Appendix: Redistributive effects of different pension systems when longevity varies by socioeconomic status

by Miguel Sanchez-Romero, Ronald D. Lee, Alexia Prskawetz

A Demographics

Individuals. Let the probability of surviving from birth to age x of an individual belonging to income group $i \in \mathcal{I} = \{1, 2, \ldots, I\}$ be

$$p_i(x) = e^{-\int_0^x \mu_i(t)dt}, \tag{A.1}$$

with $p_i(0) = 1$, $p_i(\omega) = 0$, $\omega \in (0, \infty)$ denotes the maximum age, and $\mu_i(t) \geq 0$ is the mortality hazard rate at age t of an individual of group i. The life expectancy at age x of an individual belonging to income group i is defined as

$$\mathsf{e}_i(x) = \int_x^\omega \frac{p_i(t)}{p_i(x)} dt. \tag{A.2}$$

Population. Given that all income groups grow steadily at the same rate n, the total population size at time t is

$$P(t) = B(t) \int_0^\omega e^{-nx} p(x) dx, \text{ with } p(x) = \frac{1}{I} \sum\nolimits_{i \in \mathcal{I}} p_i(x), \tag{A.3}$$

where $B(t)$ is the total number of births at time t, $p(x)$ is the average survival, which implies that the average mortality hazard rate at age x, denoted by $\mu(x)$, is $\sum_{i \in \mathcal{I}} \mu_i(x) p_i(x) / \sum_{i \in \mathcal{I}} p_i(x)$. The existence of a positive relationship between the income group and the life expectancy —see Table 1— implies that the average mortality hazard rate is biased with age towards the mortality hazard rate of higher income individuals.

B Derivation of Eq. (6)

Assuming that an individual will retire at age R_i, we define the social security wealth at age $S_i < t \leq R_i$ of an individual of type i, denoted by $\mathsf{SSW}_i(t)$, as

$$\mathsf{SSW}_i(t) = b_i(R_i) \int_{R_i}^\omega e^{-\int_t^x r + \mu_i(j) dj} dx - \int_t^{R_i} e^{-\int_t^x r + \mu_i(j) dj} \tau y_i(x) dx. \tag{B.1}$$

The first component of (B.1) is the present value at age t of the survival weighted stream of future benefits during retirement, while the second component of (B.1) is the present value at age t of the survival weighted remaining pension contributions to pay until retirement.

Eq. (6) can easily be derived from (1)–(4) by elementary algebraic manipulations of (B.1). Integrating Eq. (1) from S_i until retirement R_i gives

$$\mathsf{pp}_i(R_i) = \int_{S_i}^{R_i} e^{\int_x^{R_i} \tilde{\mathsf{r}}+\tilde{\mu}(j)dj} \phi\tau y_i(x)dx. \tag{B.2}$$

Splitting in two the integral (B.2) gives

$$\begin{aligned}\mathsf{pp}_i(R_i) &= \int_{S_i}^{t} e^{\int_x^{R_i} \tilde{\mathsf{r}}+\tilde{\mu}(j)dj} \phi\tau y_i(x)dx + \int_{t}^{R_i} e^{\int_x^{R_i} \tilde{\mathsf{r}}+\tilde{\mu}(j)dj} \phi\tau y_i(x)dx = \\ &= e^{\int_t^{R_i} \tilde{\mathsf{r}}+\tilde{\mu}(j)dj} \int_{S_i}^{t} e^{\int_x^{t} \tilde{\mathsf{r}}+\tilde{\mu}(j)dj} \phi\tau y_i(x)dx + \int_{t}^{R_i} e^{\int_x^{R_i} \tilde{\mathsf{r}}+\tilde{\mu}(j)dj} \phi\tau y_i(x)dx. \end{aligned} \tag{B.3}$$

Substituting (2) in (B.1), using the definition of $A_i(R_i, r)$, and rearranging terms, gives

$$\mathsf{SSW}_i(t) = e^{-\int_t^{R_i} r+\mu_i(j)dj} \mathsf{f}_i(R_i)A_i(R_i, r)\mathsf{pp}_i(R_i) - \int_t^{R_i} e^{-\int_t^{x} r+\mu_i(j)dj} \tau y_i(x)dx. \tag{B.4}$$

By substituting (B.3) in (B.4) and after some manipulations we have

$$\begin{aligned}\mathsf{SSW}_i(t) &= \mathsf{f}_i(R_i)A_i(R_i, r)e^{\int_t^{R_i} \tilde{\mathsf{r}}+\tilde{\mu}(j)-(r+\mu_i(j))dj} \int_{S_i}^{t} e^{\int_x^{t} \tilde{\mathsf{r}}+\tilde{\mu}(j)dj} \phi\tau y_i(x)dx \\ &+ e^{-\int_t^{R_i} r+\mu_i(j)dj} \phi\mathsf{f}_i(R_i)A_i(R_i, r) \int_t^{R_i} e^{\int_x^{R_i} \tilde{\mathsf{r}}+\tilde{\mu}(j)dj} \phi\tau y_i(x)dx \\ &- \int_t^{R_i} e^{-\int_t^{x} r+\mu_i(j)dj} \tau y_i(x)dx. \end{aligned} \tag{B.5}$$

By multiplying and dividing by ϕ the first term on the right-hand side of (B.5) we obtain that the first term of (B.5) is $\mathsf{P}_i(t)$ times the total value of all the contributions paid until age t, $\mathsf{pp}_i(t)/\phi$. Moreover, we know from (5) that $\mathsf{P}_i(t)e^{-\int_t^{R_i} \tilde{\mathsf{r}}+\tilde{\mu}(j)dj} = \phi\mathsf{f}_i(R_i)A_i(R_i, r)e^{-\int_t^{R_i} r+\mu_i(j)dj}$. Thus,

$$\begin{aligned}\mathsf{SSW}_i(t) &= \mathsf{P}_i(t)\frac{\mathsf{pp}_i(t)}{\phi} + \mathsf{P}_i(t)e^{-\int_t^{R_i} \tilde{\mathsf{r}}+\tilde{\mu}(j)dj} \int_t^{R_i} e^{\int_x^{R_i} \tilde{\mathsf{r}}+\tilde{\mu}(j)dj} \tau y_i(x)dx \\ &- \int_t^{R_i} e^{-\int_t^{x} r+\mu_i(j)dj} \tau y_i(x)dx. \end{aligned} \tag{B.6}$$

Taking into account that $\frac{\mathsf{P}_i(t)}{\mathsf{P}_i(x)} = e^{\int_t^{x} \tilde{\mathsf{r}}+\tilde{\mu}(j)-(r+\mu_i(j))dj}$, the social security wealth at time t in (B.6) can also be written as

$$\mathsf{SSW}_i(t) = \mathsf{P}_i(t)\frac{\mathsf{pp}_i(t)}{\phi} - \int_t^{R_i} e^{-\int_t^{x} r+\mu_i(j)dj} \tau(1 - \mathsf{P}_i(x))y_i(x)dx. \tag{B.7}$$

The first component of (B.7) is the monetary value given to the stream of contributions until age t, while the second component of (B.7) is the present value, survival weighted, at age t of the stream of future implicit taxes/subsidies on labor income.

By differentiating (B.7) with respect to t we have

$$\frac{\partial \mathsf{SSW}_i(t)}{\partial t} = \left(\tilde{\mathsf{r}} + \tilde{\mu}(t) + \frac{1}{\mathsf{P}_i(t)}\frac{\partial \mathsf{P}_i(t)}{\partial t}\right)\mathsf{P}_i(t)\frac{\mathsf{pp}_i(t)}{\phi} - (r + \mu_i(t))\int_t^{R_i} e^{-\int_t^x r+\mu_i(j)dj}\tau(1 - \mathsf{P}(x))y_i(x)dx + \tau y_i(t). \qquad \text{(B.8)}$$

Finally, using (7) in the second term of (B.8) and the definition of SSW in (B.7) we obtain Eq. (6).

C Parametrization

We impose the following set of assumptions with respect to the economic variables. First, we assume a risk-free market discount factor (r) of 3%. This market interest rate coincides with that assumed in the report by the National Academies of Science (NASEM, 2015). Second, the population is assumed to grow at an annual constant rate (n) of 0.5% and the growth rate of labor productivity (g) is set at 1.5% per year, corresponding closely to the US case. Third, the annual capitalization factor of the unfunded pension system ($\tilde{\mathsf{r}}$) is set at 2%(=$n + g$), which is lower than the market discount factor. Therefore, since a return of 1%(=3%-2%) is lost annually, contributions to the pension system are implicitly considered by individuals as a tax on labor income. From (7) we know that this assumption implies that the marginal value of a dollar contributed to the pension system is an increasing function with respect to age and, as a consequence, all pension systems will provide an incentive to supply more labor early in the working life and to reduce labor before retirement. Moreover, since all pension systems have a similar increase in $\overline{\mathcal{P}}_i$, pension systems will have a similar direct impact on the length of schooling. Fourth, unless otherwise indicated, we assume that the social security system uses the average survival probability to calculate the pension benefits; i.e, $\tilde{p}(x) := p(x) = \sum_{i \in \mathcal{I}} p_i(x)/I$. Fifth, we assume for the NDC systems a minimum retirement age of 55 and a maximum retirement age of 70 for all $i \in \mathcal{I}$. For the DB systems, we restrict the minimum retirement age to 62 and the maximum to 70, similar to the US pension system. Sixth, the social security budget is balanced[11]

$$\sum_{i \in \mathcal{I}} \int_{S_i}^{R_i} e^{-nt} p_i(t)\tau w_i(S_i, t)\ell_i(t)dt == \sum_{i \in \mathcal{I}} e^{-nR_i} \int_{R_i}^{\omega} e^{-(n+g)(t-R_i)} p_i(t) b_i(R_i)dt. \qquad \text{(C.1)}$$

Using Eq. (C.1) we adjust in the DB systems the social contribution rate in order to support all pension benefits claimed by the surviving retirees, while in the NDC systems we adjust the overall pension replacement rate, or generosity of the pension system. For the sake of comparison across the alternative pension systems, we use for all the NDC systems the social contribution rate

[11]Note in Eq. (C.1) that similar to the US pension system, we assume that pension benefits are held constant (in real terms) after retirement and thus they do not increase with productivity.

obtained for the US pension system (DB-II). In particular, we obtain that the necessary social contribution rate to balance the US pension system with our assumed population structure is 11.83%, under the hypothetical assumption that the population faces the survival probabilities of the cohort born in 1930. While τ must be set at 13.28% in the case of using the survival probabilities of the cohort born in 1960.

In addition, we assume all individuals have similar preferences, except for the disutility of labor (α_i) that is specific to the income quintile of the individual. This assumption reflects the fact that individuals in different quintiles have different health and labor market trajectories. The instantaneous utility of consumption is assumed to be logarithmic ($\sigma_c = 1$), as found empirically by Chetty (2006), the intertemporal elasticity of substitution on labor (σ_ℓ) is set at 0.33, so that workers supply on average thirty five percent of their available time for labor (excluding sleep time), and the subjective discount factor (ρ) is set at 0.005. As a result, the cross-sectional consumption profile increases with age by one percent, similar to the consumption pattern reported in the NTA accounts for the US in 2003 (see `www.ntaccounts.org`). The marginal utility of leisure during retirement $\varphi(\cdot)$ is assumed to be constant across income groups and birth cohorts and monotonically increasing with age. Thus, we consider the marginal utility of leisure to be a function of the average life expectancy of the 1930 birth cohort; i.e $\varphi(t) = \varphi_0 \left(\mathsf{e}(t)\right)^{-\varphi_1}$ with $\varphi_0, \varphi_1 > 0$. To match the wage rate per unit of human capital for the cohort born in 1930, we take the parameters of the Mincerian equation reported in Table 2 in Heckman et al. (2006). Nevertheless, the parameter β_0 is adjusted in order to take into account the effect of productivity growth. As in Cervellati and Sunde (2013) we fix the returns to scale in education (γ) at 0.65. Following Sánchez-Romero et al. (2016) we set the disutility of schooling (η) at 3.5 for all income groups, which corresponds to the most likely value of η for an average return to schooling between 11 and 12 percent given the life expectancy of US males born in 1930. The learning ability (θ_i) for each income quintile group is calibrated to replicate the length of schooling and the retirement age from the Health and Retirement Survey (HRS).[12] Finally, the weight of disability cost (α_i) is used to replicate the present value of lifetime benefits (PVB) at age 50 reported in NASEM (2015) for the cohort born in 1930. See Table 8 in Appendix E.

D Economic problem

We solve the problem of maximizing the lifetime utility (11) subject to the constraints (1)–(10) using the Hamiltonian before age S_i, during the working period (S_i, R_i), and after the retirement age R_i, or periods 1, 2, and 3, respectively (Tomiyama, 1985). We denote with the letter $\mathcal{H}^j$ the

[12]Data from the HRS on length of schooling and retirement age for males born in 1930 was provided by Arda Aktas and Miguel Poblete-Cazenave.

Hamiltonian associated to period $j = \{1, 2, 3\}$.

Period 1. Given a length of schooling S_i and retirement age R_i, the Hamiltonian of an individual type $i \in \mathfrak{I}$ before working ($t \leq S_i$) is defined as

$$\begin{aligned}\mathcal{H}^1 &= e^{-\int_{x_0}^{t} \rho+\mu_i(j)dj} \left[U(c_i(t)) - \eta\right] + \lambda_a^1(t) \left[(r + \mu_i(t))a_i(t) - c_i(t)\right] + \\ &+ \lambda_h[\theta_i h_i(t)^\gamma - \delta h_i(t)] \end{aligned} \tag{D.1}$$

where $\lambda_a^1(t)$ and $\lambda_h(t)$ are the costate or adjoint variables associated to the dynamics of each state variable $\{a_i(t), h_i(t)\}$ for period 1. The first-order conditions (FOCs) for an interior consumption is:

$$\frac{\partial \mathcal{H}^1}{\partial c_i} = e^{-\int_{x_0}^{t} \rho+\mu_i(j)dj} U'(c_i(t)) - \lambda_a^1(t) = 0. \tag{D.2}$$

The dynamic laws of the costate or adjoint variables are:

$$\frac{\partial \lambda_a^1(t)}{\partial t} = -\lambda_1^1(t)(r + \mu_i(t)), \tag{D.3}$$

$$\frac{\partial \lambda_h^1(t)}{\partial t} = -\lambda_h^1(t)(\gamma \theta_i h_i(t)^{\gamma-1} - \delta), \tag{D.4}$$

Period 2. Given a length of schooling S_i and a retirement age R_i, the Hamiltonian of an individual type $i \in \mathfrak{I}$ during the working period ($S_i < t < R_i$) is defined as

$$\begin{aligned}\mathcal{H}^2 &= e^{-\int_{x_0}^{t} \rho+\mu_i(j)dj} \left[U(c_i(t)) - \alpha_i v(\ell_i(t))\right] + \\ &+ \lambda_a^2(t) \left[(r + \mu_i(t))a_i(t) + (1 - \tau)h_i(S_i)\bar{w}(t - S_i)\ell_i(t) - c_i(t)\right] + \\ &+ \lambda_p^2(t) \left[(\tilde{\mathsf{r}} + \tilde{\mu}(t))\mathsf{pp}_i(t) + \phi\tau h_i(S_i)\bar{w}(t - S_i)\ell_i(t)\right], \end{aligned} \tag{D.5}$$

where $\lambda_a^2(t)$, $\lambda_h^2(t)$, and $\lambda_p^2(t)$ are the costate or adjoint variables associated to the dynamics of each state variable $\{a_i(t), h_i(S_i), \mathsf{pp}_i(t)\}$ for period 2. The first-order conditions (FOCs) for an interior consumption and hours worked are:

$$\frac{\partial \mathcal{H}^2}{\partial c_i} = e^{-\int_{x_0}^{t} \rho+\mu_i(j)dj} U'(c_i(t)) - \lambda_a^2(t) = 0, \tag{D.6}$$

$$\begin{aligned}\frac{\partial \mathcal{H}^2}{\partial \ell_i} &= -e^{-\int_{x_0}^{t} \rho+\mu_i(j)dj} \alpha_i v'(\ell_i(t)) + \lambda_a^2(t)(1 - \tau)h_i(S_i)\bar{w}(t - S_i) + \\ &+ \lambda_p^2(t)\phi\tau h_i(S_i)\bar{w}(t - S_i) = 0. \end{aligned} \tag{D.7}$$

The dynamic laws of the costate or adjoint variables are:

$$\frac{\partial \lambda_a^2(t)}{\partial t} = -\lambda_a^2(t)(r + \mu_i(t)), \tag{D.8}$$
$$\frac{\partial \lambda_h^2(t)}{\partial t} = -\lambda_a^2(t)(1-\tau)\bar{w}(t - S_i)\ell_i(t) - \lambda_p^2(t)\phi\tau\bar{w}(t - S_i)\ell_i(t), \tag{D.9}$$
$$\frac{\partial \lambda_p^2(t)}{\partial t} = -\lambda_p^2(t)(\tilde{\mathsf{r}} + \tilde{\mu}(t)). \tag{D.10}$$

Period 3. Given a length of schooling S_i and a retirement age R_i, the Hamiltonian of an individual type $i \in \mathfrak{I}$ during retirement ($t \geq R_i$) is defined as

$$\begin{aligned} \mathcal{H}^3 &= e^{-\int_{x_0}^{t} \rho+\mu_i(j)dj} \left[U(c_i(t)) + \varphi(t)\right] \\ &\quad + \lambda_a^3(t) \left[(r + \mu_i(t))a_i(t) + \mathsf{f}_i(R_i)\mathsf{pp}_i(R_i) - c_i(t)\right], \end{aligned} \tag{D.11}$$

where $\lambda_a^3(t)$ and $\lambda_p^3(t)$ are the costate or adjoint variables associated to the dynamics of the state variables $\{a_i(t), \mathsf{pp}_i(t)\}$ for period 3. The first-order conditions (FOCs) for an interior consumption and hours worked are:

$$\frac{\partial \mathcal{H}^3}{\partial c_i} = e^{-\int_{x_0}^{t} \rho+\mu_i(j)dj} U'(c_i(t)) - \lambda_a^2(t) = 0. \tag{D.12}$$

Using the definition (15), the dynamic laws of the costate or adjoint variables are:

$$\frac{\partial \lambda_a^3(t)}{\partial t} = -\lambda_a^3(t)(r + \mu_i(t)), \tag{D.13}$$
$$\frac{\partial \lambda_p^3(t)}{\partial t} = -\lambda_a^3(t)\mathsf{f}_i(R_i)(1 - \varepsilon_i). \tag{D.14}$$

Moreover, the following matching conditions hold at the switching ages S_i and R_i for the costate or adjoint variables

$$\lambda_a(S_i) := \lambda_a^1(S_i) = \lambda_a^2(S_i), \tag{D.15a}$$
$$\lambda_h(S_i) := \lambda_h^1(S_i) = \lambda_h^2(S_i), \tag{D.15b}$$
$$\lambda_a(R_i) := \lambda_a^2(R_i) = \lambda_a^3(R_i), \tag{D.15c}$$
$$\lambda_p(R_i) := \lambda_p^2(R_i) = \lambda_p^3(R_i), \tag{D.15d}$$

and for the Hamiltonians

$$\mathcal{H}(S_i) := \mathcal{H}^1(S_i) = \mathcal{H}^2(S_i), \tag{D.16a}$$
$$\mathcal{H}(R_i) := \mathcal{H}^2(R_i) = \mathcal{H}^3(R_i). \tag{D.16b}$$

Taking into account the above matching conditions, let us define the marginal rate of substitution of assets for social contributions $\overline{\mathcal{P}}(t) = \phi\lambda_p(t)/\lambda_a(t)$ for periods $\{2, 3\}$. Differentiating $\overline{\mathcal{P}}(t)$ with respect to time t, and using the dynamics of the adjoint variables, gives

$$\frac{\partial \overline{\mathcal{P}}(t)}{\partial t} = \begin{cases} \overline{\mathcal{P}}(t)(r - \tilde{\mathsf{r}} + \mu_i(t) - \tilde{\mu}(t)) & \text{for } S_i < t < R_i, \\ \overline{\mathcal{P}}(t)(r + \mu_i(t)) - \phi \mathsf{f}_i(R_i)(1 - \varepsilon_i) & \text{for } t \geq R_i. \end{cases} \tag{D.17}$$

Solving (D.17) and using the fact that $\overline{\mathcal{P}}(\omega) = 0$, the marginal rate of substitution of assets for a dollar of social contribution is:

$$\overline{\mathcal{P}}(t) = \phi \mathsf{f}_i(R_i)(1 - \varepsilon_i) A_i(R_i, r) e^{\int_t^{R_i} \tilde{\mathsf{r}} + \tilde{\mu}(j) - (r + \mu_i(j)) dj}, \tag{D.18}$$

which is equivalent to $\overline{\mathcal{P}}(t) = \mathsf{P}(t)(1 - \varepsilon_i)$.

D.1 Optimal paths of consumption (c) and hours worked (ℓ)

From the FOCs, the optimal consumption and hours worked at age x satisfy

$$e^{-\int_{x_0}^t \rho + \mu_i(j) dj} U'(c_i(t)) = \lambda_a(t), \tag{D.19}$$

$$e^{-\int_{x_0}^t \rho + \mu_i(j) dj} \alpha_i v'(\ell_i(t)) = \lambda_a^2(t) h_i(S_i) \bar{w}(t - S_i)(1 - \tau + \tau \overline{\mathcal{P}}(t)). \tag{D.20}$$

Taking logarithms in both sides of the equation and differentiating with respect to t gives

$$-(\rho + \mu_i(t)) + \frac{c_i(t) U''(c_i(t))}{U'(c_i(t))} \frac{1}{c_i(t)} \frac{\partial c_i(t)}{\partial t} = \frac{1}{\lambda_a(t)} \frac{\partial \lambda_a(t)}{\partial t}, \tag{D.21}$$

$$-(\rho + \mu_i(t)) + \frac{\ell_i(t) v''(\ell_i(t))}{v'(\ell_i(t))} \frac{1}{\ell_i(t)} \frac{\partial \ell_i(t)}{\partial t} = \frac{1}{\lambda_a(t)} \frac{\partial \lambda_a(t)}{\partial t} + \frac{\bar{w}'(t - S_i)}{\bar{w}(t - S_i)} + \frac{\tau \overline{\mathcal{P}}'(t)}{1 - \tau + \tau \overline{\mathcal{P}}(t)}. \tag{D.22}$$

Using the envelope condition on assets held and rearranging terms, we obtain (12) and (13), respectively.

D.2 Optimal length of schooling (S_i)

Given an optimal retirement age R_i we first differentiate the expected utility $V_i(x_0)$ w.r.t. S and making it equal to zero

$$\int_{x_0}^{\omega} e^{-\int_{x_0}^t \rho + \mu_i(j) dj} U'(c_i(t)) \frac{\partial c_i(t)}{\partial S} dt - \int_S^{R_i} e^{-\int_{x_0}^t \rho + \mu_i(j) dj} \alpha_i v'(\ell_i(t)) \frac{\partial \ell_i(t)}{\partial S} dt$$
$$= e^{-\int_{x_0}^S \rho + \mu_i(j) dj} (\eta - \alpha_i v(\ell_i(S))). \tag{D.23}$$

Substituting the FOCs in the previous equation gives

$$\int_{x_0}^{\omega} \lambda_a(t)\frac{\partial c_i(t)}{\partial S}dt - \int_{S}^{R_i} \lambda_a(t)(1-\tau+\tau\overline{\mathcal{P}}(t))w_i(S_i,t)\frac{\partial \ell_i(t)}{\partial S}dt = e^{-\int_{x_0}^{S}\rho+\mu_i(j)dj}\left(\eta - \alpha_i v(\ell_i(S))\right). \tag{D.24}$$

Solving the envelope condition on assets gives $\lambda_a(t) = \lambda_a(x_0)e^{-\int_{x_0}^{t} r+\mu_i(j)dj}$. Substituting this last result in (D.24) and dividing by $\lambda_a(x_0)$ gives

$$\int_{x_0}^{\omega} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}\frac{\partial c_i(t)}{\partial S}dt - \int_{S}^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}(1-\tau+\tau\overline{\mathcal{P}}(t))w_i(S_i,t)\frac{\partial \ell_i(t)}{\partial S}dt = e^{-\int_{x_0}^{S}\rho+\mu_i(j)dj}\frac{\eta - \alpha_i v(\ell(S))}{\lambda_a(x_0)}. \tag{D.25}$$

Second, we differentiate the budget constraint (10) at age x_0 w.r.t. S

$$\begin{aligned}\int_{x_0}^{\omega} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}\frac{\partial c_i(t)}{\partial S}dt &= \int_{S}^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}w_i(S,t)\frac{\partial \ell_i(t)}{\partial S}dt \\ &+ \int_{S}^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}\frac{\partial w_i(S,t)}{\partial S}\ell_i(t)dt \\ &- e^{-\int_{x_0}^{S} r+\mu_i(j)dj}w_i(S,S)\ell_i(S) + \frac{\partial \mathsf{SSW}_i(x_0)}{\partial S}.\end{aligned} \tag{D.26}$$

From (B.7) we have

$$\mathsf{SSW}_i(x_0) = -\int_{S}^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}\tau(1-\mathsf{P}(t))w_i(S,t)\ell_i(t)dt. \tag{D.27}$$

First, differentiating (D.27) with respect to S gives

$$\begin{aligned}\frac{\partial \mathsf{SSW}_i(x_0)}{\partial S} &= -\int_{S}^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}\tau(1-\mathsf{P}(t))w_i(S,t)\frac{\partial \ell_i(t)}{\partial S}dt \\ &- \int_{S}^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}\tau(1-\mathsf{P}(t))\frac{\partial w_i(S,t)}{\partial S}\ell_i(t)dt \\ &+ e^{-\int_{x_0}^{S} r+\mu_i(j)dj}\tau(1-\mathsf{P}(S))w_i(S,S)\ell_i(S) \\ &+ \int_{S}^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}\tau\frac{\partial \mathsf{P}(t)}{\partial S}w_i(S,t)\ell_i(t)dt.\end{aligned} \tag{D.28}$$

Second, differentiating (5) with respect to S gives

$$\frac{\partial \mathsf{P}_i(t)}{\partial S} = \phi\frac{\partial \mathsf{f}_i(R_i)}{\partial \mathsf{pp}_i(R_i)}\frac{\partial \mathsf{pp}_i(R_i)}{\partial S}A_i(R_i,r)e^{\int_t^{R_i}\tilde{\mathsf{r}}+\tilde{\mu}(j)-(r+\mu_i(j))dj}. \tag{D.29}$$

Using (5) and (15) in (D.29) gives

$$\frac{\partial \mathsf{P}_i(t)}{\partial S} = -\frac{1}{\mathsf{pp}_i(R_i)}\frac{\partial \mathsf{pp}_i(R_i)}{\partial S}\varepsilon_i\mathsf{P}_i(t). \tag{D.30}$$

Substituting (D.30) on the third term on the right-hand side of Eq. (D.28), and using the fact that $\mathsf{P}_i(t) = \mathsf{P}_i(R_i)e^{\int_t^{R_i} \tilde{\mathsf{r}}+\tilde{\mu}_i(j)-(r+\mu_i(j))dj}$ gives, after canceling terms,

$$\frac{\partial \mathsf{P}_i(t)}{\partial S} = -\frac{\partial \mathsf{pp}_i(R_i)}{\partial S}\frac{\varepsilon_i \mathsf{P}_i(R_i)}{\phi}e^{-\int_{x_0}^{R_i} r+\mu_i(j)dj}. \tag{D.31}$$

Differentiating the total pension points at age R_i with respect to S gives

$$\begin{aligned}\frac{\partial \mathsf{pp}_i(R_i)}{\partial S} = \int_S^{R_i} e^{\int_t^{R_i} \tilde{\mathsf{r}}+\tilde{\mu}(j)dj}\phi\tau w_i(S,t)\frac{\partial \ell_i(t)}{\partial S}dt \\ + \int_S^{R_i} e^{\int_t^{R_i} \tilde{\mathsf{r}}+\tilde{\mu}(j)dj}\phi\tau\frac{\partial w_i(S,t)}{\partial S}\ell_i(t)dt \\ - e^{\int_S^{R_i} \tilde{\mathsf{r}}+\tilde{\mu}(j)dj}\phi\tau w_i(S,S)\ell_i(S).\end{aligned} \tag{D.32}$$

Plugging (D.32) in (D.31), and using the fact that $\mathsf{P}_i(R_i) = \mathsf{P}_i(t)e^{-\int_t^{R_i} \tilde{\mathsf{r}}+\tilde{\mu}(j)-(r+\mu_i(j))dj}$, gives

$$\begin{aligned}\frac{\partial \mathsf{P}_i(t)}{\partial S} &= -\frac{\partial \mathsf{pp}_i(R_i)}{\partial S}\frac{\mathsf{P}_i(R_i)}{\phi}\varepsilon_i e^{-\int_{x_0}^{R_i} r+\mu_i(j)dj} \\ &= -\int_S^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}\tau\varepsilon_i\mathsf{P}_i(t)w_i(S,t)\frac{\partial \ell_i(t)}{\partial S}dt \\ &\quad - \int_S^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}\tau\varepsilon_i\mathsf{P}_i(t)\frac{\partial w_i(S,t)}{\partial S}\ell_i(t)dt \\ &\quad + e^{-\int_{x_0}^{S} r+\mu_i(j)dj}\tau\varepsilon_i\mathsf{P}_i(S)w_i(S,S)\ell_i(S).\end{aligned} \tag{D.33}$$

Substituting (D.33) in the last term in (D.28) gives

$$\begin{aligned}\frac{\partial \mathsf{SSW}_i(x_0)}{\partial S} = -\int_{x_0}^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}\tau(1-\overline{\mathcal{P}}_i(t))w_i(S,t)\frac{\partial \ell_i(t)}{\partial S}dt \\ - \int_{x_0}^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}\tau(1-\overline{\mathcal{P}}_i(t))\frac{\partial w_i(S,t)}{\partial S}\ell_i(t)dt \\ + e^{-\int_{x_0}^{S} r+\mu_i(j)dj}\tau(1-\overline{\mathcal{P}}_i(S))w_i(S,S)\ell_i(S).\end{aligned} \tag{D.34}$$

By plugging (D.34) into (D.26) we have

$$\begin{aligned}\int_{x_0}^{\omega} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}\frac{\partial c_i(t)}{\partial S}dt - \int_S^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}(1-\tau+\tau\overline{\mathcal{P}}_i(t))w_i(S,t)\frac{\partial \ell_i(t)}{\partial S}dt \\ = \int_S^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}(1-\tau+\tau\overline{\mathcal{P}}_i(t))\frac{\partial w_i(S,t)}{\partial S}\ell_i(t)dt \\ - e^{-\int_{x_0}^{S} r+\mu_i(j)dj}(1-\tau+\tau\overline{\mathcal{P}}_i(t))w_i(S,S)\ell_i(S).\end{aligned}$$

Using the fact that the left-hand side of (D.25) and (D.31) are equal, then

$$
\begin{aligned}
e^{-\int_{x_0}^{S}\rho+\mu_i(j)dj}\frac{\eta-\alpha_i v(\ell_i(S))}{\lambda_a(x_0)} \\
&= \int_S^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}(1-\tau+\tau\overline{\mathcal{P}}_i(t))\frac{\partial w_i(S,t)}{\partial S}\ell_i(t)dt \\
&\quad - e^{-\int_{x_0}^{S} r+\mu_i(j)dj}(1-\tau+\tau\overline{\mathcal{P}}_i(t))w_i(S,S)\ell_i(S). \qquad \text{(D.35)}
\end{aligned}
$$

Now, differentiating $w_i(S,t)$ w.r.t. S gives

$$
\begin{aligned}
\frac{\partial w_i(S,t)}{\partial S} &= \frac{\partial \bar{w}(t-S)}{\partial S}h_i(S) + \bar{w}(t-S)\frac{\partial h_i(S)}{\partial S} \\
&= -\frac{1}{\bar{w}(t-S)}\frac{\partial \bar{w}(t-S)}{\partial t}w_i(S,t) + \frac{1}{h_i(S)}\frac{\partial h_i(S)}{\partial S}w_i(t,S). \qquad \text{(D.36)}
\end{aligned}
$$

Third, we use (D.36) in (D.35)

$$
\begin{aligned}
e^{-\int_{x_0}^{S}\rho+\mu_i(j)dj}\frac{\eta-\alpha_i v(\ell_i(S))}{\lambda_a(x_0)} \\
&= \frac{\frac{\partial h_i(S)}{\partial S}}{h_i(S)}\int_S^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}(1-\tau+\tau\overline{\mathcal{P}}_i(t))w_i(S,t)\ell_i(t)dt \\
&\quad - \int_S^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}\frac{\frac{\partial \bar{w}(t-S)}{\partial t}}{\bar{w}(t-S)}(1-\tau+\tau\overline{\mathcal{P}}_i(t))w_i(S,t)\ell_i(t)dt \\
&\quad - e^{-\int_{x_0}^{S} r+\mu_i(j)dj}(1-\tau+\tau\overline{\mathcal{P}}_i(t))w_i(S,S)\ell_i(S). \qquad \text{(D.37)}
\end{aligned}
$$

Therefore, after rearranging terms, the optimal length of schooling satisfies the following condition

$$
\begin{aligned}
\frac{\frac{\partial h_i(S)}{\partial S}}{h_i(S)}\int_S^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}(1-\tau+\tau\overline{\mathcal{P}}_i(t))w_i(S,t)\ell_i(t)dt \\
&= \int_S^{R_i} e^{-\int_{x_0}^{t} r+\mu_i(j)dj}\frac{\frac{\partial \bar{w}(t-S)}{\partial t}}{\bar{w}(t-S)}(1-\tau+\tau\overline{\mathcal{P}}_i(t))w_i(S,t)\ell_i(t)dt \\
&\quad + e^{-\int_{x_0}^{S} r+\mu_i(j)dj}(1-\tau+\tau\overline{\mathcal{P}}_i(t))w_i(S,S)\ell_i(S) \\
&\quad + e^{-\int_{x_0}^{S} r+\mu_i(j)dj}\frac{\eta-\alpha_i v(\ell_i(S))}{\lambda_a(x_0)}. \qquad \text{(D.38)}
\end{aligned}
$$

Defining human capital at age S net of effective labor income taxes as

$$
W_i(S,R_i) = \int_S^{R_i} e^{-\int_S^t r+\mu_i(j)dj}(1-\tau+\tau\overline{\mathcal{P}}_i(t))w_i(S,t)\ell_i(t)dt \qquad \text{(D.39)}
$$

and dividing both sides of (D.38) by $W_i(S,R_i)$, multiplying by $e^{\int_{x_0}^{S} r+\mu_i(j)dj}$, we obtain the optimal length of schooling condition

$$
r_i^h(S) = \bar{r}_i(S,R_i) + \frac{\eta-\alpha_i v(\ell_i(S))}{U'(c_i(S))W_i(S,R_i)}. \qquad \text{(D.40)}
$$

$r_i^h(S)$ is the return to education at age S for an individual of type i

$$r_i^h(S) = \frac{1}{h_i(S)} \frac{\partial h_i(S)}{\partial S}, \tag{D.41}$$

$\bar{r}_i(S, R)$ is the rate of return lost from not working at age S or the marginal cost of the Sth unit of schooling for an individual of type i

$$\bar{r}_i(S, R_i) = \int_S^{R_i} \frac{\frac{\partial \bar{w}(t-S)}{\partial t}}{\bar{w}(t-S)} \psi_i(t) dt + \psi_i(S), \tag{D.42}$$

where

$$\psi_i(t) = \frac{e^{-\int_{x_0}^t r+\mu_i(j)dj}(1-\tau+\tau\overline{\mathcal{P}}_i(t))w_i(S,t)\ell_i(t)}{\int_S^{R_i} e^{-\int_{x_0}^u r+\mu_i(j)dj}(1-\tau+\tau\overline{\mathcal{P}}_i(u))w_i(S,u)\ell_i(u)du}. \tag{D.43}$$

From (D.42) and (D.43) we have $\int_S^{R_i} \psi_i(t)dt = 1$ and $\lim_{S \to R_i} \bar{r}_i(S, R_i) = 1$.

D.3 Optimal retirement age (R_i)

Similar to the previous subsection we start assuming that the optimal length of schooling S_i is given. Then, we differentiate the expected utility $V_i(x_0)$ w.r.t. the optimal retirement age R and equate the result to the derivative of the lifetime budget constrain w.r.t. to the optimal retirement age.

Proof. Given an optimal length of schooling S_i we first differentiate the expected utility $V_i(x_0)$ w.r.t. R and making it equal to zero

$$\int_{x_0}^{\omega} e^{-\int_{x_0}^t \rho+\mu_i(j)dj} U'(c_i(t)) \frac{\partial c_i(t)}{\partial R} dt - \alpha_i \int_{S_i}^{R} e^{-\int_{x_0}^t \rho+\mu_i(j)dj} v'(\ell_i(t)) \frac{\partial \ell_i(t)}{\partial R} dt$$
$$= e^{-\int_{x_0}^R \rho+\mu_i(j)dj} \left(\alpha_i v(\ell_i(R)) + \varphi(R)\right). \tag{D.44}$$

Substituting the FOCs in the previous equation gives

$$\int_{x_0}^{\omega} \lambda_a(t) \frac{\partial c_i(t)}{\partial R} dt - \int_{S_i}^{R} \lambda_a(t)(1-\tau+\tau\overline{\mathcal{P}}_i(t))w_i(S_i,t) \frac{\partial \ell_i(t)}{\partial R} dt$$
$$= e^{-\int_{x_0}^R \rho+\mu_i(j)dj} \left(\alpha_i v(\ell_i(R)) + \varphi(R)\right). \tag{D.45}$$

Using the envelope conditions on assets, which gives $\lambda_a(t) = \lambda_a(x_0)e^{-\int_{x_0}^t r+\mu_i(j)dj}$, dividing both sides of the equation by $\lambda_a(x_0)$ and rearranging terms gives

$$\int_{x_0}^{\omega} e^{-\int_{x_0}^t r+\mu_i(j)dj} \frac{\partial c_i(t)}{\partial R} dt - \int_{S_i}^{R} e^{-\int_{x_0}^t r+\mu_i(j)dj}(1-\tau+\tau\overline{\mathcal{P}}_i(t))w_i(S_i,t) \frac{\partial \ell_i(t)}{\partial R} dt$$
$$= e^{-\int_{x_0}^R \rho+\mu_i(j)dj} \frac{\alpha_i v(\ell_i(R)) + \varphi(R)}{\lambda_a(x_0)}. \tag{D.46}$$

Second, we differentiate the budget constraint (10) at age x_0 w.r.t. R

$$\int_{x_0}^{\omega} e^{-\int_{x_0}^{t} r+\mu_i(j)dj} \frac{\partial c_i(t)}{\partial R} dt = \int_{S_i}^{R} e^{-\int_{x_0}^{t} r+\mu_i(j)dj} w_i(S_i,t) \frac{\partial \ell_i(t)}{\partial R} dt + e^{-\int_{x_0}^{R} r+\mu_i(j)dj} w_i(S_i,R)\ell_i(R) + \frac{\partial \mathsf{SSW}_i(x_0)}{\partial R}. \quad \text{(D.47)}$$

Differentiating (B.7) at age x_0 with respect to R gives

$$\begin{aligned} \frac{\partial \mathsf{SSW}_i(x_0)}{\partial R} = &- \int_{S_i}^{R} e^{-\int_{x_0}^{t} r+\mu_i(j)dj} \tau(1-\mathsf{P}_i(t)) w_i(S_i,t) \frac{\partial \ell_i(t)}{\partial R} dt \\ &- e^{-\int_{x_0}^{R} r+\mu_i(j)dj} \tau(1-\mathsf{P}_i(R)) w_i(S_i,R)\ell_i(R) \\ &+ \int_{S_i}^{R} e^{-\int_{x_0}^{t} r+\mu_i(j)dj} \tau \frac{\partial \mathsf{P}_i(t)}{\partial R} w_i(S_i,t)\ell_i(t)dt. \end{aligned} \quad \text{(D.48)}$$

Now, by differentiating (5) with respect to R we have

$$\begin{aligned} \frac{\partial \mathsf{P}_i(t)}{\partial R} &= \mathsf{P}_i(t) \left(\frac{1}{\mathsf{f}_i(R)} \frac{\partial \mathsf{f}_i(R)}{\partial R} + \frac{1}{A_i(R,r)} \frac{\partial A_i(R,r)}{\partial R} + \tilde{\mathsf{r}} + \tilde{\mu}(R) - (r+\mu_i(R)) \right) \\ &= \mathsf{P}_i(t) \left(\frac{1}{\mathsf{f}_i(R)} \frac{\partial \mathsf{f}_i(R)}{\partial R} - \frac{1}{A_i(R,r)} + \tilde{\mathsf{r}} + \tilde{\mu}(R) \right). \end{aligned} \quad \text{(D.49)}$$

Given that $\mathsf{f}_i(R)$ depends on both R and pp, the total derivative of $\mathsf{f}_i(R)$ with respect to R is

$$\frac{\partial \mathsf{f}_i(R)}{\partial R} = \left. \frac{\partial \mathsf{f}_i(R)}{\partial R} \right|_{\mathsf{pp}} + \frac{\partial \mathsf{f}_i(R)}{\partial \mathsf{pp}_i(R)} \frac{\partial \mathsf{pp}_i(R)}{\partial R}. \quad \text{(D.50)}$$

Thus, from (15) and (D.50), Eq. (D.49) can be rewritten as follow

$$\frac{\partial \mathsf{P}_i(t)}{\partial R} = \mathsf{P}_i(t) \left(\frac{1}{\mathsf{f}_i(R)} \left. \frac{\partial \mathsf{f}_i(R)}{\partial R} \right|_{\mathsf{pp}} - \frac{1}{A_i(R,r)} + \tilde{\mathsf{r}} + \tilde{\mu}(R) \right) - \mathsf{P}_i(t)\varepsilon_i \frac{1}{\mathsf{pp}_i(R)} \frac{\partial \mathsf{pp}_i(R)}{\partial R}. \quad \text{(D.51)}$$

Substituting (D.51) on (D.48) we get

$$\begin{aligned} \frac{\partial \mathsf{SSW}_i(x_0)}{\partial R} = &- \int_{S_i}^{R} e^{-\int_{x_0}^{t} r+\mu_i(j)dj} \tau(1-\mathsf{P}_i(t)) w_i(S_i,t) \frac{\partial \ell_i(t)}{\partial R} dt \\ &- e^{-\int_{x_0}^{R} r+\mu_i(j)dj} \tau(1-\mathsf{P}_i(R)) w_i(S_i,R)\ell_i(R) \\ &+ \left(\frac{1}{\mathsf{f}_i(R)} \left. \frac{\partial \mathsf{f}_i(R)}{\partial R} \right|_{\mathsf{pp}} - \frac{1}{A_i(R,r)} + \tilde{\mathsf{r}} + \tilde{\mu}(R) \right) \int_{S_i}^{R} e^{-\int_{x_0}^{t} r+\mu_i(j)dj} \tau \mathsf{P}_i(t) w_i(S_i,t)\ell_i(t)dt \\ &- \frac{1}{\mathsf{pp}_i(R)} \frac{\partial \mathsf{pp}(R)}{\partial R} \int_{S_i}^{R} e^{-\int_{x_0}^{t} r+\mu_i(j)dj} \tau \mathsf{P}_i(t)\varepsilon_i w_i(S_i,t)\ell_i(t)dt. \end{aligned} \quad \text{(D.52)}$$

Using the relation $\mathsf{P}_i(t) = \mathsf{P}_i(R) e^{\int_t^R \tilde{\mathsf{r}} + \tilde{\mu}(j) - (r + \mu_i(j)) dj}$ in the last two terms of (D.52), and rearranging terms, we have

$$\begin{aligned}\frac{\partial \mathsf{SSW}_i(x_0)}{\partial R} = &- \int_{S_i}^{R} e^{-\int_{x_0}^{t} r + \mu_i(j) dj} \tau (1 - \mathsf{P}_i(t)) w_i(S_i, t) \frac{\partial \ell_i(t)}{\partial R} dt \\ &- e^{-\int_{x_0}^{R} r + \mu_i(j) dj} \tau (1 - \mathsf{P}_i(R)) w_i(S_i, R) \ell_i(R) \\ &+ e^{-\int_{x_0}^{R} r + \mu_i(j) dj} \left(\frac{1}{\mathsf{f}_i(R)} \frac{\partial \mathsf{f}_i(R)}{\partial R} \Big|_{\mathsf{pp}} - \frac{1}{A_i(R, r)} + \tilde{\mathsf{r}} + \tilde{\mu}(R) \right) \mathsf{P}_i(R) \frac{\mathsf{pp}_i(R)}{\phi} \\ &- e^{-\int_{x_0}^{R} r + \mu_i(j) dj} \frac{\partial \mathsf{pp}(R)}{\partial R} \frac{\mathsf{P}_i(R) \varepsilon_i}{\phi}. \end{aligned} \tag{D.53}$$

Now, from (B.2) we differentiate the total pension points at age R with respect to R, i.e. $\frac{\partial \mathsf{pp}(R)}{\partial R}$, which gives

$$\begin{aligned}\frac{\partial \mathsf{pp}_i(R_i)}{\partial S} = &\int_{S_i}^{R} e^{\int_t^R \tilde{\mathsf{r}} + \tilde{\mu}(j) dj} \phi \tau w_i(S, t) \frac{\partial \ell_i(t)}{\partial R} dt \\ &+ (\tilde{\mathsf{r}} + \tilde{\mu}(R)) \mathsf{pp}_i(R) + \phi \tau w_i(S_i, R) \ell_i(R). \end{aligned} \tag{D.54}$$

Plugging (D.54) in (D.53) and rearranging terms gives

$$\begin{aligned}\frac{\partial \mathsf{SSW}_i(x_0)}{\partial R} = &- \int_{S_i}^{R} e^{-\int_{x_0}^{t} r + \mu_i(j) dj} \tau (1 - \mathsf{P}_i(t)(1 - \varepsilon_i)) w_i(S_i, t) \frac{\partial \ell_i(t)}{\partial R} dt \\ &- e^{-\int_{x_0}^{R} r + \mu_i(j) dj} \tau (1 - \mathsf{P}_i(R)(1 - \varepsilon_i)) w_i(S_i, R) \ell_i(R) \\ &+ e^{-\int_{x_0}^{R} r + \mu_i(j) dj} \left(\frac{1}{\mathsf{f}_i(R)} \frac{\partial \mathsf{f}_i(R)}{\partial R} \Big|_{\mathsf{pp}} - \frac{1}{A_i(R, r)} + (\tilde{\mathsf{r}} + \tilde{\mu}(R))(1 - \varepsilon_i) \right) \mathsf{P}_i(R) \frac{\mathsf{pp}_i(R)}{\phi}. \end{aligned} \tag{D.55}$$

Next, plugging (D.55) in (D.47) gives

$$\begin{aligned}&\int_{x_0}^{\omega} e^{-\int_{x_0}^{t} r + \mu_i(j) dj} \frac{\partial c_i(t)}{\partial R} dt - \int_{S_i}^{R} e^{-\int_{x_0}^{t} r + \mu_i(j) dj} (1 - \tau + \tau \overline{\mathcal{P}}_i(t)) w_i(S_i, t) \frac{\partial \ell_i(t)}{\partial R} dt \\ &\quad = e^{-\int_{x_0}^{R} r + \mu_i(j) dj} (1 - \tau + \tau \overline{\mathcal{P}}_i(t)) w_i(S_i, R) \ell_i(R) \\ &+ e^{-\int_{x_0}^{R} r + \mu_i(j) dj} \left(\frac{1}{\mathsf{f}_i(R)} \frac{\partial \mathsf{f}_i(R)}{\partial R} \Big|_{\mathsf{pp}} - \frac{1}{A_i(R, r)} + (\tilde{\mathsf{r}} + \tilde{\mu}(R))(1 - \varepsilon_i) \right) \mathsf{P}_i(R) \frac{\mathsf{pp}_i(R)}{\phi}. \end{aligned} \tag{D.56}$$

Using the fact that the left-hand side of (D.3) and (D.56) are equal, we obtain

$$\begin{aligned}&e^{-\int_{x_0}^{R} \rho + \mu_i(j) dj} \frac{\alpha_i v\left(\ell_i(R)\right) + \varphi(R)}{\lambda_a(x_0)} = e^{-\int_{x_0}^{R} r + \mu_i(j) dj} (1 - \tau + \tau \overline{\mathcal{P}}_i(t)) w_i(S_i, R) \ell_i(R) \\ &+ e^{-\int_{x_0}^{R} r + \mu_i(j) dj} \left(\frac{1}{\mathsf{f}_i(R)} \frac{\partial \mathsf{f}_i(R)}{\partial R} \Big|_{\mathsf{pp}} - \frac{1}{A_i(R, r)} + (\tilde{\mathsf{r}} + \tilde{\mu}(R))(1 - \varepsilon_i) \right) \mathsf{P}_i(R) \frac{\mathsf{pp}_i(R)}{\phi}. \end{aligned} \tag{D.57}$$

Multiplying both sides of (D.57) by $e^{\int_{x_0}^{R} r+\mu_i(j)dj}$ gives

$$e^{(r-\rho)(R-x_0)}\frac{\alpha_i v\left(\ell_i(R)\right)+\varphi(R)}{\lambda_a(x_0)} = (1-\tau+\tau\overline{\mathcal{P}}_i(t))w_i(S_i,R)\ell_i(R) + \left(\frac{1}{\mathsf{f}_i(R)}\left.\frac{\partial \mathsf{f}_i(R)}{\partial R}\right|_{\mathsf{pp}} - \frac{1}{A_i(R,r)} + (\tilde{\mathsf{r}}+\tilde{\mu}(R))(1-\varepsilon_i)\right)\mathsf{P}_i(R)\frac{\mathsf{pp}_i(R)}{\phi}. \quad \text{(D.58)}$$

Defining the implicit tax rate on retirement $\uptau_i^{GW}(R)$ as

$$\uptau_i^{GW}(R) = \tau(1-\overline{\mathcal{P}}_i(t)) - \left(\frac{1}{\mathsf{f}_i(R)}\left.\frac{\partial \mathsf{f}_i(R)}{\partial R}\right|_{\mathsf{pp}} - \frac{1}{A_i(R,r)} + (\tilde{\mathsf{r}}+\tilde{\mu}(R))(1-\varepsilon_i)\right)\frac{\mathsf{P}_i(R)\frac{\mathsf{pp}_i(R)}{\phi}}{w_i(S_i,R)\ell_i(R)}, \quad \text{(D.59)}$$

where the first term is the implicit tax paid by working one additional period and the second term is the relative change in the social security wealth caused by postponing the retirement age.

Then, using (D.59) and the fact that $U'(c_i(R)) = e^{-(r-\rho)(R-x_0)}\lambda_a(x_0)$ we obtain that the optimal retirement age condition satisfies

$$\alpha_i v\left(\ell_i(R)\right)+\varphi(R) = U'(c_i(R))w_i(S_i,R)\ell_i(R)(1-\uptau_i^{GW}(R)), \quad \text{(D.60)}$$

which coincides with (18).

■

E Additional simulated data

Length of schooling S_i.

Table 6: Optimal length of schooling by income quintile (S_i), US male birth cohorts 1930 and 1960

	Defined Contribution (NDC)			Defined Benefit		
	Avg. LT NDC-I	Corrected Avg. LT NDC-II	i–th LT NDC-III	Non–progressive DB-I	Progressive DB-II	Progressive Corrected DB-III
Cohort 1930						
Quintile 1	11.5	11.5	11.6	12.0	11.4	11.4
Quintile 2	11.8	11.9	11.9	12.4	11.8	11.8
Quintile 3	12.3	12.3	12.3	12.8	12.3	12.3
Quintile 4	14.1	14.1	14.1	14.5	13.0	13.0
Quintile 5	16.1	16.1	16.0	16.3	15.0	15.1
Cohort 1960						
Quintile 1	11.3	11.3	11.3	11.6	10.7	10.7
Quintile 2	12.3	12.4	12.5	12.9	12.1	12.4
Quintile 3	14.7	14.8	14.8	15.4	14.7	14.8
Quintile 4	17.6	17.6	17.5	18.5	16.3	16.4
Quintile 5	18.9	18.8	18.8	19.8	17.9	18.0

Retirement ages R_i.

Table 7: Optimal retirement age by income quintile (R_i), US male birth cohorts 1930 and 1960

	Defined Contribution (NDC)			Defined Benefit		
	Avg. LT NDC-I	Corrected Avg. LT NDC-II	i–th LT NDC-III	Non–progressive DB-I	Progressive DB-II	Progressive Corrected DB-III
Cohort 1930						
Quintile 1	61.4	61.4	61.7	63.9	63.3	63.4
Quintile 2	61.7	61.8	61.4	64.2	63.6	63.6
Quintile 3	62.1	62.2	62.2	64.5	63.9	64.0
Quintile 4	63.4	63.5	63.3	65.0	63.8	63.8
Quintile 5	64.8	64.8	64.5	65.5	64.9	64.8
Cohort 1960						
Quintile 1	60.9	60.9	61.4	62.9	62.0†	62.0†
Quintile 2	61.8	62.0	62.3	64.6	63.1	64.6
Quintile 3	63.9	64.1	64.0	66.5	65.9	66.0
Quintile 4	66.1	66.0	65.7	69.2	66.5	66.4
Quintile 5	66.9	66.7	66.4	69.6	67.7	67.5

Notes: Symbol '†' represents cases in which individuals exit the labor market before they start claiming retirement benefits. In particular, individuals subject to the mortality regime of the 1960 cohort retire on average at age 60.7 in the DB-II system and at age 60.5 in the DB-III system.

Present value of lifetime benefits.

Table 8: Present value of lifetime benefits at age 50 by income quintile and pension system, US males, birth cohorts 1930 and 1960 (in $1 000s)

	Defined Contribution (NDC)			Defined Benefit (DB)		
	Avg. LT NDC-I	Corrected Avg. LT NDC-II	i–th LT NDC-III	Non–progressive DB-I	Progressive DB-II	Progressive Corrected DB-III
Cohort 1930						
Quintile 1	121.14	132.10	140.65	131.18	130.44	140.90
Quintile 2	143.23	154.27	160.84	158.53	151.42	159.75
Quintile 3	161.60	167.89	171.11	180.16	169.95	174.93
Quintile 4	210.66	207.91	204.48	237.64	193.46	188.46
Quintile 5	276.77	255.63	241.26	307.96	232.82	212.61
Cohort 1960						
Quintile 1	101.96	135.24	153.90	96.99	103.51	132.05
Quintile 2	137.69	171.77	188.27	144.39	139.19	177.13
Quintile 3	214.28	228.64	230.16	251.80	232.15	238.26
Quintile 4	330.44	305.37	290.96	446.24	298.78	265.81
Quintile 5	389.47	346.37	328.62	538.19	343.82	294.86

Lifetime wealth.

Table 9: Distribution of lifetime wealth (LW) at age 14 by income quintile and mortality regime in the NDC-III system (in \$1 000s)

	Cohort 1930			Cohort 1960		
	SSW	HK	LW	SSW	HK	LW
	I	II	III=I+II	I	II	III=I+II
Quintile 1	-16.80	541.45	525.65	-18.14	521.38	503.24
Quintile 2	-19.51	630.52	611.01	-23.56	679.61	656.05
Quintile 3	-21.23	686.83	665.60	-32.55	933.29	900.74
Quintile 4	-26.62	871.88	845.26	-45.77	1 312.79	1 267.01
Quintile 5	-33.64	1 113.04	1 079.39	-53.38	1 543.85	1 490.47

Notes: SSW stands for social security wealth, HK denotes the stock of human capital, and LW is the lifetime wealth.

Table 9 reports the lifetime wealth (detrended by productivity) by income quintile relative to that obtained for the income group q3 under the mortality regime 1930. We can see in Tab. 9 that the higher income quintiles q3–q5 experience an average increase over twenty percent in their lifetime wealth with the more unequal mortality regime, q2 experiences an increase of seven percent in the lifetime wealth, and q1 has five percent less wealth. Moreover, since the NDC-III system provides the same internal rate of return across income groups, we have that the ratio between the social security wealth and the stock of human capital is the same across income group.