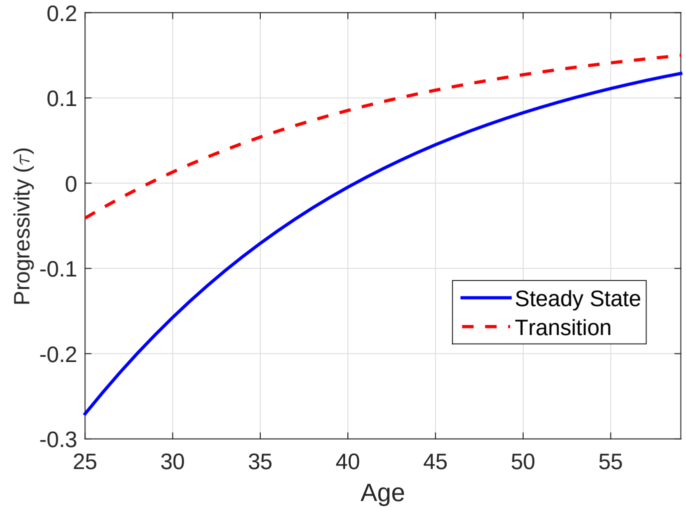

Figure 1: Optimal age dependence of taxes with no uninsurable risk ($v_\omega = 0$), a flat age profile of wages ($x_a = 0$), and discounting ($\beta < 1$) given a steady-state welfare objective (solid line) and a welfare objective that incorporates transition (dashed line). The exact parameterization for this example is the one displayed in Table 1.

studies the optimal taxation of capital (see, e.g., Chari and Kehoe 1999 for a survey; and Hassler, Krusell, Storesletten, and Zilibotti 2008 for optimal Ramsey taxation in a human capital setting with overlapping generations).

What about the slope of the progressivity age profile? Absent age variation in productivity or uninsurable risk, the optimal age profile for tax progressivity is upward sloping given $\beta < 1$, mirroring the result in part (vi) of Proposition 3. However, the optimal age profile incorporating transition is not as steep as the one that maximizes steady-state welfare. This result arises because the optimal age profile here balances two considerations. On the one hand, an upward-sloping age profile delivers short-run benefits, in that high progressivity at older ages delivers valuable consumption compression to the existing old, at a relatively small cost in terms of disincentivizing skill investment by new entrants, because of the discounting channel described previously. This short-run consideration suggests an upward-sloping progressivity profile. On the other hand, in the long run, the planner cannot use this trick of twisting the progressivity profile to expropriate the fruits of past investment without greatly affecting new investment. To see this, consider what the optimal policy would be for an economy with a single cohort. Here, there would be no reason to deviate from a flat progressivity profile, because while higher old age values for τ_a would be less costly in terms of reduced investment – the discounting channel – they would be equally less valuable –

β	δ	χ	σ	v_φ	θ	v_ω	ψ	τ_{US}
0.950	0.971	0.233	2.00	0.036	3.124	0.003	0.65	0.181

Table 1: Parameterization derived from Heathcote et al. (2017).

again because of discounting – in terms of future consumption compression. Thus, the optimal once-and-for-all profile for τ_a is a compromise between the impulse to make τ increase with age in the short run and the impulse to make it flat in the long run. Quantitatively, how these effects balance out depends on β. As $\beta \to 1$, the discounting channel vanishes, and the optimal progressivity profile becomes flat at the value that maximizes steady-state welfare.

5 Quantitative Application

In this section, we describe the model parameterization and explore the quantitative implications of the theory. We begin with the problem of the planner that maximizes steady-state welfare, as in Section 4. Next, we solve for the optimal age-dependent tax system that takes into account the full transitional dynamics.

5.1 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. The discount factor β is set to 0.95 and the survival rate δ to 0.971, corresponding to an expected working life of 35 years. The weight on public good χ in preferences is identified from the size of the U.S. government, assuming that the choice of public good provision in the data is optimal.[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, instead, structurally estimated from the model. Namely, we identify and estimate preference heterogeneity (v_φ), the elasticity of substitution between skills (θ), and the variance of uninsurable wage risk (v_ω) using cross-sectional variances and covariances of consumption, hours, and wages (measured from the Consumer Expenditure Survey (CEX) and the Panel Study of Income Dynamics (PSID)). The identification follows from the closed-form expressions for consumption, hours, and earnings derived in eqs. (6), (14), and (16). Our estimation procedure allows

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

for classical measurement error in all variables and is based on a minimum distance approach.

We identify the elasticity of human capital investments ψ from a combination of changes over time in the skill premium, in the tax progressivity, and in the upper tail of labor earnings, exploiting the expression for the skill premium in eq. (18).

Finally, the initial tax progressivity τ_{US} is 0.181, as estimated by Heathcote et al. (2017) based on TAXSIM applied to PSID income data, and data on tax deductions and on Social Security contribution and benefits. This parameter value is only relevant when we analyze transitional dynamics.

The only addition relative to the parameterization in Heathcote et al. (2017) is the age profile of wage rates. The life-cycle profile of individual efficiency is estimated on data from the PSID for the years 2000, 2002, 2004, and 2006. This is the same sample as that used for the rest of the calibration. We regress individual log wages on year dummies, years of education, and a quartic in age. This quartic identifies the age profile $\{x_a\}_{a=1}^{35}$ for men aged 25 to 59.

Figure 2 plots the wage-experience profile: note that it is not monotonically increasing with age, a feature that is quantitatively important. Table 1 summarizes the parameter values. We refer the reader to Heathcote et al. (2017) for further details.

We set a maximum model age $A = 35$, corresponding to a real age of 59. One motivation for this choice is that our focus is on the design of a tax and transfer system for working-age people.[17] One could imagine our model individuals living past 60 under a non-modeled post-retirement tax and benefit system.

5.2 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 then we consider the transitional dynamics.

Recall that we identified four different forces that shape the optimal age profile of tax progressivity: uninsurable risk, life-cycle productivity, discounting, and the sunk skill investment channel that is operational only during the transition. To understand the quantitative role of each of these, we start by studying the effect of uninsurable risk and then add the other factors cumulatively, one by one.

For all of our welfare calculations, we report the additional welfare gains of switching

[17] A practical motivation is that it is computationally infeasible to solve for an arbitrarily long sequence τ_a^*.

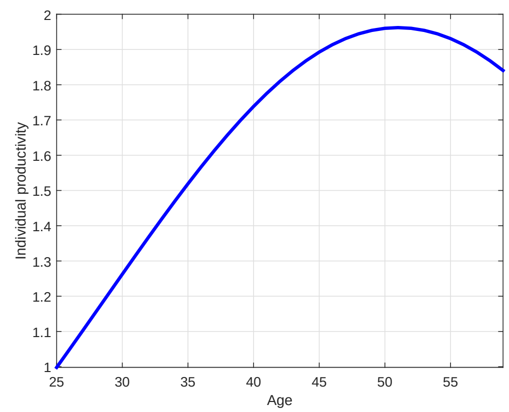

Figure 2: Estimated life-cycle profile of individual productivity.

from the optimal tax system when the progressivity parameter τ is restricted to be constant over the life cycle, as in Heathcote et al. (2017), to the optimal age-dependent tax system. Note, however, that even our age-independent tax system allows for age variation in λ_a. If we were to take as a baseline the case in which λ is also restricted to be age-invariant, the welfare gains from tagging by age would be larger.[18]

5.2.1 Uninsurable risk channel

Part (ii) of Proposition 3 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 isolate this effect, we consider the case with the calibrated amount of uninsurable risk ($v_\omega = 0.007$), but shut down the other channels by assuming a flat age profile for efficiency units ($x_a = 1$ for all a). We initially set $\beta = 1$ so that there is no difference between the policies that maximize welfare including or excluding transition.

The top left panel of Figure 3 plots the optimal sequence $\{\tau_a^*\}$. The increasing magnitude of uninsurable risk over the life cycle induces substantial variation in optimal tax progressivity: for the youngest cohorts, the optimal taxes are regressive ($\tau < 0$), whereas for the oldest cohorts, taxes are even more progressive than the current U.S. system (recall that $\tau_{US} = 0.181$). Moreover, the implied income-weighted average marginal tax rate increases by roughly 14 percentage points over the life cycle. The top right panel illustrates that, even though average pre-government income falls with

[18]The optimal age-invariant value for progressivity varies in the different cases we analyze, but it is always below the existing value estimated from the U.S. data ($\tau_{US} = 0.181$) and sometimes much below.

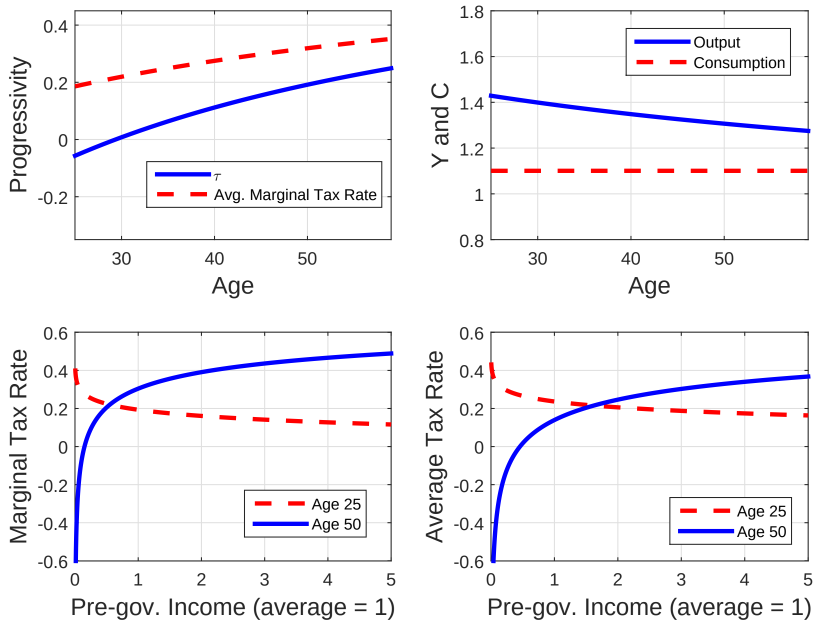

Figure 3: Optimal age dependence of taxes in the presence of uninsurable risk ($v_\omega > 0$), a flat age profile of efficiency units ($x_a = 1$), and no discounting ($\beta = 1$). Top left panel: $\{\tau_a\}$ and income-weighted average marginal tax rate. Top right panel: average income and consumption by age group. Bottom panels: marginal and average tax rate schedules at ages 25 and 50.

age because of the distortions implied by higher progressivity, the optimal tax scheme redistributes across groups so as to equalize average consumption. In fact, it is because of the rising sequence $\{\lambda_a^*\}$ that the average marginal tax rate rises less steeply than τ_a in the top left panel.

Finally, the lower two panels of Figure 3 plot the marginal and average tax rates implied by the age-dependent schedule at ages 25 and 50: the rise in optimal progressivity is stark.

These results are reminiscent of findings in recent quantitative applications of dynamic Mirrleesian optimal taxation, according to which, when income shocks are persistent, the average labor wedge 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 a flat age profile of efficiency units and no endogenous skill accumulation.[19] Thus, the relevant comparison to our model is precisely this case in which uninsurable risk is the only force influencing the age profile of

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

taxation. They find that the optimal history-dependent tax scheme has all the same qualitative features as in our model (see Farhi and Werning 2013, Figure 2). Namely, average output is decreasing in age, consumption is invariant to age, and the labor wedge is increasing in age.[20]

In terms of welfare, we find that moving from the optimal age-invariant tax system, which implies a value of $\tau^* = 0.099$, to the optimal age-dependent tax schedule would induce gains on the order of 0.28% of lifetime consumption.

5.2.2 Life-cycle productivity channel

We now add the second channel for age-varying taxes identified in Section 4. We retain the assumption $\beta = 1$. Tax progressivity should now be lower for the older cohorts because their average wages are higher, so it is more costly for the planner to distort their labor supply.

As suggested by Proposition 3, introducing a rising age profile of efficiency units is a strong force for lowering tax progressivity with age. Now the optimal tax progressivity is U-shaped (see the top left panel of Figure 4). The age profile of τ_a is decreasing until the mid-40s, a phase of the life cycle when productivity is rising steeply. Once the productivity profile flattens out —leaving only the uninsurable risk channel active— progressivity starts increasing with age.

The average marginal tax rate is sharply increasing over the life cycle, from roughly zero for the youngest cohorts to average marginal taxes near 40% for the oldest cohorts (top left panel of Figure 4). The system involves substantial intergenerational redistribution, since earnings are rising over the life cycle while consumption is constant.[21]

The welfare gains of moving from the optimal age-invariant tax system to the optimal age-dependent tax schedule are 0.15% of lifetime consumption, that is, about half as large as the previous case with a flat age-wage profile. The logic is that a constant τ now approximates quite well the mild U shape of optimal age-dependent progressivity.

[20] Golosov, Troshkin, and Tsyvinski (2016) show that with negatively skewed income shocks, the positive drift in the labor wedge is stronger in the left tail of the income distribution.

[21] The implied intergenerational redistribution is influenced by our assumption that there are no possibilities to save over the life cycle. It would be interesting to extend the analysis to allow for life-cycle savings. In this case, one could solve the model numerically and evaluate how the implications for the optimal tax system would change relative to our no-savings benchmark. We leave this for future work.

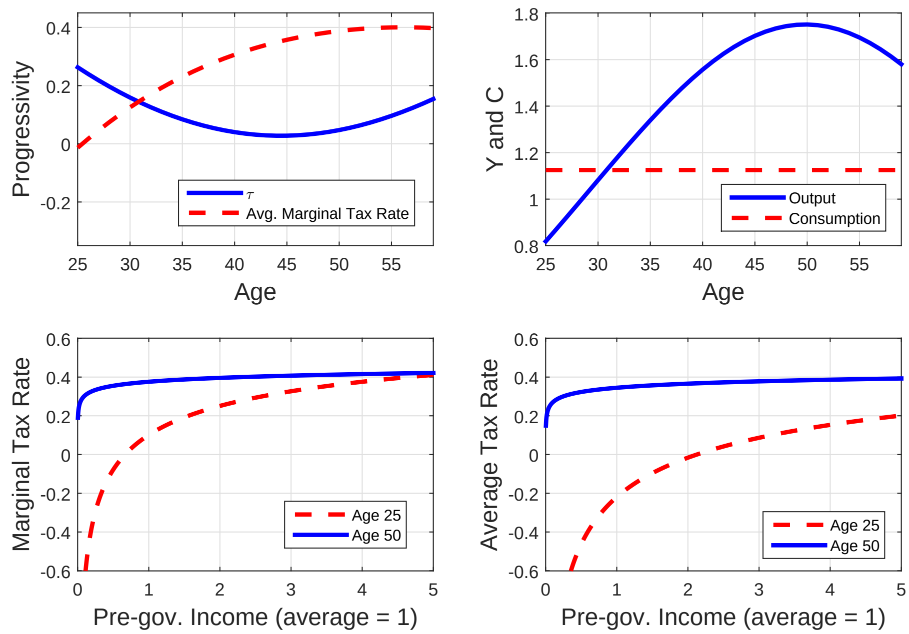

Figure 4: Optimal age dependence of taxes with uninsurable risk ($v_\omega > 0$), the empirical age profile of wages in Figure 2, and no discounting ($\beta = 1$). Top left panel: $\{\tau_a\}$ and income-weighted average marginal tax rate. Top right panel: average income and consumption by age group. Bottom panels: marginal and average tax rate schedules at ages 25 and 50.

5.2.3 Discounting channel

We now turn to the third channel for age-varying taxes identified in Section 4, namely that progressive taxation later in the life cycle is less distortive for skill investment because individuals discount future taxes when choosing their optimal amount of human capital. To understand this effect, we change value of the discount factor from $\beta = 1$ to $\beta = 0.95$. All channels are now operative, so this case should be viewed as our benchmark when focusing on steady-state welfare.

The main change relative to the previous case is that the rise in tax progressivity over the life cycle again becomes 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 (see the top left panel and the lower two panels of Figure 5).

The potential welfare gains from age-dependent taxes, relative to optimal age-invariant taxes, are now 0.25% of lifetime consumption and thus similar to the case with only the uninsurable risk channel.

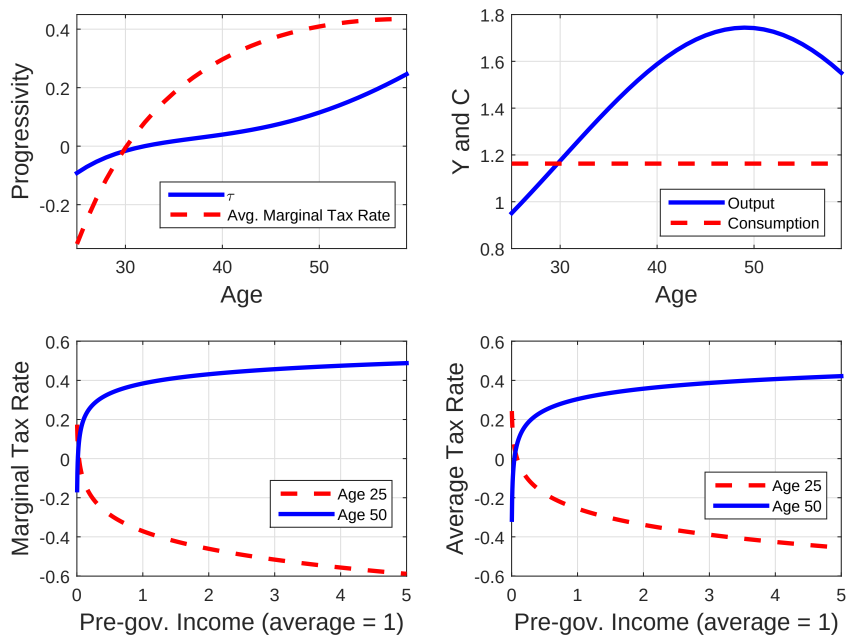

Figure 5: Optimal age dependence of taxes with uninsurable risk ($v_\omega > 0$), the empirical age profile of wages in Figure 2, and discounting ($\beta = 0.95$). Top left panel: $\{\tau_a\}$ and income-weighted average marginal tax rate. Top right panel: average income and consumption by age group. Bottom panels: marginal and average tax rate schedules at ages 25 and 50.

5.2.4 Transitional dynamics and the sunk skill investment channel

We now compute the age-dependent tax system that maximizes welfare taking into account transitional dynamics. All channels are now operative: uninsurable risk, life-cycle productivity, discounting, and sunk investment. The importance of the last depends on the tax system in place in the initial steady state. We assume that this system features the age-invariant value for progressivity $\tau_{US} = 0.181$, the estimated value for the U.S. economy, and $g = g^*$.

The optimal age-dependent tax system is plotted in Figure 6. The plots for the consumption and earnings profiles (upper right panel of Figure 6) refer to the profiles associated with cohorts living in the corresponding post-transition steady state.

A comparison of the top left panel in this figure with that of Figure 5 illustrates two key differences that arise when we incorporate transition in the analysis. First, the transition case involves a higher average level of tax progressivity. Second, the age profile of progressivity is flatter when maximizing welfare along the transition. As discussed previously, this is because the discounting channel is weaker once transition

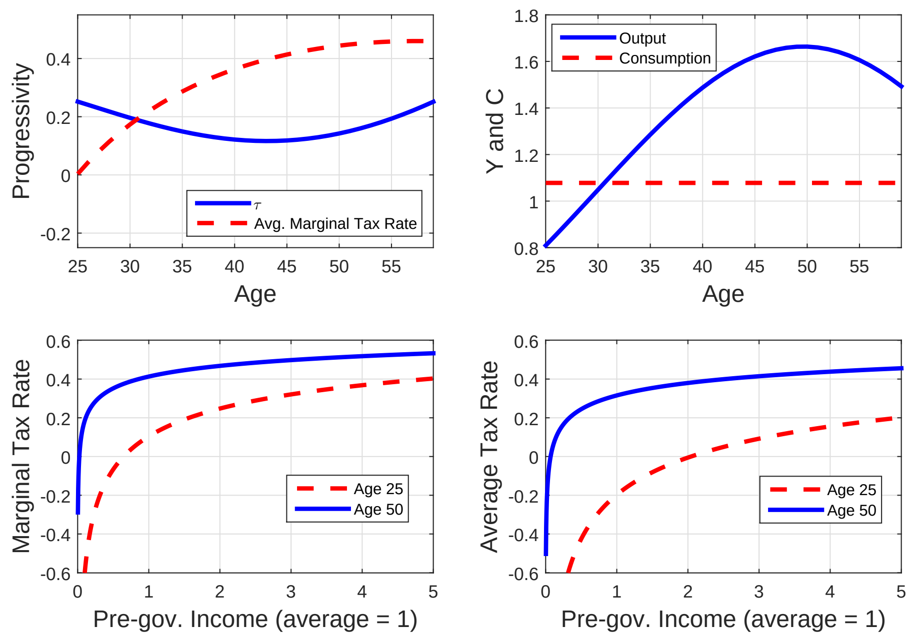

Figure 6: Optimal tax progressivity incorporating transition. This case features uninsurable risk ($v_\omega > 0$), the empirical age profile for wages, and $\beta = 0.95$. The initial steady state features $\tau_{US} = 0.181$. Top left panel: $\{\tau_a\}$ and income-weighted average marginal tax rate. Top right panel: average income and consumption by age in the final steady state. Bottom panels: marginal and average tax rate schedules at ages 25 and 50 in the final steady state.

is incorporated. The net result is that the optimal age profile of taxes is U-shaped, as in the case with $\beta = 1$.

The welfare gain of implementing optimally age-varying progressivity, relative to the best age-invariant value for progressivity, is now only 0.07% of consumption. Thus, the various channels for varying progressivity over the life cycle turn out to approximately cancel each other out. This is the main finding of our quantitative analysis.

6 Conclusions

This paper develops 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 provides 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 economy to the United States we find that when all of these channels are operative, they approximately cancel each other out. In particular, the optimal age profile for tax progressivity is mildly U-shaped when the transition is taken into account. Our main quantitative finding is that the potential welfare gains from introducing age-dependent progressivity are quantitatively small, at around one-tenth of a 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.

Chari, V. V., and P. J. Kehoe (1999): "Optimal Fiscal and Monetary Policy," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1, Part C, pp. 1671–1745. Elsevier.

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 Taxes," *Journal of Monetary Economics*, 55(4), 692–709.

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

——— (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.

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.

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. For a proof of Proposition 1, see Heathcote et al. (2017).

A.1 Proof of Proposition 2 [skill price and skill choice]

The education cost is given by $v(s) = \frac{\kappa^{-1/\psi}}{1+1/\psi}(s)^{1+1/\psi}$, where κ is exponentially distributed, $\kappa \sim \eta \exp(-\eta\kappa)$. Recall from eq. (12) in the main text that the optimality condition for skill investment is

$$v'(s) = \left(\frac{s}{\kappa}\right)^{\frac{1}{\psi}} = (1-\beta\delta)\,\mathbb{E}_0 \sum_{a=0}^{\infty} (\beta\delta)^a \frac{\partial u\left(c\left(\varphi, \alpha, s; \lambda_a, \tau_a, \bar{\tau}\right), h\left(\varphi; \tau_a\right), g\right)}{\partial s}. \tag{A1}$$

The skill level s affects only the consumption allocation (not the hours allocation) and only through the price $p(s; \boldsymbol{\tau})$, where we recall that the boldface notation $\boldsymbol{\tau}$ indicates the vector of progressivity rates $\boldsymbol{\tau} \equiv \{\tau_a\}_{a=0}^{\infty}$, assumed to be fixed over time. Hence, using (16), (A1) can be simplified as

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

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

$$\log p(s; \boldsymbol{\tau}) = \pi_0(\boldsymbol{\tau}) + \pi_1(\boldsymbol{\tau}) \cdot s, \tag{A2}$$

which implies that the skill allocation has the form[22]

$$s(\kappa; \boldsymbol{\tau}) = \left[\pi_1(\boldsymbol{\tau}) \cdot (1-\bar{\tau})\right]^{\psi} \cdot \kappa, \tag{A3}$$

[22] To see this, note that per assumption $\partial \log p(s; \boldsymbol{\tau}) / \partial s = \pi_1(\boldsymbol{\tau})$, so (A1) can be written as

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

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

$$\bar{\tau} \equiv (1 - \beta\delta) \sum_{a=0}^{\infty} (\beta\delta)^a \tau_a$$

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

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

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

$$\begin{aligned} \log p(s) &= \log\left[\frac{\partial Y}{\partial\left[N \cdot m(s)\right]}\right] = \frac{1}{\theta}\log Y - \frac{1}{\theta}\log\left[N \cdot m(s)\right] \\ &= \frac{1}{\theta}\log\left(\frac{Y}{N}\right) - \frac{1}{\theta}\log\zeta + \frac{\zeta}{\theta}s. \end{aligned} \tag{A4}$$

Equating coefficients across equations (A2) and (A4) implies $\pi_1(\boldsymbol{\tau}) = \frac{\zeta}{\theta} = \frac{\eta}{\theta}\left[(1-\bar{\tau})\,\pi_1(\boldsymbol{\tau})\right]^{-\psi}$, which yields

$$\pi_1(\boldsymbol{\tau}) = \left(\frac{\eta}{\theta}\right)^{\frac{1}{1+\psi}} (1-\bar{\tau})^{-\frac{\psi}{1+\psi}} \tag{A5}$$

and thus the equilibrium density of s is

$$m(s) = (\eta)^{\frac{1}{1+\psi}} \left(\frac{\theta}{1-\bar{\tau}}\right)^{\frac{\psi}{1+\psi}} \exp\left(-(\eta)^{\frac{1}{1+\psi}} \left(\frac{\theta}{1-\bar{\tau}}\right)^{\frac{\psi}{1+\psi}} s\right). \tag{A6}$$

Similarly, the base skill price is

$$\pi_0(\boldsymbol{\tau}) = \frac{1}{\theta}\log\left(\frac{Y}{N}\right) - \frac{\log\left(\frac{\eta}{\theta}\right)}{\theta(1+\psi)} + \frac{\psi}{\theta(1+\psi)}\log(1-\bar{\tau}). \tag{A7}$$

We derive a fully structural expression for $\pi_0(\boldsymbol{\tau})$ below in the proof of Corollary 2.2 when we solve for Y and N explicitly. From now on, we drop the boldface notation 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.2 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 (A3), and the second equality follows from (A5). Thus, log skill premia are exponentially distributed with parameter θ. The variance of log skill prices is

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

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

A.3 Proof of Corollary 2.2 [aggregate quantities]

From equation (14) aggregate hours worked by individuals of age a are

$$\begin{aligned} H(\tau_a) &= \int h(\varphi;\tau_a)\, dF(\varphi) = \exp\left(\frac{\log(1-\tau_a)}{(1+\hat{\sigma}_a)(1-\tau_a)}\right) \int \exp(-\varphi)\, dF(\varphi) \\ &= (1-\tau_a)^{\frac{1}{1+\sigma}}. \end{aligned}$$

Since α and φ are independent, it follows that $N_a(\tau_a) = \exp(x_a) \cdot \mathbb{E}[\exp(\alpha)] \cdot \mathbb{E}[h(\varphi;\tau_a)] = (1-\tau_a)^{\frac{1}{1+\sigma}} \cdot \exp(x_a)$. Finally, aggregate earnings are given by the sum of individual earnings,

$$\begin{aligned} Y &= (1-\delta)\sum_{a=0}^{\infty}\delta^a \mathbb{E}\left(y_a(\alpha,\varphi;\tau_a,\bar{\tau})\right) = (1-\delta)\sum_{a=0}^{\infty}\delta^a \mathbb{E}\left\{p(s;\bar{\tau})\exp(x_a+\alpha)\, h(\varphi;\tau_a)\right\} \\ &= \mathbb{E}\{p(s;\bar{\tau})\}(1-\delta)\sum_{a=0}^{\infty}\delta^a N_a(\tau_a), \end{aligned}$$

where

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

A.4 Proof of Proposition 3 [optimal age-dependent taxation in steady state]

From equation (25) the planner's problem can be written as:

$$\begin{aligned}
\max_{\{g,\lambda_a,\tau_a\}} \mathcal{W}(g,\{\lambda_a,\tau_a\}) &= (1-\delta)\sum_{a=0}^{\infty}\delta^a \bar{u}(a,\lambda_a,\boldsymbol{\tau}) - \frac{1-\delta}{1-\beta\delta}\bar{v}(\boldsymbol{\tau}) + \chi\log\left(g\sum_{a=0}^{\infty}\delta^a Y(a,\boldsymbol{\tau})\right) \qquad \text{(A8)}\\
&\quad \text{subject to}\\
\sum_{a=0}^{\infty}\delta^a\lambda_a\tilde{Y}(a,\lambda_a,\boldsymbol{\tau}) &= (1-g)\sum_{a=0}^{\infty}\delta^a Y(a,\boldsymbol{\tau}),\\
\int (y_{i,a})^{1-\tau_a}\,di &= \tilde{Y}(a,\lambda_a,\boldsymbol{\tau}) = \mathcal{K}(a,\tau_a,\bar{\tau})\exp\left(-\tau_a(1-\tau_a)a\frac{v_\omega}{2}\right)\\
Y(a,\boldsymbol{\tau}) &= (1-\tau_a)^{\frac{1}{1+\sigma}}\exp(x_a)\left(\frac{\theta}{\theta-1}\right)^{\frac{\theta}{\theta-1}}\left(\frac{1-\bar{\tau}}{\theta}\right)^{\frac{\psi}{(1+\psi)(\theta-1)}}\left(\frac{1}{\eta}\right)^{\frac{1}{(1+\psi)(\theta-1)}}
\end{aligned}$$

where the first additive component of equation (A8) is the utility from consumption net of the disutility of work effort, the second component is the indirect disutility of skill investment, and the third is the utility from the public good. After some tedious algebra, one obtains:

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

Letting ϑ denote the multiplier on the government budget constraint, and recognizing that $\partial\bar{u}(a,\lambda_a,\tau_a,\bar{\tau})/\partial\lambda_a = \lambda_a^{-1}$ from (16), the first-order condition with respect to λ_a gives:

$$\frac{1}{\lambda_a} = \vartheta\cdot\mathcal{K}(\tau_a,\bar{\tau})\exp\left(-\tau_a(1-\tau_a)a\frac{v_\omega}{2}\right). \qquad \text{(A9)}$$

Since $\lambda_a\tilde{Y}(a,\lambda_a,\boldsymbol{\tau}) = C(a,\lambda_a,\boldsymbol{\tau})$, average consumption of age group a, the first-order eq. (A9) implies that the planner wants to equate average consumption across ages. This proves statement (ii) of the proposition. Therefore, average consumption in the economy is $C(\boldsymbol{\tau}) = (1-g)(1-\delta)\sum_{a=0}^{\infty}\delta^a Y(a,\boldsymbol{\tau})$.

To make further progress on the exact expression for the social welfare function in (A8), we analyze each of its components. The first term can be written as:

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

Note that average log consumption for age group a is:

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

where the term in brackets is:

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

and the last term, from the optimality condition with respect to λ_a, is: $\log C\left(\boldsymbol{\tau}\right) = \log\left[\left(1-g\right)\left(1-\delta\right)\sum_{a=0}^{\infty}\delta^a Y\left(a, \boldsymbol{\tau}\right)\right]$.

Average disutility of hours worked in age group a is:

$$\int \frac{\exp\left(\left(1+\sigma\right)\varphi\right) h\left(\varphi; \tau_a\right)^{1+\sigma}}{1+\sigma} dF_\varphi = \frac{1-\tau_a}{1+\sigma}.$$

The average cost of skill investment in age group a is:

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

Combining these components and exploiting the fact that $Y\left(a, \boldsymbol{\tau}\right) = N_a\left(\tau_a\right) \cdot \mathbb{E}\left[p\left(s; \bar{\tau}\right)\right]$, we can rewrite the social welfare function (up to a constant) as:

$$\begin{aligned}
\mathcal{W}(g,\boldsymbol{\tau}) \quad = \quad & \log(1-g)+\chi\log g-(1-\delta)\sum_{a=0}^{\infty}\delta^{a}\underbrace{\frac{1-\tau_a}{1+\sigma}}_{\text{disutility of labor}} \\
& +(1+\chi)\log\{\underbrace{(1-\delta)\sum_{a=0}^{\infty}\delta^{a}(1-\tau_a)^{\frac{1}{1+\sigma}}\cdot\exp(x_a)\}}_{\text{Effective hours } N} \\
& +(1+\chi)\underbrace{\frac{1}{(1+\psi)(\theta-1)}\left[\psi\log(1-\bar{\tau})+\log\left(\frac{1}{\eta\theta^{\psi}}\left(\frac{\theta}{\theta-1}\right)^{\theta(1+\psi)}\right)\right]}_{\text{Productivity: log(average skill price) = log}(E(p(s)))} \\
& -\underbrace{\frac{1-\delta}{1-\beta\delta}\frac{\psi}{1+\psi}\frac{1}{\theta}(1-\bar{\tau})}_{\text{avg. education cost}}-(1-\delta)\sum_{a=0}^{\infty}\delta^{a}\underbrace{\left[\log\left(1-\left(\frac{1-\tau_a}{\theta}\right)\right)+\left(\frac{1-\tau_a}{\theta}\right)\right]}_{\text{cost of consumption dispersion across skills}} \\
& -(1-\delta)\sum_{a=0}^{\infty}\delta^{a}\cdot\frac{1}{2}\underbrace{(1-\tau_a)^2(v_\varphi+av_\omega)}_{\text{cons. dispersion due to unins. shocks and pref.}}
\end{aligned} \tag{A10}$$

The optimal choice of public good yields $g^* = \chi/(1+\chi)$, which proves statement (i) of the proposition. Substituting this optimal choice back into (A10), yields an expression for welfare that is only a function of the sequence $\{\tau_a\}$. Taking the first-order condition of this social welfare function with respect to τ_a (i.e., setting $\frac{\partial \mathcal{W}}{\partial \tau_a}=0$), we arrive at equation (26) in the text. Standard algebra establishes that the second-order condition is satisfied,

$$\begin{aligned}
\frac{\partial^2\mathcal{W}}{\partial^2\tau_a} \quad = \quad & -\frac{1}{(\theta-1+\tau_a)^2}-(v_\varphi+av_\omega) \\
& -\frac{1+\chi}{\theta-1}\frac{(1-\beta\delta)^2}{1-\delta}\frac{\psi}{1+\psi}\frac{(\delta\beta^2)^a}{(1-\bar{\tau})^2} \\
& -\frac{(1+\chi)(1-\delta)}{(1+\sigma)^2}\cdot(1-\tau_a)^{-\frac{2\sigma+1}{\sigma+1}}\exp(x_a)\cdot \\
& \cdot\frac{\delta^a(1-\tau_a)^{\frac{1}{1+\sigma}}\exp(x_a)+\sigma\sum_{t=0}^{\infty}\delta^t(1-\tau_t)^{\frac{1}{1+\sigma}}\cdot\exp(x_t)}{\left[(1-\delta)\sum_{t=0}^{\infty}\delta^t(1-\tau_t)^{\frac{1}{1+\sigma}}\cdot\exp(x_t)\right]^2} \\
< \quad & 0
\end{aligned}$$

This establishes that the social welfare function is globally concave in $\{\tau_a\}$, so the first-order condition (26) is necessary and sufficient for an optimum.

By inspecting (26), 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 sequence $\{x_a\}$ is constant, and one of the following conditions is satisfied: either $\beta \to 1$ or $\theta \to \infty$. Therefore, the sequence of τ_a must be independent of age in this case.

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 $(1-\tau_a)$. Similarly, 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 falling in τ_a. It follows that τ_a must increase with age.

It remains to be proven that, if the optimal sequence $\{\tau_a^*\}$ is increasing, so is $\{\lambda_a^*\}$, provided that the sequence $\{x_a\}$ is growing sufficiently slowly. This result simply derives from the fact that with slow growth in $\{x_a\}$, the average consumption of age a is decreasing in τ_a and increasing in λ_a (see the allocation in eq. (16)). Since the optimality condition (A9) requires average consumption to be equalized across ages, then λ_a must be increasing provided that the sequence $\{x_a\}$ is not rising too fast.