Online Appendix - Not For Publication

This online Appendix contains three sections: section A provides proofs omitted in the main text; section B shows the results in the general setup with risk aversion, discounting, heterogeneous life span, and overlapping generations; section C provides further details of the quantitative analysis and robustness checks.

A Proofs for Section 3

We suppress explicit dependence on θ throughout the appendix whenever it does not inhibit clarity.

A.1 Existence of retirement age

We first show that there must exist a retirement age if in the constrained optimum the Lagrange multiplier associated with incentive constraints (8) is non-negative. This can be interpreted as constraints (3) binding only downward, which is guaranteed by social welfare being redistributive from individuals with higher productivity toward those with lower productivity, i.e., $G(\theta) \geq F(\theta)$.

Proposition 5 *Suppose that in the constrained-efficient allocation the Lagrange multiplier on (8) is non-negative, i.e., (3) is binding only for $\hat{\theta} \leq \theta$. Then there exists age $R(\theta)$ such that individuals of type θ prefer to work if and only if $t < R(\theta)$.*

Proof.

Suppose that for an individual of type θ there exists an interval $[t_1, t_2]$ such that for values of t in this interval $y(t, \theta) = 0$ and another interval $[t_2, t_3]$ such that $y(t, \theta) > 0$. We show that any such allocation can be improved upon.

Consider a perturbation of the allocation for which $\tilde{y}(t, \theta) > 0$ for $t \in [t_1, t_1 + \epsilon]$ and $\tilde{y}(t, \theta) = 0$ for $t \in [t_2, t_2 + \epsilon]$. Furthermore, suppose that $\forall t \in [t_1, t_1 + \epsilon]$, $\tilde{y}(t, \theta) = \xi y(t + t_2 - t_1, \theta)$ where ξ is such that

$$\int_{t_1}^{t_1+\epsilon} \left(\frac{\tilde{y}(t,\theta)}{\varphi(t,\theta)} \right)^{1+1/\varepsilon} dt = \int_{t_2}^{t_2+\epsilon} \left(\frac{y(t,\theta)}{\varphi(t,\theta)} \right)^{1+1/\varepsilon} dt$$

Note that this perturbation keeps the length of working life and the disutility of leisure unchanged. However, since any nonworking must occur after $t^*(\theta)$ and as a result $\varphi(t_1, \theta) >$

$\varphi(t_2, \theta)$, the total output from type θ must increase. Therefore, if the allocations remain incentive compatible, such perturbation improves welfare.

In order to show incentive compatibility of the perturbed allocation, it is sufficient to show that

$$\int_{t_1}^{t_1+\epsilon} \left(\frac{\tilde{y}(t, \theta)}{\varphi(t, \theta')} \right)^{1+1/\varepsilon} dt > \int_{t_2}^{t_2+\epsilon} \left(\frac{y(t, \theta)}{\varphi(t, \theta')} \right)^{1+1/\varepsilon} dt, \forall \theta' > \theta \tag{16}$$

This implies that the value of reporting θ for all types $\theta' > \theta$ goes down in response to this perturbation and completes the argument since by assumption only downward incentive constraints bind.

To show (16), note that from Assumption 1 we have that

$$\frac{\varphi(t', \theta)}{\varphi(t, \theta)} < \frac{\varphi(t', \theta')}{\varphi(t, \theta')}, \theta' > \theta, t' > t$$

This inequality implies that

$$\left(\frac{\varphi(t', \theta)}{\varphi(t', \theta')} \right)^{1+1/\varepsilon} < \left(\frac{\varphi(t, \theta)}{\varphi\left(t, \theta'\right)} \right)^{1+1/\varepsilon}, \theta' > \theta, t' > t$$

Hence, multiplying the integrands in (16) by $\left(\frac{\varphi(t, \theta')}{\varphi(t, \theta)} \right)^{1+1/\varepsilon}$, makes the left-hand side larger, i.e., we must have that

$$\int_{t_1}^{t_1+\epsilon} \left(\frac{\tilde{y}(t, \theta)}{\varphi(t, \theta)} \right)^{1+1/\varepsilon} \left(\frac{\varphi(t, \theta)}{\varphi\left(t, \theta'\right)} \right)^{1+1/\varepsilon} dt > \int_{t_2}^{t_2+\epsilon} \left(\frac{y(t, \theta)}{\varphi(t, \theta)} \right)^{1+1/\varepsilon} \left(\frac{\varphi(t, \theta)}{\varphi\left(t, \theta'\right)} \right)^{1+1/\varepsilon} dt$$

This inequality implies (16) and concludes the proof. ■

A.2 Proof of Proposition 1

First note that when $\tilde{\varphi}\left(t, \theta\right)$ is the productivity profile and $\tilde{\eta}\left(\theta\right)$ is the fixed cost of working, optimal hours worked and the retirement decision under public information are given by

$$\psi\frac{y\left(t,\theta\right)^{1/\varepsilon}}{\tilde{\varphi}\left(t,\theta\right)^{1+1/\varepsilon}} = 1 \tag{17}$$

$$y\left(R\left(\theta\right),\theta\right) = \psi\frac{y\left(R\left(\theta\right),\theta\right)^{1+1/\varepsilon}}{\left(1+1/\varepsilon\right)\tilde{\varphi}\left(R\left(\theta\right),\theta\right)^{1+1/\varepsilon}}+\tilde{\eta}\left(\theta\right) \tag{18}$$

The planning problem with private information is to maximize $\int U(\theta)dG(\theta)$ subject to

$$U(\theta)=c(\theta)\bar{T}-\int_0^{R(\theta)}v(l(t,\theta))dt-(R(\theta)-t^*(\theta))\eta(\theta),$$

as well as (8) and (2). Suppressing θ, the first-order conditions are given by

$$g-\alpha-\mu' = 0 \tag{19}$$

$$\alpha-\lambda f = 0 \tag{20}$$

$$\forall t\leq R, -\psi\frac{y\left(t\right)^{1/\varepsilon}}{\varphi\left(t\right)^{1+1/\varepsilon}}\alpha+\lambda f-\mu\psi\left(1+1/\varepsilon\right)\frac{y\left(t\right)^{1/\varepsilon}}{\varphi\left(t\right)^{1+1/\varepsilon}}\frac{\varphi_\theta\left(t\right)}{\varphi\left(t\right)} = 0 \tag{21}$$

$$-\left[\frac{\psi}{\left(1+1/\varepsilon\right)}\frac{y\left(R\right)^{1+1/\varepsilon}}{\varphi\left(R\right)^{1+1/\varepsilon}}+\eta\right]\alpha+y\left(R\right)\lambda f-\left[\psi\frac{\varphi_\theta\left(R\right)}{\varphi\left(R\right)}\frac{y\left(R\right)^{1+1/\varepsilon}}{\varphi\left(R\right)^{1+1/\varepsilon}}-\eta'\right]\mu = 0 \tag{22}$$

$$\mu\left(\underline{\theta}\right)=\mu\left(\bar{\theta}\right) = 0,$$

where $\alpha\left(\theta\right)$ is the multiplier on the $U(\theta)$ constraint, λ is the multiplier on feasibility (2), and $\mu\left(\theta\right)$ is the multiplier on incentive constraints (8). Integrating over equation (19) and using the boundary conditions we have

$$\int_{\underline{\theta}}^{\bar{\theta}}g\left(\theta\right)d\theta-\lambda\int_{\underline{\theta}}^{\bar{\theta}}f\left(\theta\right)d\theta=0,$$

and hence $\lambda=1$ since $G\left(\bar{\theta}\right)=F\left(\bar{\theta}\right)=1$. Moreover, we also have

$$\begin{aligned}\mu\left(\theta\right) &= \int_{\underline{\theta}}^{\theta}\left[g\left(\theta'\right)-f\left(\theta'\right)\right]d\theta' \\ &= G\left(\theta\right)-F\left(\theta\right)\end{aligned}$$

and we can rewrite equation (21) as

$$\psi \frac{y(t)^{1/\varepsilon}}{\varphi(t)^{1+1/\varepsilon}} \left[1 + (1+1/\varepsilon) \frac{G(\theta)-F(\theta)}{f(\theta)} \frac{\varphi_\theta(t)}{\varphi}\right] = 1,$$

while (22) becomes

$$\begin{aligned} y(R) &= \left[\psi \frac{\varphi_\theta(R)}{\varphi(R)} \frac{y(R)^{1+1/\varepsilon}}{\varphi(R)^{1+1/\varepsilon}} - \eta'\right] \frac{G(\theta)-F(\theta)}{f(\theta)} + \left[\frac{\psi}{(1+1/\varepsilon)} \frac{y(R)^{1+1/\varepsilon}}{\varphi(R)^{1+1/\varepsilon}} + \eta\right] \\ &= \psi \frac{y(R)^{1+1/\varepsilon}}{(1+1/\varepsilon)\,\varphi(R)^{1+1/\varepsilon}} \left[1 + (1+1/\varepsilon) \frac{G(\theta)-F(\theta)}{f(\theta)} \frac{\varphi_\theta(R)}{\varphi(R)}\right] \\ &\quad + \eta - \eta' \frac{G(\theta)-F(\theta)}{f(\theta)}. \end{aligned}$$

Defining modified productivities and fixed costs

$$\begin{aligned} \tilde{\varphi}(t,\theta) &= \varphi(t,\theta) \left[1 + (1+1/\varepsilon) \frac{G(\theta)-F(\theta)}{f(\theta)} \frac{\varphi_\theta(t,\theta)}{\varphi(t,\theta)}\right]^{-\frac{\varepsilon}{1+\varepsilon}} \\ \tilde{\eta}(\theta) &= \eta(\theta) \left[1 - \frac{G(\theta)-F(\theta)}{f(\theta)} \frac{\eta'(\theta)}{\eta(\theta)}\right], \end{aligned}$$

it can be easily seen that (17) and (18) are satisfied for these productivity profiles and fixed costs, establishing the claim. ■

A.3 Proof of Proposition 2

The first-order conditions associated with the mechanism design problem are given by (19)-(22) and the boundary conditions. Note that integrating over equations (19) and (20) and using the boundary conditions $\mu(\underline{\theta}) = \mu(\bar{\theta}) = 0$ implies that $\lambda = 1$.

We can then write the third condition (21) as

$$\psi \frac{y(t)^{1/\varepsilon}}{\varphi(t)^{1+1/\varepsilon}} \left[1 + (1+1/\varepsilon) \frac{G(\theta)-F(\theta)}{f(\theta)} \frac{\varphi_\theta}{\varphi}\right] = 1 \rightarrow y(t) = \psi^{-\varepsilon} \tilde{\varphi}(t,\theta)^{1+\varepsilon}$$

Replacing the above in (22), we have

$$\tilde{\varphi}(R(\theta),\theta)^{1+\varepsilon} = \tilde{\eta}(\theta)(1+\varepsilon)\psi^{\varepsilon}$$

This proves the claim. ■

A.4 Proof of Proposition 3

It is sufficient to show equation (11) in the proposition since the rest of the argument is given in the main text. Note that the definitions of wedges (5) and (6) imply

$$\begin{aligned}\tau_R(\theta)\, y(R(\theta),\theta) &= y(R(\theta),\theta) - \left[\frac{\psi}{1+1/\varepsilon}\frac{y(R(\theta),\theta)^{1+1/\varepsilon}}{\varphi(R(\theta),\theta)^{1+1/\varepsilon}} + \eta(\theta)\right] \\ \tau_y(R(\theta),\theta)\, y(R(\theta),\theta) &= y(R(\theta),\theta) - \psi\frac{y(R(\theta),\theta)^{1+1/\varepsilon}}{\varphi(R(\theta),\theta)^{1+1/\varepsilon}},\end{aligned}$$

and hence, using (21) and (22) we arrive at

$$\begin{aligned}&\tau_R(\theta)\, y(R(\theta),\theta) - \frac{1}{1+1/\varepsilon}\tau_y(R(\theta),\theta)\, y(R(\theta),\theta) \\ &= (1+\varepsilon)\, y(R(\theta),\theta) - \eta(\theta) \\ &= -\frac{G(\theta)-F(\theta)}{f(\theta)}\eta'(\theta),\end{aligned}$$

where the last equality follows from (17) and (18). ∎

A.5 Proof of Lemma 1

Since $y(t,\theta)$ is increasing and continuous in θ, then the range of $y(t,\cdot): \Theta \mapsto \mathbb{R}$ is an interval, $\left[\underline{y}(t), \bar{y}(t)\right]$. The envelope condition associated with the maximization in (13) is given by

$$v'(\theta) = \psi\frac{\varphi_\theta(t,\theta)}{\varphi(t,\theta)}\frac{y(t,\theta)^{1+1/\varepsilon}}{\varphi(t,\theta)^{1+1/\varepsilon}},$$

where $v(\theta) = y(t,\theta) - \mathcal{T}(t, y(t,\theta)) - \frac{\psi}{1+1/\varepsilon}\frac{y(t,\hat{\theta})^{1+1/\varepsilon}}{\varphi(t,\theta)^{1+1/\varepsilon}}$. Hence, we must have

$$v(\theta) = \underline{v} + \int_{\underline{\theta}(t)}^{\theta}\psi\frac{\varphi_\theta(t,\hat{\theta})}{\varphi(t,\hat{\theta})}\frac{y(t,\hat{\theta})^{1+1/\varepsilon}}{\varphi(t,\hat{\theta})^{1+1/\varepsilon}}d\hat{\theta},$$

where $\underline{\theta}(t)$ is the lowest type that works at age t. As a result

$$y(t,\theta) - \mathcal{T}(t, y(t,\theta)) - \frac{\psi}{1+1/\varepsilon}\frac{y(t,\hat{\theta})^{1+1/\varepsilon}}{\varphi(t,\theta)^{1+1/\varepsilon}} = \underline{v} + \int_{\underline{\theta}(t)}^{\theta}\psi\frac{\varphi_\theta(t,\hat{\theta})}{\varphi(t,\hat{\theta})}\frac{y(t,\hat{\theta})^{1+1/\varepsilon}}{\varphi(t,\hat{\theta})^{1+1/\varepsilon}}d\hat{\theta}$$

This equation uniquely defines $\mathcal{T}(t,\cdot)$ up to a constant. Furthermore, since $y(t,\theta)$ is increasing in θ, standard mechanism design arguments establish that for any such $\mathcal{T}$, equation (13) is indeed satisfied (see, e.g., Fudenberg and Tirole (1991)). ■

A.6 Proof of Lemma 2

Given the tax schedule and benefit formula, a household of type θ's optimization problem is given by

$$\begin{array}{ll} \max_{R(\theta),y(t)} & \int_0^{R(\theta)} \left[y(t,\theta) - \mathcal{T}(t, y(t,\theta))\right] dt + b(R) \\ & - \int_0^{R(\theta)} \left[\frac{\psi}{1+1/\varepsilon} \frac{y(t,\theta)^{1+1/\varepsilon}}{\varphi(t,\theta)^{1+1/\varepsilon}}\right] dt - \eta(\theta)\left(R(\theta) - t^*(\theta)\right) \end{array} \tag{23}$$

We refer to the solution for (23) as $\tilde{y}(t,\theta)$ and $\tilde{R}(\theta)$. Our goal is to show that $\tilde{y}(t,\theta) = y(t,\theta)$ and $\tilde{R}(\theta) = R(\theta)$. We start by showing the following lemma:

Lemma 3 *If $\tilde{y}(t,\theta) \geq 0$, then $\tilde{y}(t,\theta) = y(t,\theta)$.*

Proof. The proof simply follows from the definition of $\mathcal{T}$. Since $\theta = \arg\max_{\hat{\theta}} y\left(t,\hat{\theta}\right) - \mathcal{T}\left(t, y\left(t,\hat{\theta}\right)\right) - v\left(\frac{y(t,\hat{\theta})}{\varphi(t,\theta)}\right)$, it is optimal for an individual of type θ to choose $y(t,\theta)$ at age t. ■

Using the above lemma, we can show the following:

Lemma 4 *Choosing $\{y(t,\theta)\}_{t\leq R(\theta)}, R(\theta)$ for an agent of type θ is a local optimum for an individual of type θ in (23).*

Proof. Suppose on the contrary that the individual chooses $R\left(\hat{\theta}\right) \neq R(\theta)$, then given the definition of b, the utility for the household is given by

$$\int_0^{R(\hat{\theta})} \left[y(t,\theta) - \mathcal{T}(t, y(t,\theta))\right] dt - \int_0^{R(\hat{\theta})} \left[\frac{\psi}{1+1/\varepsilon} \frac{y(t,\theta)^{1+1/\varepsilon}}{\varphi(t,\theta)^{1+1/\varepsilon}}\right] dt - \eta(\theta)\left(R(\theta) - t^*(\theta)\right)$$
$$+c\left(\hat{\theta}\right)\bar{T} - \int_0^{R(\hat{\theta})} \left[y\left(t,\hat{\theta}\right) - \mathcal{T}\left(t, y\left(t,\hat{\theta}\right)\right)\right] dt \tag{24}$$

Taking a derivative with respect to $\hat{\theta}$, we have

$$\left[y\left(R\left(\hat{\theta}\right),\theta\right)-\mathcal{T}\left(R\left(\hat{\theta}\right),y\left(R\left(\hat{\theta}\right),\theta\right)\right)-\frac{\psi}{1+1/\varepsilon}\frac{y\left(R\left(\hat{\theta}\right),\theta\right)^{1+1/\varepsilon}}{\varphi\left(R\left(\hat{\theta}\right),\theta\right)^{1+1/\varepsilon}}-\eta\left(\theta\right)\right]R'\left(\hat{\theta}\right)$$
$$+c'\left(\hat{\theta}\right)\bar{T}-\left[y\left(R\left(\hat{\theta}\right),\hat{\theta}\right)-\mathcal{T}\left(R\left(\hat{\theta}\right),y\left(R\left(\hat{\theta}\right),\hat{\theta}\right)\right)\right]R'\left(\hat{\theta}\right)$$
$$-\int_0^{R(\hat{\theta})}\frac{\partial}{\partial\theta}y\left(t,\hat{\theta}\right)\left[1-\frac{\partial}{\partial y}\mathcal{T}\left(t,y\left(t,\hat{\theta}\right)\right)\right]dt$$

Evaluating the above expression when $\hat{\theta}=\theta$,

$$c'\left(\theta\right)\bar{T}-\left[\frac{\psi}{1+1/\varepsilon}\frac{y\left(R\left(\theta\right),\theta\right)^{1+1/\varepsilon}}{\varphi\left(R\left(\theta\right),\theta\right)^{1+1/\varepsilon}}+\eta\left(\theta\right)\right]R'\left(\theta\right)-\int_0^{R(\theta)}\frac{\partial}{\partial\theta}y\left(t,\theta\right)\left[1-\frac{\partial}{\partial y}\mathcal{T}\left(t,y\left(t,\theta\right)\right)\right]dt,$$

and by static incentive compatibility (13) the above expression becomes

$$c'\left(\theta\right)\bar{T}-\left[\frac{\psi}{1+1/\varepsilon}\frac{y\left(R\left(\theta\right),\theta\right)^{1+1/\varepsilon}}{\varphi\left(R\left(\theta\right),\theta\right)^{1+1/\varepsilon}}+\eta\left(\theta\right)\right]R'\left(\theta\right)-\int_0^{R(\theta)}\psi\frac{y\left(t,\theta\right)^{1/\varepsilon}}{\varphi\left(t,\theta\right)^{1+1/\varepsilon}}\frac{\partial}{\partial\theta}y\left(t,\theta\right)dt,$$

which is zero by incentive compatibility of the original allocation. This implies that $\hat{\theta}=\theta$ is a local optimum point of the function (24), and hence $R=R\left(\hat{\theta}\right)$ is a local optimum of (23). ■

Together the above lemmas establish the claim in Lemma 2. ■

B Proofs for the general setup of Section 2

We extend here the qualitative results in the main text to the general setup with risk aversion and discounting.

B.1 Equivalence

We start by showing that the equivalence result generalizes in an intuitive way and then show that the sufficient-statistic-type expressions remain unchanged.

The first-order conditions (19)-(22) associated with the mechanism design problem now become

$$g-\alpha-\mu'=0 \tag{25}$$

$$\alpha u'(c)-\lambda f=0 \tag{26}$$

$$\forall t \leq R, \quad -\psi \frac{y(t)^{1/\varepsilon}}{\varphi(t)^{1+1/\varepsilon}} \alpha + \lambda f - \mu\psi (1 + 1/\varepsilon) \frac{y(t)^{1/\varepsilon}}{\varphi(t)^{1+1/\varepsilon}} \frac{\varphi_\theta(t)}{\varphi(t)} = 0 \tag{27}$$

$$-\left[\frac{\psi}{1+1/\varepsilon} \frac{y(R)^{1+1/\varepsilon}}{\varphi(R)^{1+1/\varepsilon}} + \eta\right] \alpha + y(R)\lambda f - \left[\psi \frac{\varphi_\theta(R)}{\varphi(R)} \frac{y(R)^{1+1/\varepsilon}}{\varphi(R)^{1+1/\varepsilon}} - \eta'\right] \mu = 0 \tag{28}$$

$$\mu(\underline{\theta}) = \mu(\bar{\theta}) = 0$$

Note that the previous arguments for treating consumption levels as constant over time are unaffected. Thus the first two equations (25) and (26) combined with the boundary conditions give

$$\mu(\theta) = \int_{\underline{\theta}}^{\theta} \left[g(\theta') - \frac{\lambda f(\theta')}{u'(c(\theta'))}\right] d\theta'$$

Recall that λ is the marginal value of public funds for the planner. Following Saez (2001) and Saez and Stantcheva (2016), let us denote $\bar{g}(\theta)$ the ratio of the marginal value of public funds for the planner to utilitarian marginal welfare from per-capita consumption of θ-type individuals, i.e., $\bar{g}(\theta) = \frac{\lambda f(\theta)}{u'(c(\theta))}$. In other words, a utilitarian planner is indifferent between $\bar{g}(\theta)$ more public funds and a marginal decrease in θ-type consumption. The larger $\bar{g}(\theta)$, the less a utilitarian planner values consumption by θ-types and therefore $\bar{g}(\theta)$ is a parameter reflecting exogenous motive to redistribute from θ. The same motive for all types up to θ is then $\bar{G}(\theta) = \int_{\underline{\theta}}^{\theta} \frac{\lambda f(\theta')}{u'(c(\theta'))} d\theta'$. Thus

$$\mu(\theta) = G(\theta) - \bar{G}(\theta).$$

We can then re-write equation (27) as

$$\frac{\psi y(t)^{1/\varepsilon}}{\varphi(t)^{1+1/\varepsilon}} \left[1 + \left(1 + \frac{1}{\varepsilon}\right) \frac{G(\theta) - \bar{G}(\theta)}{\bar{g}(\theta)} \frac{\varphi_\theta(t)}{\varphi}\right] \frac{\bar{g}(\theta)}{\lambda f(\theta)} = 1,$$

and (28) becomes

$$\begin{aligned} y(R) \quad = \quad & \frac{\psi y(R)^{1+1/\varepsilon}}{(1+1/\varepsilon)\varphi(R)^{1+1/\varepsilon}} \left[1 + \left(1 + \frac{1}{\varepsilon}\right) \frac{G(\theta) - \bar{G}(\theta)}{\bar{g}(\theta)} \frac{\varphi_\theta(R)}{\varphi(R)}\right] \frac{\bar{g}(\theta)}{\lambda f(\theta)} \\ & + \left[\eta - \eta' \frac{G(\theta) - \bar{G}(\theta)}{\bar{g}(\theta)}\right] \frac{\bar{g}(\theta)}{\lambda f(\theta)}. \end{aligned}$$

Then the equivalence result in the general environment follows if the virtual productivities and fixed costs are given by

$$\begin{aligned}\tilde{\varphi}\left(t,\theta\right) &= \varphi(t,\theta)\left[\left(1+\left(1+\frac{1}{\varepsilon}\right)\frac{G\left(\theta\right)-\bar{G}\left(\theta\right)}{\bar{g}\left(\theta\right)}\frac{\varphi_{\theta}\left(t,\theta\right)}{\varphi\left(t,\theta\right)}\right)\frac{\bar{g}(\theta)}{\lambda f(\theta)}\right]^{-\frac{\varepsilon}{1+\varepsilon}},\\ \tilde{\eta}\left(\theta\right) &= \eta\left(\theta\right)\left[\left(1-\frac{G\left(\theta\right)-\bar{G}\left(\theta\right)}{\bar{g}\left(\theta\right)}\frac{\eta'\left(\theta\right)}{\eta\left(\theta\right)}\right)\frac{\bar{g}(\theta)}{\lambda f(\theta)}\right].\end{aligned}$$

That is, to maintain the full information optimality (17) and (18) the modification of productivities and fixed costs must be given by the above.

Intuitively, these are analogous modifications to the ones we derived in the main text for the quasi-linear example with the exception of what is being used as a benchmark of utilitarian redistribution. With risk aversion it is no longer enough to simply compare redistributive motives, G, against the distribution of types, F, as was the case in the quasi-linear example. Since risk aversion introduces curvature into the utility of consumption, even a utilitarian benchmark would take that into account trading it off against funds available to the rest of the population. This is precisely what $\bar{G}$ captures.

On the other hand, if curvature is removed the planner is able to transform public funds one for one into the utility of consumption of a given type, making $\bar{G}$ equal to F and making the virtual types above identical to the ones derived in the main text.

Finally, using these definitions of virtual types we can re-write (28) as

$$\tilde{\varphi}(R(\theta),\theta)^{1+\varepsilon} = \tilde{\eta}(\theta)\left(1+\varepsilon\right)\psi^{\varepsilon}$$

Thus, the sufficient statistic expression is unchanged and as in the main text the constrained-efficient retirement age is increasing if and only if $\frac{\partial}{\partial\theta}\frac{\tilde{\varphi}(t,\theta)^{1+\varepsilon}}{\tilde{\eta}(\theta)}\Big|_{t=R(\theta)} \geq 0$.

B.2 Retirement incentives vs. hours incentives

Note first that an implication of the first-order conditions (25) and (26) combined with the boundary conditions can be re-written as

$$\mu\left(\theta\right) = \int_{\theta}^{\bar{\theta}}\left[\frac{\lambda f\left(\theta'\right)}{u'\left(c\left(\theta'\right)\right)} - g\left(\theta'\right)\right]d\theta'$$

Multiplying the third equation (27) for $t = R\left(\theta\right)$ by $\frac{y(R)}{1+1/\varepsilon}$ and subtracting from (28) we have

$$-\eta\alpha + \left(1+\varepsilon\right)y\left(R\right)\lambda f + \eta'\mu = 0$$

Hence,

$$\begin{aligned}(1+\varepsilon)\,y(R) &= \frac{\eta}{u'(c(\theta))} - \eta' \frac{\mu}{\lambda f} \\ &= \frac{\eta}{u'(c(\theta))} - \eta' \frac{1-F(\theta)}{f(\theta)} \int_{\theta}^{\bar{\theta}} \left[\frac{1}{u'(c(\theta'))} - \frac{g(\theta')}{\lambda f(\theta')} \right] \frac{dF(\theta')}{1-F(\theta)} \end{aligned} \quad (29)$$

Note that the wedges are given by

$$\begin{aligned}\tau_R(\theta)\,y(R(\theta),\theta) &= y(R(\theta),\theta) - \frac{1}{u'(c(\theta))}\left[\frac{\psi}{1+1/\varepsilon}\frac{y(R(\theta),\theta)^{1+1/\varepsilon}}{\varphi(R(\theta),\theta)^{1+1/\varepsilon}} + \eta(\theta)\right] \\ \tau_y(R(\theta),\theta)\,y(R(\theta),\theta) &= y(R(\theta),\theta) - \frac{1}{u'(c(\theta))}\psi\frac{y(R(\theta),\theta)^{1+1/\varepsilon}}{\varphi(R(\theta),\theta)^{1+1/\varepsilon}}\end{aligned}$$

and hence

$$\begin{aligned}&\tau_R(\theta)\,y(R(\theta),\theta) - \frac{1}{1+1/\varepsilon}\tau_y(R(\theta),\theta)\,y(R(\theta),\theta) \\ &= (1+\varepsilon)\,y(R(\theta),\theta) - \frac{\eta(\theta)}{u'(c(\theta))} \\ &= -\eta'\frac{1-F(\theta)}{f(\theta)}\int_{\theta}^{\bar{\theta}}\left[\frac{1}{u'(c(\theta'))} - \frac{g(\theta')}{\lambda f(\theta')}\right]\frac{dF(\theta')}{1-F(\theta)},\end{aligned}$$

where the last equality follows from (29). ■

B.3 Implementation

Here we construct tax and pension benefit policies to implement an incentive-compatible allocation in our general model. Our goal is to construct an age-dependent tax function, $\mathcal{T}(t,y)$, as well as a benefit function given by present value $b(R,Y)$, where Y is a measure of life-time earnings. Throughout this section we keep the assumption that $v(l) = \psi l^{1+1/\varepsilon}/(1+1/\varepsilon)$. We make the following assumptions about the implemented allocation:

Assumption 2 *The function $y(t,\cdot) : \Theta \longmapsto \mathbb{R}$ is increasing in θ.*

This assumption implies that a tax function that only depends on income at age t is well defined. Note that in this implementation each individual is solving the following optimization problem:

$$\max_{y(t),R,c} \int_0^{\bar{T}} e^{-\rho t} u(c)\, dt - \int_0^R e^{-\rho t} \left[v(y(t)/\varphi(t,\theta)) + \eta(t,\theta) \right] dt$$

subject to

$$\int_0^{\bar{T}} e^{-\rho t} c dt = \int_0^R e^{-\rho t} \left[y(t) - \mathcal{T}(t, y(t)) \right] dt + e^{-\rho R} b \left(R, \frac{1}{R} \int_0^R y(t)\, dt \right)$$

As in the main text, we start by constructing the tax function. For any tax function, $\mathcal{T}(t,y)$, we consider the following auxiliary optimization problem:

$$\mathcal{V}(b,R,Y;\theta,\mathcal{T}) = \max_{\{\tilde{y}(t)\}_{t\in[0,R]},c} \int_0^{\bar{T}} e^{-\rho t} u(c)\, dt - \int_0^R e^{-\rho t} \left[v(\tilde{y}(t)/\varphi(t,\theta)) + \eta(t,\theta) \right] dt \tag{30}$$

subject to

$$\begin{aligned} \int_0^{\bar{T}} e^{-\rho t} c dt &= \int_0^{\bar{T}} e^{-\rho t} \left[\tilde{y}(t) - \mathcal{T}(t, \tilde{y}(t)) \right] dt + e^{-\rho R} b \\ Y &= \frac{1}{R} \int_0^{\bar{T}} \tilde{y}(t)\, dt \end{aligned} \tag{31}$$

We say a tax function, $\mathcal{T}(t,y)$, *partially implements* the allocation $c(\theta), \{y(t,\theta)\}_{t\in[0,R(\theta)]}, R(\theta)$, if $\{y(t,\theta)\}_{t\in[0,R(\theta)]}$ is the solution to (30) when $R = R(\theta)$,

$$b = e^{\rho R(\theta)} \left[\int_0^{\bar{T}} e^{-\rho t} c(\theta)\, dt - \int_0^{R(\theta)} e^{-\rho t} \left[y(t,\theta) - \mathcal{T}(t, y(t,\theta)) \right] dt \right],$$

and $Y = Y(\theta) = \frac{1}{R(\theta)} \int_0^{R(\theta)} y(t,\theta)\, dt$. The following lemma establishes that such a tax function exists.

Lemma 5 *Consider the constrained-efficient allocation with extended earnings* $\left\{ c(\theta), \{y(t,\theta)\}_{t\in[0,1]}, R(\theta) \right\}_{\theta\in\Theta}$. *Then*

1. *There exists a tax function* $\mathcal{T}(t,y)$ *so that* $\{y(t,\theta)\}_{t\in[0,R(\theta)]}$ *is a local optimum in (30) for* $R = R(\theta)$, $b = e^{\rho R(\theta)} \left[\int_0^{\bar{T}} e^{-\rho t} c(\theta)\, dt - \int_0^{R(\theta)} e^{-\rho t} \left[y(t,\theta) - \mathcal{T}(t, y(t,\theta)) \right] dt \right]$, *and* $Y = Y(\theta) = \frac{1}{R(\theta)} \int_0^{R(\theta)} y(t,\theta)\, dt$.

2. *Suppose that for all* t, $\tau_y(t,\theta) < 1$ *and that* $\frac{y(t,\theta)^{1+\frac{1}{\varepsilon}}}{1-\tau_y(t,\theta)}$ *and* $y(t,\theta)$ *move together. That is,*

$y(t,\theta) > y(t,\theta')$ if and only if $\frac{y(t,\theta)^{1+\frac{1}{\varepsilon}}}{1-\tau_y(t,\theta)} > \frac{y(t,\theta')^{1+\frac{1}{\varepsilon}}}{1-\tau_y(t,\theta')}$. Then there exists a tax function that partially implements the allocation $c(\theta), \{y(t,\theta)\}_{t\in[0,R(\theta)]}, R(\theta)$.

Proof. 1. We prove the first part by constructing one such tax function. Let $\tau_y(t,\theta)$ be the labor wedge faced by individual θ at age t. We define the tax function $\mathcal{T}$ as follows:

$$\begin{aligned} \mathcal{T}(t,y) &= \underline{\mathcal{T}}(t) + \int_{\underline{y}(t)}^{\min\{y,\bar{y}(t)\}} \tau_y\left(t,\hat{\theta}(y,t)\right) dy \\ &\quad + (y - \min\{y,\bar{y}(t)\}), \forall y \in \left[\underline{y}(t), \infty\right) \\ \mathcal{T}(t,y) &= \underline{\mathcal{T}}(t), \forall y \in \left[0, \underline{y}(t)\right), \end{aligned} \tag{32}$$

where $\underline{y}(t) = \min_{\theta\in\Theta} y(t,\theta)$, $\bar{y}(t) = \max_{\theta\in\Theta} y(t,\theta)$, and $y\left(t,\hat{\theta}(t,y)\right) = y$. Note that the function $\hat{\theta}(t,y)$ is well-defined since $y(t,\theta)$ is a one-to-one function of θ. In addition, $\underline{\mathcal{T}}(t)$ is an arbitrary function of age. The intercept of the tax function $\underline{\mathcal{T}}(t)$ is an arbitrary continuous function. Note that (32) implies that the marginal tax rate is 1 for values of y above $\bar{y}(t)$. Given our construction, it is clear that $\{y(t,\theta)\}_{t\in[0,R(\theta)]}$ satisfies the first-order conditions associated with (30) when the last constraint is slack – since marginal tax rate evaluated at $y(t,\theta)$ is equal to the labor wedge.

2. To prove the second claim, we conjecture that the constraint (31) is slack. Then any solution to (30) must satisfy the following local optimality condition:

$$u'(c)\left(1 - \mathcal{T}_y(t,\tilde{y}(t))\right) = \psi \frac{\tilde{y}(t)^{1/\varepsilon}}{\varphi(t,\theta)^{1+1/\varepsilon}}$$

We can rewrite this equation as

$$\left(1 - \mathcal{T}_y(t,\tilde{y}(t))\right)\tilde{y}(t)^{-1/\varepsilon} = \psi \frac{u'(c)}{\varphi(t,\theta)^{1+1/\varepsilon}}$$

By the assumption stated above, the left-hand side of the above equation is decreasing in $\tilde{y}(t)$. Now suppose that $c > c(\theta)$, then $u'(c) < u'(c(\theta))$ and hence $\tilde{y}(t) < y(t,\theta)$. Since marginal tax rates are less than 1,

$$\int_0^{R(\theta)} e^{-\rho t}\left[\tilde{y}(t) - \mathcal{T}(t,\tilde{y}(t))\right] dt \leq \int_0^{R(\theta)} e^{-\rho t}\left[y(t,\theta) - \mathcal{T}(t,y(t,\theta))\right] dt,$$

which implies that the budget constraint cannot hold. A similar argument implies that $c < c(\theta)$ cannot hold, and hence the unique solution to a relaxed version of (30) is

$\{y(t,\theta)\}_{t\in[0,R(\theta)]}, c(\theta)$ which satisfies (31). This establishes the claim. ■

The properties assumed above are satisfied in all of our numerical simulations. Note also that there are potentially many tax functions that partially implement the constrained-efficient allocations. Our construction of benefits below works for all tax functions that partially implement the constrained-efficient allocations.

Consider a tax function $\mathcal{T}(t,y)$ that partially implements the constrained-efficient allocations and its associated pension benefits defined by

$$\hat{b}(\theta) = e^{\rho R(\theta)} \left[\int_0^{\bar{T}} e^{-\rho t} c(\theta)\, dt - \int_0^{R(\theta)} e^{-\rho t} \left[y(t,\theta) - \mathcal{T}(t, y(t,\theta)) \right] dt \right]$$

Our goal is to find a function $b(R,Y)$ such that $b(R(\theta), Y(\theta)) = \hat{b}(\theta)$ and

$$\mathcal{V}(b(R,Y), R, Y; \theta, \mathcal{T}) \leq \mathcal{V}(b(\theta), R(\theta), Y(\theta); \theta, \mathcal{T})$$

This would imply that an individual of type θ finds it optimal to choose $R(\theta)$ and $Y(\theta)$. Lemma 5 then implies that having chosen $R(\theta)$ and $Y(\theta)$, the optimal choice for earnings and consumption is the constrained-efficient allocation. The following proposition establishes that function $b(\cdot,\cdot)$ exists.

Proposition 6 *Suppose that Θ is a compact set. There exists a function $b(\cdot,\cdot)$ such that $b(R(\theta), Y(\theta)) = \hat{b}(\theta)$ and $\mathcal{V}(b(R,Y), R, Y; \theta, \mathcal{T}) \leq \mathcal{V}(b(\theta), R(\theta), Y(\theta); \theta, \mathcal{T})$.*

Proof. Our proof of this proposition is in two steps. First, we show that

$$\theta \in \arg\max_{\theta'} \mathcal{V}\left(b\left(\theta'\right), R\left(\theta'\right), Y\left(\theta'\right); \theta, \mathcal{T}\right) \tag{33}$$

We show this by first showing that $\theta' = \theta$ is the local optimum in the above optimization. To see this, we take a derivative of the above function and apply envelope condition to get

$$\begin{aligned} \frac{\partial}{\partial \theta'} \mathcal{V}\left(b\left(\theta'\right), R\left(\theta'\right), Y\left(\theta'\right); \theta, \mathcal{T}\right)\Big|_{\theta'=\theta} &= \\ \mathcal{V}_b b'(\theta) + \mathcal{V}_R R'(\theta) + \mathcal{V}_Y Y'(\theta) &= \end{aligned}$$

$$
\begin{aligned}
& e^{-\rho R(\theta)} u'(c(\theta)) b'(\theta) \\
& + \Bigg\{ -e^{-\rho R(\theta)} \left[v\left(\frac{y(t,\theta)}{\varphi(t,\theta)} \right) + \eta(\theta) \right] \\
& + u'(c(\theta)) \left[e^{-\rho R(\theta)} \left(y(R(\theta),\theta) - \mathcal{T}(R(\theta), y(R(\theta),\theta)) \right) - \rho e^{-\rho R(\theta)} b(\theta) \right] \\
& + \mu \left[\frac{1}{R(\theta)^2} \int_0^{R(\theta)} y(t,\theta)\, dt - \frac{1}{R(\theta)} y(R(\theta),\theta) \right] \Bigg\} R'(\theta) - \mu Y'(\theta) =
\end{aligned}
$$

$$
\begin{aligned}
& e^{-\rho R(\theta)} u'(c(\theta)) \left[\begin{array}{c} \rho R'(\theta) b(\theta) + e^{\rho R(\theta)} \int_0^{\bar{T}} e^{-\rho t} c'(\theta)\, dt \\ -e^{\rho R(\theta)} \int_0^{R(\theta)} e^{-\rho t} \left[1 - \mathcal{T}_y(t, y(t,\theta)) \right] y_\theta(t,\theta)\, dt \\ -\left[y(R(\theta),\theta) - \mathcal{T}(R(\theta), y(R(\theta),\theta)) \right] R'(\theta) \end{array} \right] \\
& + \Bigg\{ -e^{-\rho R(\theta)} \left[v\left(\frac{y(t,\theta)}{\varphi(t,\theta)} \right) + \eta(\theta) \right] \\
& \quad + u'(c(\theta)) \left[e^{-\rho R(\theta)} \left(y(R(\theta),\theta) - \mathcal{T}(R(\theta), y(R(\theta),\theta)) \right) - \rho e^{-\rho R(\theta)} b(\theta) \right] \\
& \quad + \mu \left[\frac{1}{R(\theta)^2} \int_0^{R(\theta)} y(t,\theta)\, dt - \frac{1}{R(\theta)} y(R(\theta),\theta) \right] \Bigg\} R'(\theta) \\
& - \mu \left\{ \left[\frac{1}{R(\theta)^2} \int_0^{R(\theta)} y(t,\theta)\, dt - \frac{1}{R(\theta)} y(R(\theta),\theta) \right] R'(\theta) + \frac{1}{R(\theta)} \int_0^{R(\theta)} y_\theta(t,\theta)\, dt \right\} =
\end{aligned}
$$

$$
\begin{aligned}
& \int_0^{\bar{T}} e^{-\rho t} u'(c(\theta)) c'(\theta)\, dt - e^{-\rho R(\theta)} \left[v\left(\frac{y(t,\theta)}{\varphi(t,\theta)} \right) + \eta(\theta) \right] R'(\theta) \\
& - \int_0^{R(\theta)} y_\theta(t,\theta) \left[e^{-\rho t} u'(c(\theta)) \left(1 - \mathcal{T}_y(t, y(t,\theta)) \right) + \mu \frac{1}{R(\theta)} \right] dt =
\end{aligned}
$$

$$
\begin{aligned}
& \int_0^{\bar{T}} e^{-\rho t} u'(c(\theta)) c'(\theta)\, dt - e^{-\rho R(\theta)} \left[v\left(\frac{y(t,\theta)}{\varphi(t,\theta)} \right) + \eta(\theta) \right] R'(\theta) \\
& - \int_0^{R(\theta)} e^{-\rho t} \frac{y_\theta(t,\theta)}{\varphi(t,\theta)} v'\left(\frac{y(t,\theta)}{\varphi(t,\theta)} \right) dt,
\end{aligned}
$$

where μ is the Lagrange multiplier on (31) and the last equality follows from the first-order conditions in (30). Note that the final expression above is the same as the local incentive compatibility of the constrained-efficient allocation and hence zero. Therefore, $\theta' = \theta$ is the local optimum of the optimization problem (33). Extensive tedious algebra available upon request shows that under certain conditions $\theta' = \theta$ is also a maximum in (33).

Now, we can construct the function $b(\cdot,\cdot)$. For each value of (R,Y), let $\tilde{b}(\theta)$ be defined by $\mathcal{V}\left(\tilde{b}(\theta),R,Y;\theta,\mathcal{T}\right) = \mathcal{V}(b(\theta),R(\theta),Y(\theta);\theta,\mathcal{T})$. Then, we let $b(R,Y) = \min_\theta \tilde{b}(\theta)$. Since $\mathcal{V}(b,R,Y;\theta,\mathcal{T})$ is increasing in b, we must have

$$\mathcal{V}(b(R,Y),R,Y;\theta,\mathcal{T}) \leq \mathcal{V}\left(\tilde{b}(\theta),R,Y;\theta,\mathcal{T}\right) = \mathcal{V}(b(\theta),R(\theta),Y(\theta);\theta,\mathcal{T})$$

Note that by Assumption 2, since $Y(\theta)$ is strictly increasing in θ, $b(\cdot,\cdot)$ is well-defined. This completes the proof. ■

B.4 Heterogeneous life span

The formal statement of the mechanism design problem discussed in Section 5.4 is given by the following problem:

$$\max \int_\Theta \left[\int_0^{\bar{T}(\theta)} e^{-\rho t} u(c(t,\theta))\,dt - \int_0^{R(\theta)} e^{-\rho t}\left[v(y(t,\theta)/\varphi(t,\theta)) + \eta(t,\theta)\right]dt\right] dG(\theta)$$

subject to feasibility

$$\int_\Theta \int_0^{\bar{T}(\theta)} e^{-\rho t} c(t,\theta)\,dt dF(\theta) + H \leq \int_\Theta \int_0^{R(\theta)} e^{-\rho t} y(t,\theta)\,dt dF(\theta)$$

and incentive compatibility

$$\int_0^{\bar{T}(\theta)} e^{-\rho t} u(c(t,\theta))\,dt - \int_0^{R(\theta)} e^{-\rho t}\left[v(y(t,\theta)/\varphi(t,\theta)) + \eta(t,\theta)\right]dt \geq$$

$$\int_0^{\bar{T}(\theta')} e^{-\rho t} u\left(c\left(t,\theta'\right)\right)dt - \int_0^{R(\theta')} e^{-\rho t}\left[v\left(y\left(t,\theta'\right)/\varphi(t,\theta)\right) + \eta(t,\theta)\right]dt, \forall \theta,\theta'$$

Closely following our baseline analysis we use the first-order approach to incentive compatibility and then numerically verify global incentive compatibility ex post.

B.5 Overlapping generations

We show here that the problem considered in Section 2 is equivalent to the steady state associated with a planning problem of an overlapping generations economy.

Time is continuous, $t \in \mathbb{R}_+ \cup \{0\}$, and at each point in time a generation with a unit mass is born. The individuals born at t live until $t + \overline{T}$. The heterogeneity within a generation is represented by θ, with distribution $F(\theta)$. As in the main text, the production function is given by

$$\mathcal{F}(K(t), L(t)) = rK(t) + L(t),$$

where $\mathcal{F}$ is net output (GDP net of depreciation) and r is net capital income (capital income net of depreciation). The preferences over sequences of consumption, labor supply, and retirement are given by

$$\int_0^{\overline{T}} e^{-\rho a} u(c(a))da - \int_0^{R(\theta)} e^{-\rho a} \left[v(l(a)) + \eta(a,\theta)\right] da$$

Allocations are given by

- individual allocations $c(t,a,\theta)$, $l(t,a,\theta)$, $R(\theta)$ for all $t \geq -\overline{T}$ and $a \in [\max\{t,0\}, \overline{T}]$, where t is the time of birth and a is the age of the individual (this includes the initial generations that were born prior to $t = 0$) and
- aggregate capital given by $K(t)$ and aggregate consumption and aggregate effective labor given by

$$\begin{aligned} C(t) &= \int_0^{\overline{T}} \int_\Theta c(t-a,a,\theta) dF(\theta) da \\ L(t) &= \int_0^{\overline{T}} \int_\Theta \varphi(a,\theta) l(t-a,a,\theta) \mathbf{1}[a \leq R(\theta)] dF(\theta) da \end{aligned}$$

Feasibility requires that

$$C(t) + \hat{H} + \dot{K}(t) = rK(t) + L(t),$$

where $\hat{H}$ is the value of output purchased by the government in each period and is constant over time. Assuming as in the main text that $r = \rho$, the above can be written in its present value form eliminating capital:

$$\int_0^\infty e^{-rt} C(t) dt + \frac{\hat{H}}{r} = rK(0) + \int_0^\infty e^{-rt} L(t) dt.$$

Disaggregate this into individual consumptions and use integration by parts to arrive at the following:

$$\int_{-\overline{T}}^\infty e^{-rt} \int_\Theta \int_{\max\{-t,0\}}^{\overline{T}} e^{-ra} c(t,a,\theta) da dF(\theta) dt + \frac{\hat{H}}{r} =$$
$$rK(0) + \int_{-\overline{T}}^\infty e^{-rt} \int_\Theta \int_{\max\{-t,0\}}^{\overline{T}} e^{-ra} \varphi(t,a,\theta) l(t,a,\theta) da dF(\theta) dt$$

Incentive compatibility is as it is defined in the main text.

The objective function for the planning problem is given by

$$\int_{-\overline{T}}^{\infty} e^{-\lambda t} \int_{\Theta} U(t,\theta) dG(\theta) dt, \tag{34}$$

where $U(t,\theta)$ is the utility of an individual born at t. When $t \geq 0$, this is given by

$$U(t,\theta) = \int_0^{\overline{T}} e^{-\rho t} u(c(t,a,\theta)) da - \int_0^{R(t,\theta)} e^{-\rho a} [v(l(t,a,\theta)) + \eta(a,\theta)] da,$$

while for individuals born at $t \leq 0$, this is given by

$$U(t,\theta) = \int_{-t}^{\overline{T}} e^{-\rho t} u(c(t,a,\theta)) da - \int_{-t}^{R(\theta)} e^{-\rho a} [v(l(t,a,\theta)) + \eta(a,\theta)] da$$

The parameter λ captures the intergenerational discount rate in the social welfare function. The planning problem is then to maximize the value of the objective in (34) over the set of feasible and incentive-compatible allocations. Note that since $r = \rho$, any solution to this problem must prescribe constant consumption over the life-cycle for each individual.

The planning problem in the main text, alternatively, maximizes the value for one generation subject to incentive compatibility and a feasibility constraint of the form

$$H + \int_0^{\bar{T}} \int_{\Theta} e^{-rt} c(t,\theta) dF(\theta) dt = \int_0^{\bar{T}} \int_{\Theta} e^{-rt} \varphi(t,\theta) l(t,\theta) \mathbf{1}[t \leq R(\theta)] dF(\theta) dt$$

Label the solution to this problem $\{c^*(t,\theta;H), l^*(t,\theta;H), R^*(\theta;H)\}$. We refer to the planning problem associated with a single generation as **P1** and the planning problem in the OLG economy as **P2**. Let the value of the social welfare in **P1** be given by $U^*(H)$.

For the steady state of **P2**, let the allocations be given by $\{c_{ss}(a,\theta), l_{ss}(a,\theta), R_{ss}(\theta)\}$. In addition, define H_{ss} as the difference between the present value of labor earnings and the present value of consumption in the steady state. Our main claim is that the steady state allocation coincides with the solution to **P1** for H_{ss}. The idea is that otherwise we can simply replace the steady state allocation with the solution to **P1** for generations that are born late in time.

Claim 1 *Given a unique solution to **P1**, the steady-state allocation of **P2** coincides with that of the one generation economy with H_{ss}, i.e.,*

$$c_{ss}(t,\theta) = c^*(t,\theta;H_{ss}), l_{ss}(t,\theta) = l^*(t,\theta;H_{ss}), R_{ss}(t,\theta) = R^*(t,\theta;H_{ss}).$$

Proof. Suppose not. Consider the solution to **P2**, and let $H(t)$ be the difference between the present value of labor earnings and consumption for generation t. Given the definition of the steady state of **P2**, it must be that $H(t)$ converges to H_{ss} as t tends to ∞. Since the solution to **P1** is unique and the steady-state allocation does not coincide with it, it must be that social welfare of the steady-state allocation is less than $U^*(H_{ss})$. Since $H(t)$ converges to H_{ss} and $U^*(H)$ is continuous in H, for large enough t , the aggregate welfare of generations born at t is less than $U^*(H_t)$. Replace these allocations with the solution of **P1** for $H = H(t)$. These allocations are feasible since the difference in present value of consumption and labor earnings remains constant. Furthermore, they are incentive compatible and they deliver a higher value of social welfare. This implies that the initial allocation cannot be a solution to **P2**. ■

C Further details for Sections 4 and 5

We expand here on the main text to provide further details of the sample construction, the estimation, and the sensitivity checks of the results.

Data sets. We use two sources of longitudinal individual-level data, the HRS and the PSID. To work with the HRS, data we use version K of RAND HRS files, which are based on all publicly available surveys from 1992 to 2008. The details of cleaning, processing, and consolidating raw HRS variables are extensively documented in RAND HRS version K documentation available online from RAND Corporation.

Raw data files of the PSID contain a well-known range of inconsistencies, impossible answers, and other issues. We use the data set from Heathcote, Perri, and Violante (2010), who carefully address these issues, and refer to that paper for details. We use Sample A of Heathcote, Perri, and Violante (2010), which is their most inclusive sample and is essentially a cleaned version of the data that come from the Survey Research Center (SRC) sample of the PSID using all of the annual surveys from 1967 to 1996 and the biennial surveys for 1999, 2001, and 2003. One issue particularly well-known in the literature, potentially important for our purposes, is top coding of incomes in the PSID. Heathcote, Perri, and Violante (2010) address it by fitting a Pareto distribution in the right tail. For our baseline we take this data with the fitted Pareto tail. Among the robustness exercises below, we check the effects of removing the fitted thick right tail to make sure it does not introduce artificial qualitative features into our results.

Sample selection. As a baseline we use males of one U.S. cohort referred to as 1940 cohort. It includes males born between 1931 and 1941. This coincides with the initial HRS

Table 3: Distribution of annual hours worked, by age.

Age	Annual hours					
	[0,500)	[500,1000)	[1000,1500)	[1500,2000)	$\geq$2000	hours/worker
60	0.03	0.04	0.06	0.17	0.70	2150
61	0.04	0.03	0.06	0.17	0.66	2124
62	0.08	0.05	0.07	0.18	0.57	2015
63	0.10	0.09	0.08	0.13	0.52	1909
64	0.12	0.09	0.12	0.12	0.45	1789
65	0.11	0.10	0.14	0.14	0.43	1777
66	0.16	0.09	0.16	0.13	0.33	1642
67	0.25	0.08	0.12	0.08	0.25	1605
68	0.26	0.08	0.14	0.09	0.20	1539
69	0.28	0.08	0.11	0.09	0.19	1538
70	0.24	0.12	0.15	0.09	0.19	1418

Note: 1940-cohort males in the pooled sample of the HRS and the PSID.

cohort, which was first interviewed in 1992 and subsequently every two years, providing the longest observed cohort over time in the HRS. This also implies that the cohort have approached age 60 by the year 2000 and hence we use a stylized version of the U.S. Social Security system from 2000.

This leads us to start with 9,638 individuals in the PSID and 15,959 in the HRS. After we restrict the age to at least 20 and gender to males, we obtain in the PSID sample 6,918 and in the HRS sample 3,656 individuals. When we restrict observations per individual to be at least 23 in the PSID data (later we check sensitivity by relaxing this restriction), we are left with a sample of 1,116 individuals. We restrict observations per individual to at least 5 in the HRS to get 971 individuals. The final count of individuals in the pooled HRS-PSID data set is 2,087.

In terms of observations, we start with 93,924 in the PSID and 56,667 in the HRS. After restricting the age and gender, we obtain 77,157 observations in the PSID and 26,602 observations in the HRS. Restricting observations to at least 23 per individual in the PSID leaves 30,751 observations; restricting observations to at least 5 per individual in the HRS leaves 5,788 observations. The final count of observations in the pooled HRS-PSID data set is 36,539, providing on average 18 observations per individual.

Life span, retirement ages, and Social Security claiming ages. Each individual enters the quantitative environment at age 20. We need three additional ages for each individual: the age when the individual claims Social Security benefits, $S(\theta)$; when the individual retires, $R(\theta)$; and when the individual dies, $T(\theta)$.

As a baseline, we take $\bar{T}$ to be 81.6 from the Social Security Administration's Life

Table 4: Distributions of retirement ages, benefit claiming ages, and life spans, by earnings decile.

	Earnings percentile		
	Bottom decile	Median	Top decile
Retirement age by definition:			
baseline	67.6	66.9	68.5
alternative	66.3	66.1	66.0
Retirement age by sector:			
manufacturing and mining	67.2	66.3	64.8
retail	67.8	70.3	68.0
professional services	70.1	66.8	70.5
Retirement age by education:			
less than high school	67.7	67.0	67.5
high school and above	67.1	66.8	68.7
Social Security claiming age	62.6	63.4	63.5
Life expectancy at age 60	78.3	81.4	83.8

Note: 1940-cohort males in the pooled sample of the HRS and the PSID. Percentiles are constructed as in the main text.

Tables in Bell and Miller (2005) for 1940-born males (see also our discussion above of the comparison with the steady-state of an overlapping generations economy). Since the retirement behavior is part of the focus here, we take the life span for the individuals who survived at least until age 60.

We obtain retirement ages $R(\theta)$ by using two definitions, a baseline and an alternative. Our baseline definition of retirement uses the RAND HRS variable *RwLBRF*, which aims at consolidating all available in the HRS sources of information about the individual's labor-force status. The consolidation uses a variety of questions, depending on the wave, listed in the online RAND files documentation. The evidence of working hence comes from multiple sources reconciled and summarized in *RwLBRF*. Of importance here is that this consolidation aims at separating retirement from claiming pension benefits from Social Security, unemployment, partial retirement, or claiming to be retired while also reporting labor earnings.

We first identify the first and last non-missing labor-force status values for an individual according to *RwLBRF*. Then, within this range, only if we observe a change in status from a non-missing non-5 value to the value of 5 in two consecutive waves (to retired from any other status), we define the person as potentially retired in the latter wave. Finally, the maximum among potential retirement ages is used as the baseline retirement age.

We check the implications of alternative definitions of retirement in the data. Retire-

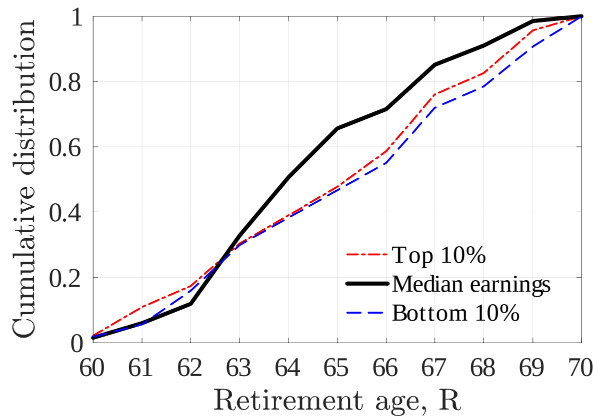

Figure 13: Cumulative retirement age distribution for selecred percentiles of lifetime earnings.

ment in our setting means no work hours (excluding unemployment) from a given age onward. As a check, we use a definition based purely on hours worked reported in the PSID and the HRS, referred to as the alternative definition in the main text. Following Guvenen (2009), we define individuals as retired if hours worked fall below 520 permanently (i.e. 10% of 5,200 hours, the likely highest sustainable annual hours according to Guvenen (2009)). The Pearson correlation coefficient between retirement age according to the baseline definition and this alternative definition is 0.67. The average retirement ages by the two definitions in the pooled sample are much less than a standard deviation apart.

The individual information about Social Security benefits claiming ages $S(\theta)$ (as opposed to retirement ages) is summarized in the RAND HRS in the variable *RASSAGEB*. We adopt it as the baseline definition of Social Security claiming age in the quantitative analysis.

Distribution of hours and retirement ages. Here we take a more detailed look at the behavior of hours worked around the time of retirement and then describe how the retirement ages vary with the definition of retirement, with education, by sector, and with labor earnings.

First, we examine the labor supply behavior around retirement to see if in our sample retirement appears to be associated with smooth transitions or with more abrupt changes from full-time work to not working. For instance, Rogerson and Wallenius (2013) show evidence of quite abrupt transitions for the CPS and the PSID separately as well as review the evidence in the literature based on the HRS. Table 3 suggests similar behavior in our pooled sample of the HRS and the PSID. At age 60, by far the most common average amount of hours worked per year is greater than 2,000. Already starting at age 67 and

Table 5: Statistical models of productivity-age profiles.

Parameter	Model 1	Model 2	Model 3
a	0.126171 (0.0309)*	0.213422 (0.0732)*	0.128442
β_2	-0.00172 (0.000233)*	-0.00214 (0.000263)*	-0.00172
β_3	7.267x10^{-6} (1.723x10^{-6})*	0.00001 (1.911x10^{-6})*	6.5403x10^{-6}
Individuals:	2,087	2,087	N/A
Observations:	36,539	22,822	N/A

Note: * Indicates significance at the 1 percent level.

quite clearly by age 70, the most common hours worked are less than 500, with an increase in that group from 16% to 25% at age 67.

Table 4 summarizes how the distribution of retirement ages changes with the definition of retirement, with education, and by sector. The baseline and the alternative definitions of retirement result in essentially similar retirement age patterns. More physically demanding sectors, such as manufacturing and mining, deviate from the general pattern, with retirement ages that decline with higher earnings. Individuals in less physically demanding sectors tend to retire at older ages, both at the bottom and at the top of the earnings distribution. More educated individuals tend to retire at younger ages throughout the distribution of earnings. Using the baseline definition of retirement, Figure 13 compares the cumulative distribution of retirement ages (retirement hazard ratios) of the median earnings decile with that of the top and the bottom deciles, summarizing the overall distributional differences.

Productivity-age profiles. The statistical model described in the main text can be more explicitly written as

$$w_{it} = \left[\beta_0 + \beta_1 x_{it} + \beta_2 x_{it}^2 + \beta_3 x_{it}^3\right] + \sum_{j=2}^{I} \gamma_{0i}\left(d_j\right) + \sum_{j=2}^{I} \gamma_{1i}\left(d_j x_{it}\right) + \varepsilon_{it},$$

where w_{it} is the logarithm of effective reported labor earnings per hour for person $i = 1, .., I$ at time t, x_{it} is a measure of experience, taken to be age in the baseline, so that the term in brackets is the logarithm of the common age component, d_i are individual dummies, and the constraints implied by our parametric assumption is given by $\beta_1 = a\beta_0$ and $\gamma_{1i} = a\gamma_{0i}$ for $i = 2, .., I$. That is, we estimate non-linear equations

$$w_{it} = \left[(1 + a x_{it})\,\beta_0 + \beta_2 x_{it}^2 + \beta_3 x_{it}^3\right] + (1 + a x_{it}) \sum_{i=2}^{I} \gamma_{0i}\left(d_i\right) + \varepsilon_{it},$$

Table 6: Summary statistics for earnings deciles used in constructing productivity-age profiles.

Earnings Decile	Average Decile Statistic			
	Annual Hours	Education (years)	Married (fraction)	Caucasian (fraction)
1st (bottom)	2227.19	11.0819672	0.85	0.85
2nd	2274.84	11.7015504	0.89	0.82
3rd	2217.93	12.0788382	0.86	0.89
4th	2304.49	12.4729730	0.93	0.92
5th	2237.65	12.8357488	0.92	0.91
6th	2175.42	13.2620321	0.88	0.93
7th	2123.48	13.6042781	0.91	0.93
8th	2160.45	14.0960452	0.88	0.94
9th	2066.72	14.5444444	0.82	0.94
10th (top)	2077.16	15.5217391	0.86	0.97

Note: deciles are based on the predictions from the baseline estimation.

where $\{a, \beta_0, \beta_2, \beta_3, \gamma_{0i}\}$ are estimated using GMM, reported in the Model 1 column in Table 5.

Given estimates $\left\{\hat{a}, \hat{\beta}_0, \hat{\beta}_2, \hat{\beta}_3, \{\hat{\gamma}_{0i}\}_{i=2}^{I}\right\}$, we construct a vector of estimates $\{\log \hat{\theta}_i\}_{i=1}^{I}$ defined as

$$\{\hat{\beta}_0, \hat{\beta}_0 + \hat{\gamma}_{02}, \hat{\beta}_0 + \hat{\gamma}_{03}, \cdots\}$$

so that the predicted log productivity-age profiles are given by

$$\log \hat{\varphi}_t^n = (1 + \hat{a}x_t) \log \bar{\theta}_i^n + \hat{\beta}_2 x_t^2 + \hat{\beta}_3 x_t^3$$

where $\log \bar{\theta}_i^n$ is the average of $\log \hat{\theta}_i$'s for a group $n = 1..N$ and $x_t = 20..80$ is age. That is, the individual fixed effects are interpreted as individual type and $\log \varphi$ is proxied with the logarithm of effective labor earnings per hour, i.e., the computed ratio of all labor earnings to total hours reported in the PSID and the HRS. The labor earnings are taken directly from our pooled sample variables, converting to constant 2000 dollars. We aim to capture total labor earnings summing over the variables containing "salaries and wages", "separate bonuses", "the labor portion of business income", "overtime pay", "tips", "commissions", "professional practice or trade payments", and "miscellaneous labor income". We similarly capture the total reported hours.

For our baseline, we set $N = 10$ where each group is defined to be a decile, later changing the number of groups to $N = 5$ and $N = 20$. Overall, the Pearson correlation coefficient between estimated type $\hat{\theta}$ and the average annual earnings is 0.85. The same correlation with average annual earnings after age 60 is 0.80. To a first approximation,

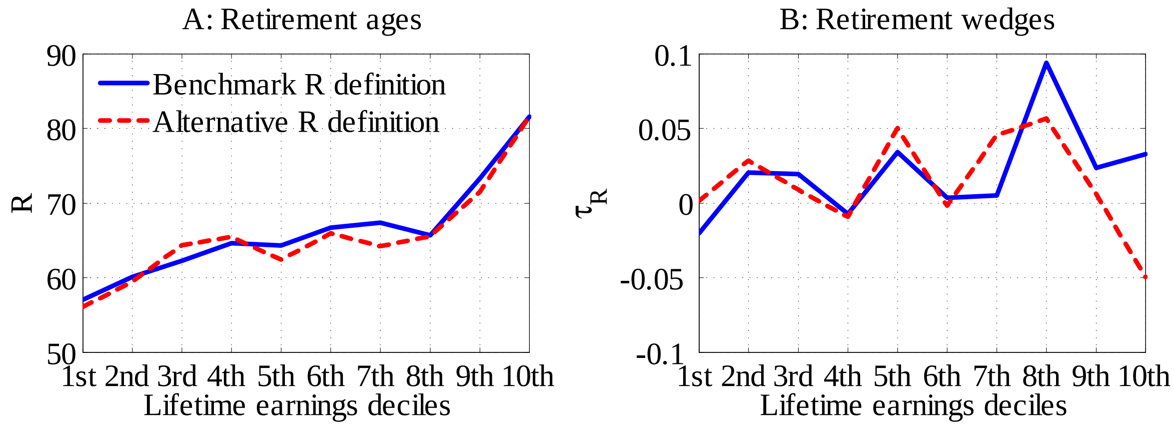

Figure 14: The effects of the variation in the definition of retirement on efficient retirement ages as a function of lifetime earnings (Panel A) and their associated optimal retirement wedges (Panel B).

the deciles are also referred to as average annual-earnings deciles. Table 6 reports the analogues of the summary statistics in the main text for each of the baseline deciles.

Estimation sensitivity checks. We implement several variants of the above estimation approach resulting in qualitatively virtually indistinguishable simulations from those reported in the main text. First, we varied the number of groups to $N = 5$ and $N = 20$ with virtually identical results. Then, instead of age, we also used two alternative measures of experience x_{it}:

$$x_{it}^1 = age_{it} - 19$$

and

$$x_{it}^2 = age_{it} - \max\{edu_{it}, 10\} - 5.$$

We find that replacing age with a measure of experience accounts for the rightward shift in the peak of the profile as the type increases, as is well-known in labor literature. That is, using x_{it}^1 instead of age produces very similar productivity profiles to the ones reported in the main text, except the peaks of the profiles of different types are much closer to each other, with x_{it}^2 resulting in virtual alignment. The latter variant is also very close to the profiles we obtained by closely following the approach of Nishiyama and Smetters (2007) and grouping individual observations by type into one of seven bins, each for a 10-year interval of ages – 25-35 years old, 34-45 years old, ..., 74-85 years old (the few remaining individuals older than 85 were put in the last group) – and extrapolating by using shape-preserving cubic splines to obtain the complete productivity-age profiles.

As another check of the sensitivity of the estimation results, we relaxed the require-

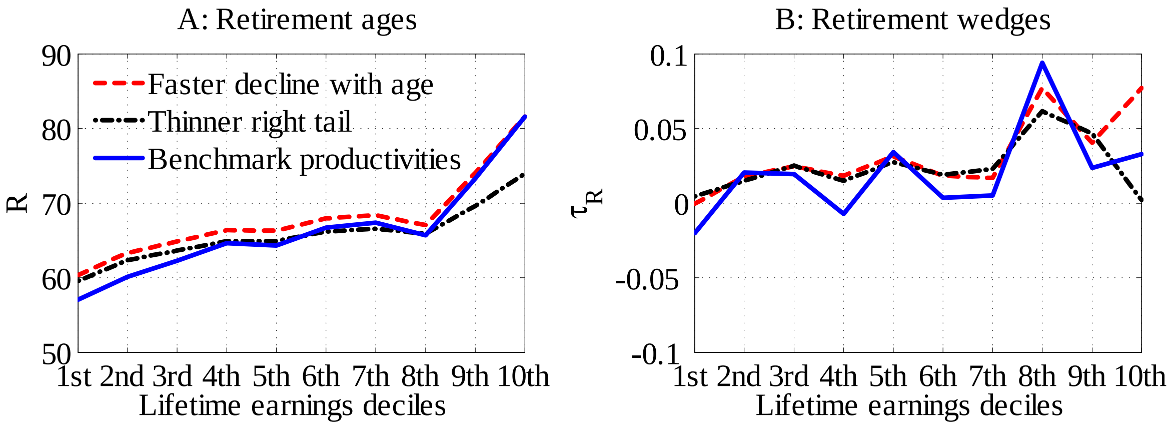

Figure 15: The effects of the variation in the estimates of the productivity-age profiles on efficient retirement ages as a function of lifetime earnings (Panel A) and their associated optimal retirement wedges (Panel B).

ment that in the pooled sample the individuals from PSID have 23 observations, and replaced it with 5 observation requirement, as used for the HRS individuals. To keep the GMM estimation computationally feasible, we randomly selected the same number of individuals as in the original sample so that the total pooled-sample size remained the same. The resulting estimates are reported in Table 5 in the Model 2 column. The estimates of the baseline Model 1 are within two standard deviations from Model 2.

Finally, given a debate in the literature about how much of the curvature in the profiles is driven by the parametric choices, especially at later ages, we study the effects on the constrained optimum of the following check: we force the curvature to vary by exogenously increasing parameter a by 20 percent and reducing parameter β_3 by 10 percent. The resulting parameters are reported in Table 5 in the Model 3 column and the constrained optimum is simulated below with very similar results to the baseline.

Further robustness results. In addition to the analysis of the sensitivity of the normative simulation results in the main text, we provide here three further sets of deviations from the baseline simulation. First, we show the robustness of the main normative simulation results with respect to the definition of retirement ages in the data. As there is no single universally accepted definition in the literature, we adopted two definitions we judged to be potentially as different as the data would allow. The definitions are discussed in the main text, with additional details above. Figure 14 illustrates the resulting comparison to the baseline simulated constrained optimum. As the main text explains, it is reasonable to focus on the profiles of retirement ages and on optimal retirement wedges since the rest of the constrained optimum follows. Figure 14 reveals qualitatively only mi-

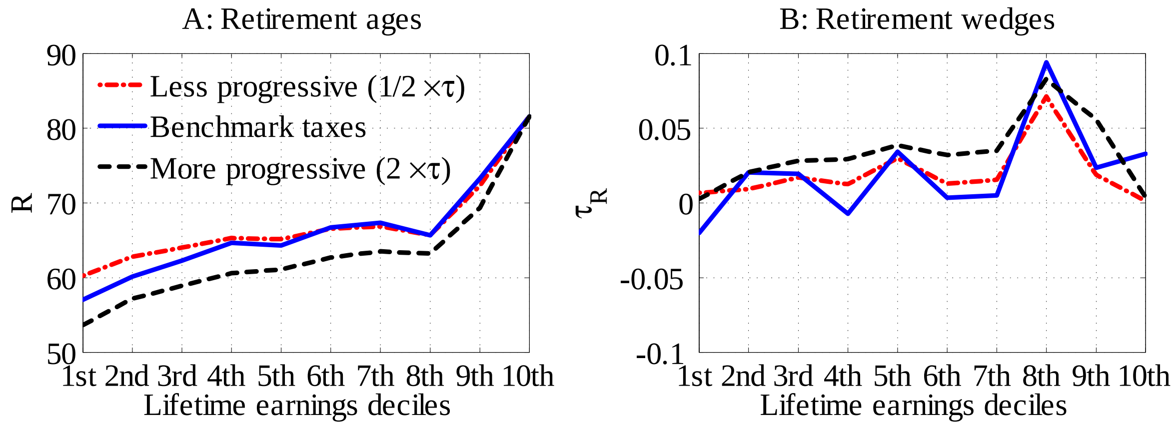

Figure 16: The effects of the variation in the estimates of the effective status-quo policies on efficient retirement ages as a function of lifetime earnings (Panel A) and their associated optimal retirement wedges (Panel B).

nor differences that would lead one to make the same conclusions as in the main text.

Second, a potentially key deviation from the baseline simulation results may be introduced by the parametric restrictions on the productivity-age profiles as we discussed above. To explore these effects, we varied our assumptions in the estimation as discussed in the previous section. Figure 15 compares the baseline simulation of the optimum to two of the most extreme cases we found. Once again the differences appear to be quantitatively minimal.

Finally, we explore how robust our main quantitative insights are with respect to the estimates of the effective tax functions. Even though the estimates in the main text follow the literature, they are at the core of the estimated fixed costs and thus it is instructive to illustrate how sensitive the simulated constrained optimum is to possible errors in the estimates. Figure 16 compares the results from the baseline calibration to a calibration based on a tax function with the progressivity parameter arbitrarily forced to be twice as large and half as large as the estimate from the literature that we used in the baseline and discussed in the main text. As with all of the cases above, this results in quantitatively minimal changes and suggests the same general conclusions we discussed in the main text.