Optimal Public Pensions: Maximum deferral, then covering all retiree consumption

Henning Bohn*

September 18, 2018

Abstract

Public pensions typically start at retirement and provide partial funding for retirement consumption. Retirees must rely on savings for the remainder. A pension system offering higher benefits at the end of life would be significantly more efficient. In an optimal system, retirees first rely entirely on their savings, and then public pensions pay for all retirement consumption. The system's funding level determines the age(s) of pension eligibility, which may vary by income if progressivity is desired. An implication is that pension reforms should focus on adjusting the pension age, with advance notice, while maintaining high replacement rates.

*University of California Santa Barbara. Department of Economics, NH 2127, UCSB, Santa Barbara, CA 93106; henning.bohn@ucsb.edu; www.econ.ucsb.edu/~bohn.

Pension systems around the world are struggling with rising longevity and declining birth rates, which create pressure for cost-reducing reforms. The standard approach to thinking about retirement is to view public and private funding as complementary pillars from which the elderly draw simultaneously throughout their retirement (see, e.g., Poterba 2014). Reducing costs then means reduced monthly payments and an expectation that younger cohorts must save more.

This paper makes a case that a sequential payout structure would be more efficient. Optimal public funding specifies a time gap between retirement and pension eligibility. During this period, retirees are expected to finance their consumption entirely from private savings. Once the public pension start, it should cover all desired consumption. Put differently, for any given present value of pension benefits, the optimal pension age is the age at which this present value matches the expected present value of the person's remaining lifetime consumption.

The main argument is that public pensions and private savings have comparative advantages at different time horizons. For private savings, key challenges are managing assets and coping with longevity risk. The resulting costs tend to be cumulative—typically annual charges—and are highest in old age. For public pensions, the main challenge is the excess burden of taxes. This may be substantial, but it is a one-time cost. This suggests that the use of public funds—any given amount—should be concentrated late in life. This would amortize the excess burden over the longest horizon and maximize the government's comparative advantage with regard to management cost and annuitization. Private savings are relatively more efficient for the period soon after retirement.

Several related issues are not modeled explicitly but worth noting. First, financial literacy is a concern. Making a financial plan that ends when public funding starts would be much easier than planning for an open-ended lifetime. This is important because medical and cognitive problems that limit decision-making are increasing with age (Gamble et al. 2015).[1] Second, public coverage of advanced-age consumption implies that the government would bear a large share of aggregate longevity risk. This is appropriate from the perspective of intergenerational risk sharing (Bohn 2006). Third, public pensions are often conditional on not working and thereby encourage early retirement (Gruber and Wise 2008). A time gap between retirement and pension eligibility would eliminate this problem.

[1] To illustrate how retirement planning would simplify, consider someone retiring at age 65 who expects to cover half of retirement consumption from own savings and the other half from public pensions. Suppose the maximum age is 120, and consumption after age 75 accounts for half of the present value of retirement consumption. Funding half of annual consumption privately means buying life annuities or taking the risk of outliving ones savings. Buying annuities means having to trust a company to remain solvent for a very long time (55 years–from 65 to 120) The government should be indifferent between paying 50 percent of consumption starting at age 65 or 100 percent starting at 75. The latter would make individuals responsible for funding all consumption between ages 65 or 75, which is easier than planning to age 120; it would exclude the advanced ages at which survival probabilities are low and cognitive problems most likely.

Existing public pension systems – in the U.S. and around the world – fail to exploit the comparative advantage of concentrating public funding at the end of life. Instead, pensions are paid as soon as a person retires. The retirement age is treated as essentially synonymous with the eligibility age for public pensions. This confounds two separate decisions. The retirement age is an individual choice – an optimal labor supply decision – whereas pension eligibility is a policy issue. The point of this paper is that the retirement age and the age at which public pensions start (*pension age*, for brevity) should be as far apart as possible, subject to resource constraints.

Some pension systems allow members to choose when to start receiving benefits. U.S. social security, for instance, makes actuarial adjustments when individuals start drawing benefits before or after their "full" retirement age, but only within the age range 62-70. Financial advisors sometimes recommend that healthy retirees should defer claiming social security, which goes in the direction public pension starting later in exchange for higher monthly benefits. However, such advice is usually motivated by private information about (good) health and comes with exhortations to defer retirement. My analysis does not rely on private information and separates retirement from claiming benefits.

The growth in pension cost resulting from rising longevity has motivates numerous policy proposal suggesting cuts in pensions and telling retirees to rely more on private savings. The options are usually presented in terms of marginal shifts between multiple income sources from which retirees are supposed to draw simultaneously. My analysis implies that benefits cuts (if any) are best implemented by raising the pension age and not by reducing the replacement rate. The analysis also implies that progressivity – more private responsibility at higher incomes – should be implemented by differentiating the pension age across income levels. High-income individuals would have to cover more retirement years with savings than low-income individuals, but the replacement rate would be the same once pensions start (optimally: 100 percent).

To study the optimal interaction of private savings and public pensions, one must avoid assumptions that would produce extreme solutions. If private financial markets were perfectly efficient and annuities available without cost, even small tax distortions would make public funding inefficient. If governments could credibly promise that payroll "contributions" are individually tracked and the full value is returned as pension, an optimal tax-transfer scheme could dominate private savings. To avoid these unrealistic extremes, the model assumes intermediation cost for private savings and it restricts payroll taxes to be not fully refundable. This setting ensures that there is an economic rationale for both private retirement savings and for public funding.

Intermediation costs and tax distortions are complicated and require simplifications. In theory, retirement savers could manage their assets are near-zero cost by holding index funds and deal with uncertain mortality by placing all their assets into fair annuities (Yaari 1965). In practice, many investors require advice to manage their assets, or risk losses due to mistakes, and annuitization requires that wealth is turned over to outside managers. The latter incurs intermediation costs and creates undiversified default risks, as annuitants are long-term creditors of the insurer.

In the model, intermediation costs are captured simply by a parameter for maximum annuitization cost. At ages for which mortality is greater than the annuitization cost, individuals place their private savings into annuities. At younger ages, individuals are assumed to earn a non-annuitized return– interpreted either as regular savings, or more conveniently, as annuitized return with intermediation cost just equal to mortality.

Intermediation costs are also affected by the availability of corporate pension plans. In theory, such plans can offer low-cost, annuitized private pensions suitable to supplement public pensions. Corporate pensions arguably motivated the paradigm of multiple pension pillars. However, regulations imposed in response to high-profile defaults have made them prohibitively costly for corporations to offer, at least in the U.S. Hence the model does not include corporate pensions.

There is an extensive literature on annuities. Brugiavini (1993) examines why annuities are not widely used, contrary to Yaari's (1965) findings. Numerous empirical studies document the money's worth of annuities, e.g., Mitchell et al (2001). There is also a literature on the optimal timing of payroll taxes, e.g., Fenge et al (2006). The main innovation of the paper is to present sequential payouts – private then public – as organizing principle for the design of public pensions. (The principle also applies to non-cash benefits. For example, Medicaid coverage for nursing homes is a retiree benefit appropriated focused on very old age.)

The paper is organized as follows. Section 1 sets up a tractable overlapping generations model with stochastic mortality and describes conditions for Pareto-efficient policies. The basic model assumes linear taxes and abstracts from cross-sectional heterogeneity and from individualized pension promises. Section 2 provides an interpretation of public versus private retirement funding as choice between investment funds that have one-time versus recurrent cost. Section 3 examines several extensions: A setting with earnings-linked benefits shows that payouts are the end of life have an additional benefit of minimizing tax distortions. A setting with cross-sectional heterogeneity shows that payouts are the end of life are optimal for each types of agent, with pensions ages generally differing across types. The results also extend to versions with non-linear taxes and with uniformity restriction on taxation. Section 4 provides a numerical calculations to quantify the welfare gains from optimizing pension payouts and to illustrate how pension policy should respond to rising longevity and to fiscal shortfalls. Section 5 concludes. Proofs are in an appendix.

1 Optimal Policy in a tractable Overlapping-Generations Model

The model in this section is designed to allow a tractable analysis of optimal pensions and taxes for a generic cohort within a fiscal structure that covers many generations. Tractability is obtained by modeling labor-leisure choices so that taxes do not distort margins other than labor supply; by allowing tax rates to differ across cohorts (by age); and by assuming a representative agent per cohort.

1.1 Population and Preferences

Let time t be discrete and let age be denoted by i. Abstracting from childhood, cohort t is "born" at the start of working age ($i = 0$) and has maximum life span $i_{\max}$. Let $\Pi_{[0,i],t} = E_t[I^{alive}_{i,t+i}]$ denote the unconditional survival probabilities to age i, where $I^{alive}_{i,t+i}$ is a 0-1 indicator for being alive at time $t+i$. Survival probabilities are known and exogenous. Conditional survival probabilities are $\Pi_{[i,j],t+i} = \Pi_{[0,j],t}/\Pi_{[0,i],t}$ for general $j > i$, and $\Pi_{i,t+i} = \Pi_{[0,i+1],t}/\Pi_{[0,i],t}$ denotes one-period survival at age i. Assume $\Pi_{i,t+i} \in (0,1)$ for all $i < i_{\max}$, decreases with age, ending with $\Pi_{i_{\max},t+i_{\max}} = 0$.

Individuals have preferences over consumption $c_{i,t+i}$:

$$U_t = \sum_{i=0}^{i_{\max}} \beta^i \cdot I^{alive}_{i,t+i} \cdot u(c_{i,t+i})$$

Period utility u is increasing and concave with $u'(0) = \infty$, and $\beta > 0$.[2]

The cohort size N_t is exogenous and members of each cohort are identical. (See Section 3 for cross-sectional heterogeneity.)

To ease notation, time subscripts are henceforth omitted when dealing with a generic cohort t. Also, lagged terms in U_t are omitted when expected utility $E_{t+i}U_t$ (or E_iU) is evaluated at age $i > 0$.

1.2 Labor Supply and Retirement

Assumption on labor supply are designed so that taxes discourage market labor without distorting other margins in the model, and so there is a well-defined retirement age.

Individuals have unit time, supply $l_i \in [0,1]$ to the market, and $1 - l_i$ to home production H, which increasing, concave, and satisfies $H'(1) < 1$ and $H'(0) = \infty$. Non-zero market labor incurs an age-dependent fixed cost $\mu_i \geq 0$.

Market labor is subject to a wage tax at rate τ_i. Then labor income after taxes is

$$y_i = w_i(1 - \tau_i)l_i + w_iH(1 - l_i) - \mu_i \cdot I_{\{l_i>0\}}, \tag{1}$$

where I is the indicator function. Assume the government cannot credibly link taxes to future transfers (for now; see Section 3 for earnings-linked pensions.) Then labor supply maximizes y_i, which implies $l(\tau_i) \equiv 1 - (H')^{-1}(1 - \tau_i)$ with $l'(\tau_i) < 0$. Define $L(\tau) \equiv (1 - \tau)l(\tau) + H(1 - l(\tau))$. Then labor income at age i is $y_i = w_iL(\tau_i) - \mu_i$. Tax payments are $w_iT(\tau_i)$, where $T(\tau) \equiv \tau l(\tau)$.

[2] The model assumes no gift or bequest motives because altruism or joy-of-giving would severely complicate the normative analysis. At the individual level, desired gifts and bequests at various ages could be subsumed into consumption in the respective periods; this is not modeled formally to avoid clutter.

Let $\hat{\tau} = \arg\max_\tau\{\tau l(\tau)\}$ be the revenue-maximizing tax rate (Laffer curve peak). For $\tau \in (0, \hat{\tau})$, let $\Phi(\tau) = \frac{-L'(\tau)-T'(\tau)}{T'(\tau)} > 0$ denote the marginal excess burden, i.e., the reduction in pre-tax income per unit revenue.

Retirement is modeled by assumptions on μ_i. For now, assume μ_i is a step function that jumps at some age $i_R \in (0, i_{\max})$ from $\mu_i = 0$ for $i < i_R$ to $\mu_i > l(0) \cdot \max_{i \geq i_R}\{w_i\}$ for $i \geq i_R$. (This is generalized in Section 3.) Then i_R is a well-defined *retirement age* in the sense that $l_i = 0$ is optimal for all $i \geq i_R$.

Retirement income from home production, $y_i = w_i H(1)$, would needlessly clutter the analysis. Hence I normalize $H(1) = 0$. Then for $i \geq i_R$, $y_i = 0$, and for $i < i_R$, y_i is market income minus the opportunity cost of forgone home production.

Overall, these assumptions ensure that income and labor supply are separable from consumption and savings, that taxes have a well-defined excess burden, and that retirement is well-defined.

1.3 Private Savings with Financial Frictions

Uncertainty and time variation in returns would needlessly complicated analysis. Assume therefore that the gross return on assets is $R = 1 + r$, where $r > 0$ is a constant interest rate. This return is available to the government and to institutional investors, notably to annuity providers.

Individuals at age i face a management cost κ^s for regular savings (amount x_i^s), a charge κ^a on annuities (x_i^a), and a borrowing cost κ^d on debt (d_i). Conditional on survival to age $i+1$, regular savings pay $(1-\kappa^s)R$ and annuities pay $(1-\kappa^a) \cdot R/\Pi_i$. Assuming debts are discharged upon death, lenders charge $(1+\kappa^d) \cdot R/\Pi_i$ conditional on survival.

Net assets at the end of a period are $x_i = x_i^a + x_i^n - d_i$. They are accumulated from initial net assets A_i, after-tax labor income y_i, government transfers (benefits) $B_i \geq 0$: $x_i^s + x_i^a - d_i = x_i \equiv A_i + y_i + B_i - c_i$, where $A_0 = 0$. Net assets at the start of the next period are

$$A_{i+1} = (1-\kappa^s)R \cdot x_i^s + (1-\kappa^a)R/\Pi_i \cdot x_i^a - (1+\kappa^d)R/\Pi_i \cdot d_i. \qquad (2)$$

The first-order conditions for optimal investments and borrowing are

$$\begin{aligned}
(1-\kappa^s) \cdot \Pi_i \cdot R\beta u'(c_{i+1}) - u'(c_i) + \Lambda_{x_i^s} &= 0 \text{ for } x_i^s \geq 0 & (3)\\
(1-\kappa^a) \cdot R\beta u'(c_{i+1}) - u'(c_i) + \Lambda_{x_i^a} &= 0 \text{ for } x_i^a \geq 0 & (4)\\
(1+\kappa^d) \cdot R\beta u'(c_{i+1}) - u'(c_i) - \Lambda_{d_i} &= 0 \text{ for } d_i \geq 0 & (5)
\end{aligned}$$

where Λ_ξ denotes the Kuhn-Tucker multiplier for a generic condition $\xi \geq 0$.

Two insights follow from (3-5). First, assets are annuitized if $1 - \kappa^a \geq (1-\kappa^s) \cdot \Pi_i$. Since mortality is increasing with age, annuitization starts at some age i_a that satisfies $\Pi_{i_a} \leq \frac{1-\kappa^a}{1-\kappa^s} < \Pi_{i_a+1}$. No annuitization at ages $i < i_a$ implies that lucky heirs receive accidental bequests. Since tracking accidental bequests would complicate the analysis, I assume instead that annuity providers charge $\kappa_i^a = \min\{\kappa^a, 1 - \Pi_i(1-\kappa^s)\}$ to annuitants at age i. This may be interpreted as limit pricing to capture the young. Then without loss of generality, all assets can be treated as annuitized ($x_i^s \equiv 0$) with age-dependent cost κ_i^a.[3] Second, debt and assets are mutually exclusive. Hence $x_i = x_i^a$ if $x_i \geq 0$ and $x_i = -d_i$ if $x_i < 0$.

To streamline the exposition, define the return on net assets conditional on survival by

$$R_{i+1}(x_i) = K_{i+1}(x_i) \cdot \frac{R}{\Pi_i}, \text{ where } K_{i+i}(x_i) = \begin{cases} 1-\kappa_i^a & \text{if } x_i > 0 \\ 1 & \text{if } x_i = 0 \\ 1+\kappa^d & \text{if } x_i < 0 \end{cases}$$

captures financial frictions. Normalizing $K_{i+i}(0) = 1$ is convenient for handling corner solutions below. For reference below, define $K_{[i_1,i_2]} = \prod_{i=i_1+1}^{i_2} K_{i+1}(x_i)$ for general $[i_1, i_2]$ and note that for savers, $K_{[i_1,i_2]} < 1$. The return modifiers $K_{i+1}(x_i)$ and $K_{[i_1,i_2]}$ succinctly summarize the financial frictions as compared to the Yaari (1965) environment.

The dynamics of net assets then reduce to

$$x_i = y_i + B_i - c_i + A_i = y_i + B_i - c_i + R_i(x_{i-1})x_{i-1}, \tag{6}$$

where $A_{i+1} = R_{i+1}(x_i)x_i$. By construction of i_R, income $y_i = w_i L(\tau_i)$ depends on taxes for $i < i_R$, and $y_i = 0$ for $i \geq i_R$. The individual problem is to maximize $E[U]$ by choice of $\{c_i, x_i\}$ subject to (6) and $A_0 = 0$, for given (rational) expectations about policy and wages.

Financial frictions will be critical for the analysis because public pensions involve a round trip of funds, from participants to the government during working age, and in reverse during retirement. For any given amount of net taxes paid or net transfers received, such a system could not be efficient if private savings were frictionless.

1.4 Optimal Fiscal Policy: Taxes and Transfers

The objective is to characterize optimal policy towards a generic cohort t. However, to separate efficiency from redistributional concerns and to ensure common initial conditions across policies, the policy problem is presented as a welfare problem starting a some initial period t_0. Government debt D_{t_0} and the path of non-pension expenditures G_t, $t \geq t_0$, are taken as given.

All spending is financed by labor income taxes. The government's choice variables are taxes $\tau_{i,t} \geq 0$ and transfers $B_{i,t} \geq 0$ applied to cohorts (born in) $t \geq t_0 - i_{\max}$ at ages $i \geq \max\{0, t_0 - t\}$. Taxes and benefits may vary over time and by age (for now; see Section 3 for restrictions). Though individual survival

[3] Modeling accidental bequests would not provide much substantive insight since the probabilities of leaving accidentals bequests $(1-\Pi_i)$ are small, being bounded by the intermediation cost that would trigger annuitization ($1 - \Pi_i < 1 - \frac{1-\kappa^a}{1-\kappa^s} = \frac{\kappa^s+\kappa^a}{1-\kappa^s}$). However, accidental bequests would create cross-sectional heterogeneity and require detailed assumptions which cohorts would be the beneficiaries.

is stochastic, aggregate taxes and benefits are treated as deterministic, invoking the law of large numbers.

The government's primary balance in period t is

$$PB_t = \sum_{i=t-i_R+1}^{t} N_{t-i}\Pi_{[0,i],t-i} w_{i,t} T(\tau_{i,t}) - \sum_{i=t-i_{\max}}^{t} N_{t-i}\Pi_{[0,i],t-i} B_{i,t} - G_t;$$

where revenues are summed over working-age cohorts (born between $t - i_R + 1$ and t) and benefits are summed over all living cohorts. Over time, debt accumulates according to $D_{t+1} = R \cdot (PB_t + D_t)$. The government's intertemporal budget constraint is $D_{t_0} = \sum_{t=t_0}^{\infty} \rho^{t-t_0} PB_t$, where $\rho = 1/R$.

Pareto efficiency requires that policy maximizes a welfare function of the form

$$W_{t_0} = \sum_{t=t_0-i_{\max}}^{\infty} \beta^{t-t_0} \omega_t E_{t_0}[U_t]$$

for some sequence of weights $\omega_t > 0$. Maximizing welfare generally involves intergenerational redistribution, which must be set aside to derive general properties of optimal pensions. Generational accounting is a convenient tool to separate efficiency and redistribution. Let

$$GA_{i_0,t+i_0} = \sum_{i=i_0}^{i_{\max}} \rho^{i-i_0} \Pi_{[i_0,i],t+i_0} [T(\tau_{i,t+i}) w_{i,t+i} - B_{i,t+i}] \tag{7}$$

denote the generational account of cohort t at age i_0, per unit population, which is the present value of future taxes minus transfers. The government's intertemporal budget constraint at time t_0 can then be written as linear combination of generational components:

$$D_{t_0} + \sum_{t=t_0}^{\infty} \rho^{t-t_0} G_t = \sum_{i=0}^{i_{\max}} N_{t-i} GA_{i,t_0} + \sum_{t=t_0+1}^{\infty} \rho^{t-t_0} N_t GA_{0,t}. \tag{8}$$

Equation (8) shows how the exogenous items on the left are allocated to current cohorts (age $i \geq 0$ at $t = t_0$) and to future generations (age $i = 0$ at $t \geq t_0 + 1$).

My generational accounts differ from Auerbach and Kotlikoff's original approach (Auerbach et al 1999) by making a distinction between debt payments and transfers. Debt is a commitment whereas transfers are generally discretionary. The distinction matters here because taxes are distortionary and commitments may force the government to impose taxes.[4]

Using generational accounts, the welfare problem can be defined as maximizing W_{t_0} by choice of $\{c_{i,t+i}, x_{i,t+i}, \tau_{i,t+i}, B_{i,t+i}\}_{i\geq 0, t\geq t_0}$, subject to (7) and (8) with given initial debt D_{t_0} and given spending plans $\{G_t\}_{t\geq t_0}$, subject to (6) and the individual optimality conditions (3)-(5) with given initial assets

[4] See Bohn (1992) for a more complete discussion of governmet accounting and the role of distortionary taxes.

($A_{0,t} = 0$ for cohorts $t > t_0$ born after t_0 and given A_{i,t_0} for cohorts $i \geq 1$ alive at t_0.

Note that with cohort-dependent taxes and benefits, each policy variable enters into exactly one of the generational accounts. Moreover, with tax distortions captured by $L(\tau)$ and $T(\tau)$, policy has no incentive to distort savings, which means (3)-(5) are satisfied automatically and can be omitted. Hence the welfare problem of maximizing W_{t_0} decomposes into three separate problems:

Proposition 1 *Solutions to the welfare problem can be obtained by solving the following three problems:*

(i) For each cohort $t \in [t_0 - i_{\max}, t_0]$, which is alive at time t_0: Maximize $E_{t_0}[U_t]$ by choice of $\{c_{i,t+i}, x_{i,t+i}, \tau_{i,t+i}, B_{i,t+i}\}_{i \geq i_0}$, for a given generational account balance GA_{i_0,t_0} at time t_0.

(ii) For each cohort $t > t_0$: Maximize $E_{t_0}[U_t]$ by choice of $\{c_{i,t+i}, x_{i,t+i}, \tau_{i,t+i}, B_{i,t+i}\}_{i \geq 0}$ for a given generational account balance $GA_{0,t}$ at birth.

(iii) Overall: Maximize W_{t_0} by choice of $\{GA_{i_0,t_0}\}_{0 \leq i_0 \leq i_{\max}}$ and $\{GA_{0,t}\}_{t > t_0}$ for current and future generations, given the relationships between generational accounts and $E_{t_0}[U_t]$ derived in parts (i)-(ii).

Parts (*i*) and (*ii*) are efficiency conditions that must necessarily hold for any Pareto-optimal policy. Only part (*iii*) depends on welfare weights. All results below are implications of parts (*i*) and (*ii*), or extensions thereof, and therefore valid regardless of welfare weights.

1.5 Properties of Efficient Pensions and Taxes

Consider the government's problem of maximizing a generic cohort's utility $E_{i_0}[U]$ at some age $i_0 \geq 0$, given the generational account GA_{i_0}. (General $i_0 \geq 0$ covers parts (*i*) and (*ii*) above, avoiding case distinctions. For cohorts alive at t_0, $GA_{i_0,t}$ is given at $i_0 = t_0 - t > 0$; otherwise $GA_{0,t}$ is given at $i_0 = 0$.)

Since the government maximizes utility, individual rationality is satisfied trivially. Efficient transfers and taxes must satisfy the first order conditions

$$\begin{aligned} \beta^i u'(c_i)\Pi_{[0,i]} - \Lambda_{GA} \cdot \rho^i \Pi_{[0,i]} + \Lambda_{B_i} &= 0 \\ \beta^i u'(c_i)\Pi_{[0,i]} \cdot L'(\tau_i) w_i + \Lambda_{GA} \cdot \rho^i \Pi_{[0,i]} \cdot T'(\tau_i) w_i + \Lambda_{\tau_i} &= 0 \end{aligned}$$

for all $i \geq i_0$, where Λ_{GA} is the shadow value of per-capita resources in (7), and where $\Lambda_{B_i} \geq 0$ and $\Lambda_{\tau_i} \geq 0$ are Kuhn-Tucker multipliers. This implies:

$$\beta^i R^i u'(c_i) = \Lambda_{GA} - \Lambda_{B_i} \cdot R^i / \Pi_{[0,i]} \text{ and} \quad (9)$$

$$\beta^i R^i u'(c_i) = \Lambda_{GA} \cdot \varphi(\tau_i). \quad (10)$$

The term $\varphi(\tau_i) = \frac{T'(\tau_i)}{-L'(\tau_i)} = \frac{1}{1+\Phi(\tau)}$ captures how tax distortions restrict the government's ability to provide pensions.[5] If higher taxes reduce income by a

[5] Deriving (10) involves an intermediate step of showing that $\Lambda_{\tau_i} = 0$. This holds because (9) implies $\beta^i R^i u'(c_i) \leq \Lambda_{GA}$, so (10) always holds with equality for some $\tau_i \geq 0$.

unit, only the fraction $\varphi(\tau_i)$ is collected by the government, where $\varphi < 1$ for $\tau_i > 0$.

Public pensions are systems that first impose taxes and later make transfers to the same people. A positive theory public pensions must explain why such round-tripping of funds can be optimal when taxes are distortionary. Costly asset management and annuitization provide a natural explanation. The optimality condition for savings can be written as

$$R^j \beta^j u'(c_j) = R^i \beta^i u'(c_i)/K_{[i,j]} \tag{11}$$

for any pair of ages $j > i$. For savers, $K_{[i,j]} = \prod_{i=i_1+1}^{i_2}(1-\kappa_i^a) < 1$ implies that the left hand sides of conditions (9) and (10) are rising with age. Hence the conditions for non-zero taxes ($\beta^i R^i u'(c_i) = \Lambda_{GA} \cdot \varphi(\tau_i)$) and non-zero transfers ($\beta^j R^j u'(c_j) = \Lambda_{GA}$) can both be satisfied if taxes and transfers are far enough apart that, for some allocations, $K_{[i,j]} \leq \varphi(\tau_i)$. An immediate second implication is that optimal policy must maximize the time interval between taxes and transfers, which means paying transfers as late as possible.

To be precise, these arguments are formalized as propositions, starting with the optimality of deferral:

Proposition 2 *If public pension benefits are nonzero at some age $i_P \geq i_R$, then the pension pays for all consumption in all subsequent periods. Private assets optimally decline to zero in the period when pension benefits start.*

Formally define the *pension age* $i_P = \min\{i \geq i_R : B_i > 0\}$ as the age at which pensions start. Then Prop.2 specifies that $B_i = c_i$ for all $i > i_P$ and $x_i = 0$ for $i \geq i_P$. Private wealth declining to zero at i_P means $B_{i_P} = c_{i_P} - A_{i_P} \leq c_{i_P}$. The age i_P is implicitly determined by the cohort's generational account, which determines the available present value of transfers.[6]

Prop.2 is sharply at odds with the conventional wisdom that retirees should tap a mixture of private and public funds throughout their retirement. Efficient transfers are either zero or 100 percent of consumption, except for at most one transition period.

Prop.2 applies if benefits are non-zero. The following propositions provide existence conditions for non-zero benefits and characterize the time gap between the end of contributions and the start of pensions.

Proposition 3 *Suppose there is an age $i \leq i_R - 1$ such that savers set $x_j > 0$ for all $j \in [i, i_R]$. If $\prod_{i=i_1+1}^{i_{\max}}(1-\kappa_i^a) = K_{[i,i_{\max}]} < \varphi(\tau_i)$, then policy is inefficient. A marginal increase in B_k for some $k \geq i_R$ would increase welfare.*

[6] The borrowing friction κ^d is needed here to make the optimal policy unique. If $\kappa^d = 0$, individuals and the government would be indifferent between the transfers described in Prop.2 and transfers with the same present value paid even later in life, because the same consumption stream could be financed by borrowing. As extreme example, consider a pension system promising a huge payment to survivors at age $i_{\max}$ and nothing for $i < i_{\max}$. If individuals can borrow against against $B_{i_{\max}}$ at the same interest rate at which pensions are discouted, and if $B_{i_{\max}}$ has the same present value as the transfers under Prop.2, individuals can consume the same as under Prop.2. Such perfect borrowing against future pensions is unrealisting and the possibility would sidetrack the analysis. Any infinitesimal $\kappa^d > 0$ breaks the tie; κ^d has no other function in the model.

Prop.3 specifies under what conditions financial frictions provide a rationale for public pensions. Zero public pensions cannot be optimal if savings frictions are greater than the excess burden of taxes. (Intuitively, optimality requires $K_{[i,i_{\max}]} \approx \varphi(\tau_i)$, but this cannot be stated as equality because $K_{[i,i_{\max}]}$ is not continuous in x_j.)

Note that without public pensions, non-zero consumption would require $x_i > 0$ for all $[i_R - 1, i_{\max}]$; hence $K_{[i_R-1,i_{\max}]} < \varphi(\tau_{i_R-1})$ is a sufficient condition for $B_{i_{\max}} > 0$. Conversely:

Proposition 4 *If tax rates are high enough in the period before retirement that $\varphi(\tau_{i_R-1}) < K_{[i_{R-1},i]}$ for any $i \geq i_R$, then optimally $B_j = 0$ for all $j \in [i_R, i]$.*

Under the conditions of Prop.4, there is a time gap between retirement and the start of pensions (if any).[7] At least the initial years of retirement are financed by private saving. In particular, $B_{i_R} > 0$ cannot be optimal unless taxes at age $i_R - 1$ are so close to zero that the loss from excess burden, $1 - \varphi(\tau_{i_R-1})$, is no greater than the savings friction $\kappa^a_{i_{R-1}}$ for one year.

Proposition 5 *Optimal tax rates satisfy $\varphi(\tau_{i-1}) = K_i(x_{i-1}) \cdot \varphi(\tau_i)$, provided both are nonzero.*

Since $K_i(x_{i-1}) \in [1-\kappa^a_i, 1+\kappa^d]$ bounded around one, optimal taxes should be nearly constant over time. For savers, $\varphi(\tau_{i-1})/\varphi(\tau_i) = (1 - \kappa^a_i) < 1$ and $\varphi' < 0$ imply a declining sequence of tax rates. This is broadly consistent with Fenge et al (2006). Moreover, Prop.5 reinforces that the pension age cannot equal the retirement age except under restrictive conditions: from Prop.4, $B_{i_R} > 0$ is optimal only if taxes at age $i_R - 1$ are near zero, and then Prop.5 implies that tax rates must be low in prior years.

In combination, Prop.4 and 5 imply a significant time gap between taxes and transfers – either a significant time interval between retirement age and pension age, or near-zero taxes before retirement, or both. Additional features of optimal policy are examined quantitatively in Section 4 below.

To conclude, the same savings friction that justifies the existence of public pensions also implies that benefits should be paid at the end of the life cycle. The latter is therefore an intrinsic property of optimal pensions.

2 A Finance Interpretation

This section provides a finance intuition for the results above and explains why both intermediation cost and tax distortions are needed to obtain a mix of public and private retirement funding.

[7] Note that Pareto-optimality does not imply contributary pensions for all cohorts. If ω_t is very small, cohort t may face taxes close to $\hat{\tau}$, so $\varphi(\tau_i) < K_{[i,j]}$ may apply at all ages and $B_j = 0$. If ω_t is very high, transfers may be positive during working age, so $\beta^i R^i u'(c_i) = \Lambda_{GA}$ and $\tau_i = 0$ at all ages.

From parts (i) and (ii) of Prop.1, optimal policy maximizes each cohort's utility conditional on the optimal (from part (iii)) generational account balance GA_0. The problem of maximizing $E[U]$ by choice of $\{c_i, B_i, \tau_i, A_i\}$ for given GA_0 has a recursive representation that will also be convenient for the numerical analysis below.

From the taxpayers' perspective, a positive generational account is a liability and a negative generational account a net asset. To focus on surviving members of a cohort and their retirement benefits, let

$$GB_i = \left(\sum_{j=i}^{i_{\max}} \rho^{j-i}\Pi_{[i,j]}B_j\right) - \left(\sum_{j=i}^{i_{\max}} \rho^{j-i}\Pi_{[i,j]}T(\tau_j)w_j\right) = -\frac{GA_i}{\Pi_{[0,i]}} \tag{12}$$

denote the present value of net transfers an individual of cohort t alive at age i_0 will receive from the government – the generational net asset, or "generational benefit" for short. Over time, GB_i accumulates according to

$$GB_{i+1} = \frac{R}{\Pi_i} \cdot [GB_i + T(\tau_i)w_i - B_i]. \tag{13}$$

Importantly, the accumulation incurs no cost of asset management or annuitization.

The cohort's problem is then equivalent to a investment problem with two funds, fund A and fund GB. The fund balances (A_i, GB_i) serves as state variables. The Bellman equation with planning horizon $k = i_{\max} - i \geq 1$ is

$$V_{k+1}(A_i, GB_i) = \max\{u(c_i) + \beta\Pi_i V_k(A_{i+1}, GB_{i+1})\}$$

subject to (13), $A_{i+1} = R_{i+1}(x_i)x_i$, $x_i = A_i + B_i - c_i + (L(\tau_i)w_i - \mu_i)$, and the inequalities $c_i \geq 0$, $B_i \geq 0$, $\tau_i \geq 0$, and $w_i L(\tau_i) \geq \mu_i$. Since all assets are consumed at $i_{\max}$, the recursion ends with $B_{i_{\max}} = GB_{i_{\max}}$ and $V_1(A_{i_{\max}}, GB_{i_{\max}}) = u(A_{i_{\max}} + GB_{i_{\max}})$.

This investment analogy does not assume an individual ability to make or withdraw contributions from the generational account. Once policy is determined, "deposits" into fund GB are collected as taxes. Collecting taxes $T(\tau_i)w_i$ reduces after-tax income from $w_i L(0)$ to $w_i L\{T^{-1}(T(\tau_i))\}$, which is a loss greater than $T(\tau_i)w_i$. The difference–the excess burden–acts like a one-time purchase fee for fund GB. Thus the choice between funds A and GB is equivalent to the choice between a no-load investment with annual fees (A) versus a mutual fund with up-front load and zero annual cost (GB).

Note that pay-as-you-go financing is an advantage in this context. If public pensions were fully funded, the model's assumption of no annual management cost for fund GB would at best be an approximation justified by economies of scale. However, most governments carry substantial debt and most public pensions operate as pay-as-you-go systems. On the margin, taxes going into such a system reduce public debt. Hence there are no assets to manage. The reduced debt may even reduce the government's underwriting costs in the bond markets.

The analogy to load versus no-load funds explains intuitively why optimal retirement planning draws down fund A before tapping fund GB (Prop.2), why the tradeoff involves a comparison between excess burden and annual costs (Prop.3), and why there must be a time gap between deposits into and withdrawals from fund GB (Prop.4-5).

3 Optimal Pensions with More General Assumptions

This section examines the structure of public pensions under several sets of alternative assumptions. In each case, assumptions are made to ensure that one form of retirement savings does not dominate the other.

3.1 Limited Commitment: Earnings-linked pensions

Public pensions in many countries are linked to individual earnings. Hence it is worth documenting that the main results above generalize to a setting with earnings-linked pensions, provided the linkage is not so strong that contributions become voluntary and non-distortionary.

Let $\phi_i \leq \bar{\phi} < 1$ denote the degree of earnings-linkage at age i, which is the present value of incremental pensions promised by the pension system per unit of payroll taxes. The upper bound $\bar{\phi}$ is meant to capture taxpayers' belief that government's commitment to differentiate pensions based on prior earnings is limited.[8]

To keep track of earnings-based promises, let each cohort's transfers be divided into an individual, earning-based component $B_i^e \geq 0$ with present value GB_i^e and regular (unearned) transfers $B_i^u \geq 0$ with present value GB_i^u. Each period, a share ϕ_i of payroll taxes $\tau_i w_i l_i$ is credited to specific taxpayers (to GB_i^e). The remainder is credited to GB_i^u. The resulting dynamics are

$$GB_{i+1}^e = \frac{R}{\Pi_i} \cdot [GB_i^e + \phi_i \tau_i w_i l_i - B_i^e] \tag{14}$$

$$GB_{i+1}^u = \frac{R}{\Pi_i} \cdot [GB_i^u + (1-\phi_i) \tau_i w_i l_i - B_i^u] \tag{15}$$

By construction, $GB_i^e + GB_i^u = GB_i = -GA_i/\Pi_{[0,i],t}$. Hence the analysis fits into the fiscal framework of Section 1.

[8] A specific bound that ensures non-zero private savings is derived below. This is an issue if the government could credibly promise a full refund of payroll "contributions" with interest ($\phi_i = 1$), which could dominate all savings. Commitment is required because if the government re-optimized welfare when pensions are due, it would disregard past earnings. Hence taxpayers may dismiss promises of $\phi_i = 1$ as time-inconsistent. The ϕ_i are modeled as age-dependent because this is empirically relevant. For example, if pensions are based on career-average earnings, early contributions have much lower present value that payments close to retirement, as $\phi_i = \phi_{i+1}/(R/\Pi_i) < \phi_{i+1}$ would be increasing. If pension credits are linked to an index (e.g, aggregate wages in the U.S.), ϕ_i varies whenever growth in the index differs from the discount rate.

Regular and earnings-linked pensions are equivalent except that the earnings-linkage reduces distortions to labor supply. Hence the decomposition of the overall welfare problem into generational components still applies, as in Prop.1. The only new element is that efficiency can be improved by minimizing tax distortions, which is achieved by maximizing the value that taxpayers assign to their earnings-based pension. Due to financial frictions, the individual value of future transfers tends to be greater than the cost to the government. The value-added turns out to depend on the payout policy, which can be exploited to characterize optimal policy, as follows.

Let v_i denote the value of a marginal increase in GB_i^e at age i. Labor supply maximizes $y_i + v_i\phi_i\tau_i w_i l_i$, which implies the optimality condition $1-\tau_i - H'(1-l_i) + v_i\phi_i\tau_i = 0$ or $l_i = 1-(H')^{-1}[1-(1-v_i\phi_i)\tau_i]$. This is identical to labor supply in the basic model with tax rate $(1-v_i\phi_i)\tau_i$. Thus the earnings-linkage reduces tax distortions.

Payout policy can be defined by the shares of earned benefits $GB_{i_R}^e$ at retirement paid out at different dates. Let $b_j = \rho^{j-i_R}\Pi_{[i_R,j]}B_j^e/GB_{i_R}^e$ denote the share paid at age $j \geq i_R$. where $\sum_{j=i_R}^{i_{\max}} b_i = 1$. Since individuals discount future transfers by $R_{k+1}(x_k)$, they assign value

$$v_{i_R} = \sum_{j=i_R}^{i_{\max}} b_j \prod_{k=i_R}^{j} (\frac{R/\Pi_k}{R_{k+1}(x_k)}) = \sum_{j=i_R}^{i_{\max}} (b_i/K_{[i_R,j]}) \tag{16}$$

to a unit increase in $GB_{i_R}^e$. Similar discounting during working age implies $v_i = v_{i_R}/K_{[i,i_R]}$ for $i < i_R$, so all v_i are proportional to v_{i_R}.

Minimizing tax distortions by choice of ϕ_i and b_j requires maximizing v_i for any given ϕ_i, which in turn requires maximizing v_{i_R} in (16).[9] Since for savers $K_{[i_R,j]}$ is less than one and strictly decreasing in j, deferring payouts increases v_{i_R} for as long as individuals have positive private savings.

If a cohort receives no unearned transfers in retirement, this argument implies a unique optimal payout policy for B_j^e: No benefits are paid until a pension age i_P. The pension age is determined by the conditions that $B_j^e = c_j$ must pay for *all* consumption at ages $i > i_P$ and that total payments add up to $\sum_{j\geq i_P}^{i_{\max}} b_j = 1$, which implies an amount $B_{j_P}^e \leq c_{j_P}$. This is identical to the optimal policy for regular transfers derived in Section 1.

If a cohort receives unearned transfers in addition to earnings-linked benefits, the argument for deferral applies to total transfers, leaving indeterminate which type of benefits is paid in which period, provided the present values add up to $GB_{i_R}^e$ and $GB_{i_R}^u$, respectively. The indeterminacy arises because financial frictions vanish once private savings are exhausted. To summarize:

Proposition 6 *In a system with earnings-linked pensions, there is at most one transitional period in which consumption is financed from a mix of public benefits*

[9] The focus here is on payout policy for given ϕ_i, because the optimal choice of ϕ_i involves a non-trivial tradeoff between minimizing tax distortions (which suggests high ϕ_i) and funding the generational account (which is proportional to $1-\phi_i$). Hence there are no general results about optimal ϕ_i.

(regular or earnings-linked) and private savings. Before the transitional period, public benefits are zero. After the transitional period, public benefits pay for all consumption.

The central result that pensions should be paid late in life thus extends to earnings-linked pensions. The argument is strengthened because deferral provides value-added not only by avoiding financial friction but also by reducing tax distortions.

Note that the existence of private savings is non-trivial in this setting. If $v_i\phi_i \geq 1$, "taxes" could be reinterpreted as voluntary "contributions" to a public pension system that dominates private savings. To rule out allocations without any private savings (in which case Prop.6 would still apply, but trivially with $i_P = i_R$), one must assume that the government's credibility to return contributions is limited in the sense that the earnings-linkage is bounded by $\phi_i \leq \bar{\phi}$ where $\bar{\phi} < 1/\max\{v_i\}$.

3.2 Cross-Sectional Heterogeneity

Now suppose each cohort has a distribution F of types η that differ by labor productivity (and possibly other characteristics). Let $w_{i,t}$ now denote a cohort's average wage and let types be labeled by their productivity relative to the mean, so type η earns $\eta w_{i,t}$ and $\int \eta dF(\eta) = 1$. For simplicity, assume relative earnings are fixed over the life-cycle.[10]

Individuals maximize utility $E_t[U_t(\eta)]$, earn income $y_{i,t+i}(\eta)$ as in (1), and accumulate assets according to

$$\begin{aligned} x_{i,t+i}(\eta) &= A_{i,t+i}(\eta) + B^e_{i,t+i}(\eta) + B^u_{i,t+i}(\eta) - c_{i,t+i}(\eta) + y_{i,t+i}(\eta) \\ A_{i,t+i}(\eta) &= R_{i,t+i}[x_{i-1,t+i-1}(\eta)] \cdot x_{i-1,t+i-1}(\eta) \end{aligned}$$

Assume a welfare function $W_{t_0} = \sum_{t=t_0-i_{\max}}^{\infty} \beta^{t-t_0} \cdot \int \omega_t(\eta) E_{t_0}[U_t(\eta)] dF(\eta)$ with arbitrary weights $\omega_t(\eta) > 0$. Let taxes τ_t be linear and assume an earnings linkage $\phi_i(\eta)$ that may depend on type (and on time, though time-subscripts are omitted). Regular transfers $B^u_i(\eta)$ may also depend on age and on type. Earning-linked benefits are again defined by payout parameters $b_j(\eta)$, so $B^e_j(\eta)$ $= b_j GB^e_{i_R}(\eta) R^{j-i_R}$, where $GB^e_{i_R}(\eta)$ is accumulated as in (14).

In this setting, optimal pension ages and amounts will generally vary across types, but one can show that for each type η, benefits have the same structure as in the homogenous agents case. A proof is omitted since the result is a special case of Prop.7 below. The point of this section is to note a key difference between taxes and benefits: Wheras taxes are linked across types, pension benefits can optimized separately for each type, given the present value allocated to the respective type.

[10] One-dimensional heterogeneity should suffice here, as this section is meant to sketch how the basic model generalizes, not to maximize generality.

3.3 Optimal Public Pension Payouts in General

Many assumptions above were motivated by the ambition to model pensions together with the tax system required to finance them. More sweeping results can be obtained if one takes the tax system as given and simply assumes that a certain present value of pension payments is part of the optimal policy. One finds:

Proposition 7 *Regardless of the tax system applicable during working age, suppose optimal policy includes paying non-zero transfers $GB_{i_R}(\eta) > 0$ to a particular type of retiree (η), and suppose type η has savings $A_{i_R}(\eta) \geq 0$ at retirement. Then a necessary conditions for optimality is that pension benefits maximize expected retiree utility $U^R(\eta) = \sum_{i=i_R}^{i_{\max}} \beta^i I_{i,t+i}^{alive} u(c_i)$ subject to given $A_{i_R}(\eta)$ and $GB_{i_R}(\eta)$. If benefits are nonzero at some age $i_P(\eta) \geq i_R$, the pension pays for all consumption in all subsequent periods.*

Prop. 2, 6 and 7 express the same principal result in complementary ways. Prop.7 makes no assumptions about taxes but is silent about conditions under which non-zero transfers are part of an optimal policy. Prop.2 and 6 invoke specific assumptions on taxes to show that $GB_{i_R}(\eta) > 0$ is part of an optimal policy under plausible assumptions.

The pension ages $i_P(\eta)$ generally differ cross-sectionally. While it is beyond the scope of this paper to characterize cross-sectional policy differences, standard utilitarian logic suggests more generous benefits for low-productivity than for high-productivity types. This would imply that optimal retirement ages $i_P(\eta)$ are increasing in productivity. This intuition would be strengthened if income were positively correlated with longevity (not modeled here). Importantly, regardless of welfare weights, once pension benefits start for type η, they optimally cover all consumption in all subsequent periods.

Note that the argument for deferring pension is a marginal argument. This implies:

Proposition 8 *Starting from any policy that violates Prop.7, welfare is increased by any shift of transfers to a later date at which individuals use private savings for consumption, holding $GB_{i_R}(\eta)$ constant.*

Prop.8 allows an evaluation of real-world pension reforms. One strong implication is that if pension parameters must be changed to reduce cost, an increase in the pension age is always preferable to a cut in replacement rates.

Prop.7-8 apply to models with non-linear taxes, such as Mirrleesian taxation. In dynamic Mirrleesian taxation, incentive constraints are sometimes framed in terms of promised utility. Prop.7 could easily be restated in these terms: suppose type η enters retirement with promised utility $U_t^R(\eta)$ and the government seeks to deliver this utility at minimal cost. Since minimizing $GB_{i_R}(\eta)$ for given $U_t^R(\eta)$ and $A_{i_R}(\eta)$ is the dual of the problem in Prop.7, the optimal payout policy has the same properties.

Prop.7-8 also apply under arbitrary restrictions on tax rates, including uniformity across age/cohorts. This is important to clarify that the age-specific taxes assumed in Section 1 are without loss of generality. To make this explicit, the appendix shows how Prop.1 can be modified with age-independent taxes and why Prop.2 still applies.

One concern about pension deferrals may be that promises of extremely high transfers to very high income retirees might not be credible. If so, optimal policy could be modified straightforwardly. For example, if there were a maximum credible transfer $B_{\max}$, the optimal policy for types consuming $c_i(\eta) > B_{\max}$ would be to set $i_P(\eta)$ such that $GB_{i_R}(\eta)$ is paid out with $B_i(\eta) = B_{\max}$ for $i > i_P(\eta)$.

The central intuition is that holding constant the present value of transfers from the government's perspective, a marginal deferral of payments will increase a pension's value to individuals who face costs of asset management and annuitization. This argument applies until benefits are deferred to the maximum extent possible. The argument does not involve or depend on cross-sectional heterogeneity or redistributional issues. This motivates why the basic model assumes homogenous agents for simplicity.

3.4 Endogenous Retirement Age

Suppose the fixed cost of market work increases gradually with age so $L(\hat{\tau}) < \mu_i/w_i \leq l(0)$ applies for some age range $[i_{R1}, i_{R2})$. Then the individually-rational retirement age depends on taxes. Because retirees pay no taxes, $i_R = i_{R2}$ is optimal. Hence optimal labor taxes must satisfy $\tau_i \leq L^{-1}(\mu_i/w_i)$ to avoid premature retirement. Apart from this additional constraint on taxes, optimal policy is determined as in the previous sections.

In many countries, pensions systems seems to provide inefficient incentives to retire by providing pension benefits too early (Gruber and Wise 2008). This problem is difficult to avoid in traditional pension systems that make payments starting at retirement. Even if benefits are initially set to start at $i_P = i_R$, improvements in health may increase i_R, and then pensions end up starting prematurely. A pension system with gap between the retirement age i_R and the pension age i_P, as proposed in this paper, would eliminate the problem of inducing premature retirement.

4 Illustrative Welfare Calculations

This section presents numerical analysis to document the quantitative relevance of sequential pension funding. The scenarios are roughly calibrated to current U.S. and European conditions but use round numbers for clarity and abstract from cross sectional heterogeneity and from earnings linkages. The focus is on comparisons across scenarios and their implications for policy responses to changes in the economic and demographic environment.

4.1 Benchmark: Conventional Public Pensions

In most developed countries, public pension systems pay benefits early enough that virtually all members draw benefits as soon as they retire. Hence conventional pensions are interpreted here as systems with pension eligibility that starts at retirement, setting $i_P = i_R$.[11]

Public pensions typically pay constant or formula-fixed benefits. Formulaic adjustments, such as growth factors, would needlessly complicate the analysis. Conventional public pensions are therefore best interpreted as a *fixed-parameter system* that pay a constant benefit $B_i = \bar{B}$ for all $i \geq i_R$ and impose a constant tax rate $\tau_i = \bar{\tau}$ for all $i < i_R$.

For each specification of the model, welfare gains will be computed by comparing the optimal pension system to the optimal fixed-parameter system. That is, parameters $(\bar{\tau}, \bar{B})$ that are selected to maximize generational utility.

4.2 Baseline Assumptions

Survival rates are taken from the U.S. Social Security Administration for the 2010 cohort, averaged over males and females. Assume working age starts at age 20, retirement at age 66, and the life span is 120 years. Population is constant, $N_t = 1$. Bonds pay $r = 0.02$ and the rate of time preference equals the bond rate ($\beta = \rho$), which means consumption would be constant in the absence of financial frictions. Assume $\kappa^a = 0.01$ as maximum annuitization cost, which becomes effective at age 60. Assume negligible borrowing costs, $\kappa^d = .0001$, non-zero only to avoid indeterminacies that would arise from costless borrowing against pensions. Wages w_i are proxied by an age-earning profile taken from Rupert and Zanella (2015), normalized to one at peak earnings (age 50). Work costs are $\mu_i = 0$ for $i < i_R$.

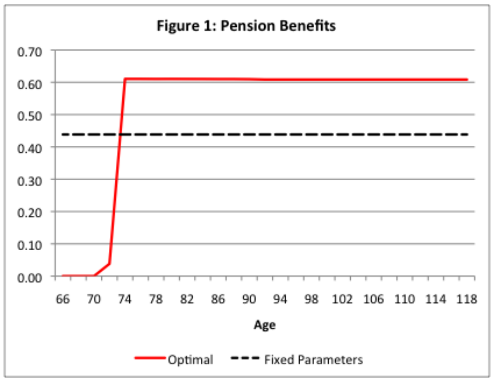
Home production is $H(1-l) = -hl^{1+\varepsilon}$ with parameters $h = 0.5$ and $\varepsilon = 2$.[12] This implies a labor supply elasticity of $1/\varepsilon = 0.5$ and $\hat{\tau} = \frac{\varepsilon}{1+\varepsilon} = 2/3$. Assume power utility over market consumption, $u(c) = \frac{1}{1-\gamma}c^{1-\gamma}$ with $\gamma = 2$.

All calculations are done with one period representing two calendar years, starting at age 20 (i.e., natural ages 20-21 are $i = 0$ in the calculations, ages 22-23 are $i = 1$, etc.). However, in the figures and tables below, age is converted back into natural units to facilitate the interpretation.

The generational account is calibrated so that the initial balance per-capita is equal to the revenue from a constant tax rate $\tau_{other} = 0.1$. Non-pension expenditures influence the optimal level of pensions by influencing the marginal excess burden of taxes; to document this dependence, the sensitivity analysis will consider alternative values of τ_{other}.

[11] This is arguably a sympathetic interpretation, because there is considerable evidence that pensions systems may encourage premature retirement (suggesting $i_e < i_R$), e.g., by making benefits conditional on not working; see Gruber and Wise (2008).

[12] The normalization $H(1) = 0$ implies $H(1-l) < 0$ for $l > 0$. If the latter is objectionable (counterintuitive), one could redefine $H(1-l) = h - hl^{1+\varepsilon}$ and $u(c) = \frac{1}{1-\gamma}(c-h)^{1-\gamma}$. Then home production would be non-negative and provide subsistence consumption; the model would remain unchanged otherwise.



4.3 Baseline Results

Figures 1-6 show the time paths of the main variables in the baseline scenario. Since peak earnings are normalized to one, all real values should be interpreted as fractions or multiples of two-year peak earnings.

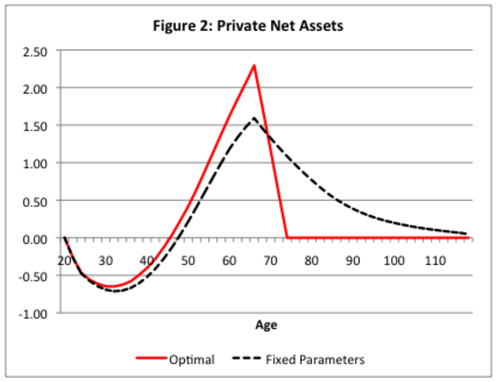
Optimal pensions (figure 1) are zero until age 70 and jump to 0.61 at age 74. The transitional value of 0.04 at age 72 is best interpreted as full benefit for a fractional period. Since 0.04 for two years is equivalent to 0.61 for $\frac{4}{61}$ of the 2-year period and 0 for the remainder, optimal benefits can be expressed as zero until age 73.9, and 0.61 thereafter. All pension ages below are expressed in this form.

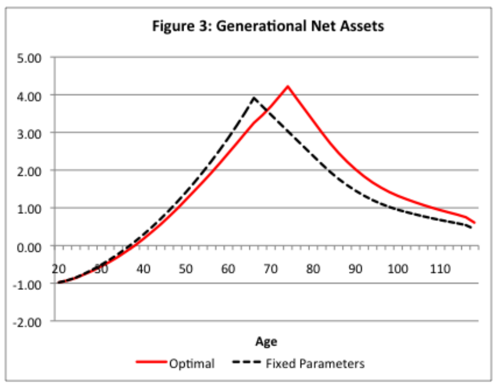
The pension with optimally-chosen fixed parameters offer a constant benefit of 0.44, starting earlier but providing far less at older ages than the optimal pension.

Figure 1 illustrates how optimal benefits jump from zero to a value greater than under a fixed parameter pension. The exact timing of the jump depends on model parameters—see sensitivity analysis below. A specific scenario is nonetheless instructive to convey the model's full implications.

Figures 2-4 show private net assets (A), generational benefits (GB) – here called generational net assets since they act like assets – and total net assets ($A+GB$). During working age, asset holdings in the optimal and fixed-parameter cases follow similar paths. In retirement, the optimal plan draws on private assets first and depletes them by age 74. Until then, generational net assets accumulate. In contrast, retirement financing in a fixed-benefit setting uses all funding sources in rough proportions. The resulting consumption profiles (figure 5) are similar until the optimal pension payments start. Then optimal consumption is constant and ends up significantly higher than with a fixed pension.

Optimal tax rates (figure 6) are fairly stable over the life cycle, though declining as retirement approaches. The intuition is that imposing taxes early maximizes the time interval between taxes and transfers over which the excess burden is amortized. For comparisons below, the overall level of optimal taxes is usefully summarized by an average tax rate τ_{av}, defined as constant rate





that would yield the same lifetime revenue; here $\tau_{av} = 0.238$. In this scenario (though not in general), fixed pensions have a higher tax rate ($\bar{\tau} = 0.27$), which is due to the cost of paying benefits for a longer period.

The welfare gain from optimal as compared to fixed-parameter pensions is about 0.6 percent of lifetime consumption. Compared to having no public pensions at all, optimal pensions provide a welfare gain of 1.8 percent. Fixed-parameter pensions would provide a welfare gain of only 1.2 percent. The main differences arise during from optimal payouts during retirement: a system with a constant tax rate and variable benefits would provide a welfare gains of 1.7 percent, which is close to the fully optimal system. The age-dependent tax rates contribute only 0.1 percent.

Another perspective on welfare is to consider an unexpected reform at the time of retirement, taking taxes and asset positions as given (in the spirit of Section 3.3). Shifting from fixed-parameter to optimal benefits would provide welfare gains equal to 2.4 percent of retirement consumption. These are substantial gains.

4.4 Sensitivity Analysis

This section examines the ramifications of alternative parametric assumptions.

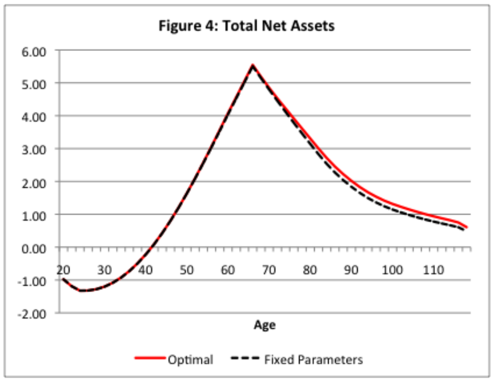


Figure 4: Total Net Assets
6.00
5.00
4.00
3.00
2.00
1.00
0.00
-1.00
-2.00
20 30 40 50 60 70 80 90 100 110
Age
Optimal
Fixed Parameters

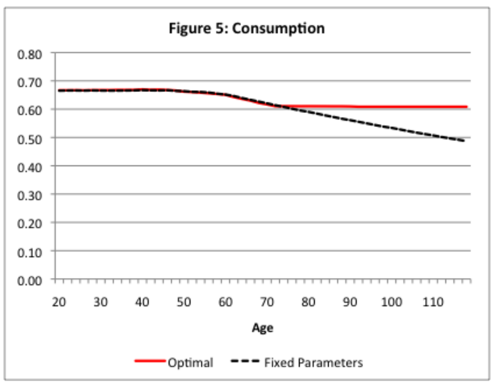


Figure 5: Consumption
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
20 30 40 50 60 70 80 90 100 110
Age
Optimal
Fixed Parameters

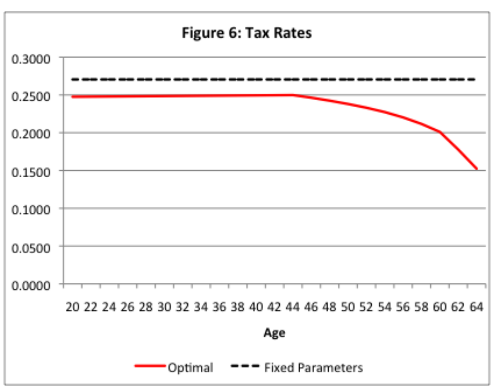


Figure 6: Tax Rates
0.3000
0.2500
0.2000
0.1500
0.1000
0.0500
0.0000
20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64
Age
Optimal
Fixed Parameters

Table 1: Sensitivity Analysis

Setting	i_P	τ_{avw}	$B_{i\|i>i_P}$	$\bar{\tau}$	$\bar{B}$
Baseline	73.9	0.24	0.61	0.27	0.44
$\kappa^a = 0.5\%$	77.5	0.20	0.62	0.20	0.28
$\kappa^a = 5\%$	73.9	0.24	0.60	0.31	0.53
$\varepsilon = 1$	79.8	0.17	0.56	0.16	0.18
$\varepsilon = 3$	71.2	0.28	0.66	0.33	0.60
$\gamma = 4$	73.2	0.25	0.63	0.27	0.44

For annuitization cost, consider a low-cost setting with $\kappa^a = 0.005$ and a high-cost setting with $\kappa^a = 0.05$. For labor supply, consider a higher or lower labor supply elasticity by setting $\varepsilon = 1$ or $\varepsilon = 3$. For preferences, consider a reduced elasticity of intertemporal substitution by assuming $\gamma = 4$. Table 1 shows the optimal and fixed-parameter pension systems.

Optimal benefits stay remarkably stable across scenarios, again reflecting the optimality of covering end-of-life consumption fully. A lower (or higher) cost of annuitization reduces (increases) the value of public pensions. In the optimal system, this leads to a higher (lower) pension age, whereas in a fixed parameters, benefits are scaled down (up). A higher labor supply elasticity increases (decreases) the excess burden of pensions and other taxes, leading to later pension eligibility in the optimal system and to lower benefits in a fixed-parameter system; a reduced lower labor supply elasticity has the opposite effects. A reduced elasticity of intertemporal substitution triggers slightly higher pension benefits in the optimal system, while the fixed-parameter system remains essentially unchanged. The intuition is that optimal consumption is less variable over the life cycle when the elasticity of intertemporal substitution is low, and this requires higher old-age consumption.

Overall, the optimal "size" of the pension system is clearly sensitive to various institutional and behavioral parameters. With fixed benefits and taxes, variations in size translate into higher or lower optimal benefits. In contrast, when the pension age is chosen optimally, variations in "size" and in taxes translate into earlier or later pension ages. The benefit amounts vary only to the extent that optimal consumption varies. For all scenarios considered here, the pension age appears to be in the early- or mid-70s, which is higher than the eligibility age of most existing pension systems.

Put differently, if one views a country's observed mix of private and public retirement funding as revealing the size of the pension system appropriate for the country's institutional and behavioral parameters, then the analysis here implies that welfare could be improved if one took the present value of current pensions (which typically start at retirement) and reallocated the funds to provide higher public pensions at a higher starting age.

Table 2: Optimal pensions when mortality declines

Mortality	i_R	i_P	τ_{av}	$B_{i\|i>i_P}$	$\bar{\tau}$	$\bar{B}$
Baseline	66	73.9	0.238	0.610	0.270	0.44
Improved	66	75.9	0.239	0.604	0.267	0.39
Improved	68	76.1	0.235	0.607	0.273	0.45

4.5 Optimal Responses to Increasing Longevity

This section returns to the baseline parameters and uses the model for policy analysis. One issue is the optimal response to increasing longevity.

The answers turn out to depend significantly on the question to what extent the changes driving longevity also improve individuals' ability to work longer. Table 2 compares the baseline to two alternative scenarios. Both assume a 25 percent reduction in mortality, which means life expectancy at age 20 increases by about 3.3 years. One scenario assumes an unchanged retirement age of $i_R = 66$, the other assumes that improved mortality also increases the optimal retirement age to $i_R = 68$.

If these changes occur unexpectedly, intergenerational risk sharing suggests that young and future cohorts should at least partially insure older cohorts (see Bohn 2006). As polar cases, consider (*i*) an unchanged generational account and (*ii*) adjustments in the generational account GA_{i_R} at retirement that keep benefits unchanged.

Table 2 shows for each scenario the optimal pension age, the average tax rate in the optimal system,[13] optimal benefits for $i > i_P$, and for comparison, the tax rates and benefits in a fixed-parameter system.

If mortality is reduced and the retirement age stays unchanged at 66, the optimal benefit amount and optimal taxes remain almost unchanged. The main adjustment is that the pension age rises from 73.9 to 75.9. The economic intuition is that an essentially unchanged lifetime labor income is stretched over a longer horizon, so consumption in every period decreases slightly. Optimal pensions must cover end-of-life consumption fully, so the optimal pension amount must not decline more than consumption. In a fixed-parameter setting, in contrast, longevity can be absorbed only by higher taxes and/or lower benefits. In the scenario here, an exogenous pension age of 66 implies a cut in benefits from 0.44 to 0.39.

In the scenario with retirement at 68, the optimal pension system offers slightly greater benefits are an even higher pension age. The intuition is that two more years with labor income raise the optimal consumption profile at all ages, though only slightly. The optimal pension age rises because the benefits to support higher consumption can only be paid for a somewhat shorter period. In a fixed-parameter system, the pension age would simply rise to 68 and benefits would be similar to the baseline; thus benefits are sensitive to the retirement

[13] Optimal taxes vary as shown in Figure 6. To express the level of taxation concisely, the table reports the tax rates that would raise the same present value of revenues as the optimal tax system.

Table 3: Optimal pensions in relation to other taxes

Setting	τ_{other}	GA_0	i_P	τ_{av}	$B_{i\|i>i_P}$	$\bar{\tau}$	$\bar{B}$	Δc
High-tax	0.20	1.85	79.3	0.29	0.54	0.28	0.18	0.7
Baseline	0.10	0.98	73.9	0.24	0.61	0.27	0.44	0.6
Low-tax	0	0	70.0	0.18	0.68	0.22	0.64	0.2

age, which is inefficient. Intuitively, adding two more years to a 46-year career is a minor change in economic conditions that should not trigger major policy changes.

Turning to a slightly different setting, suppose the decline in mortality is unexpected and occurs at the time of retirement. The cost of financing the original consumption stream would increase by about 9 percent. If benefits are kept unchanged, the present value of optimal benefits would increase by 14.6 percent and cover about 96 percent of the funding needs, keeping consumption essentially unchanged. The intuition is that 96 percent of the mortality improvements occur at the end of life when optimal pensions pay for all consumption. In a fixed-parameter system, the present value of benefits would increase by about 9 percent, roughly in proportion to the present value of consumption. Per-period consumption would have to decline by about 2 percent.

In summary, this scenario shows that the optimal pension system can provide almost complete insurance against longevity risk, and it can provide significantly more intergenerational sharing of such risks than a fixed-parameter system.

4.6 Optimal Responses to Budget Problems and Low Birth Rates

The optimal taxes and pensions for each generation are affected by the government's budget position through the generational account balance $GA_{0,t}$ that is imposed on cohort t as it enters the labor force. Within the overall welfare problem (part *iii*), higher non-pension expenditures and higher initial debt would tighten the budget constraint (8), and – ceteris paribus – require an increase in $GA_{0,t}$ for all cohorts.

If future generations are expected to make positive contributions to the budget (if $GA_{0,t} > 0$), a small cohort size would also tighten (8). Then low birth rates cause budgetary problem and imply an increase in the per-capital generational account for all cohorts.

Table 3 documents the impact of budget problems by comparing three scenarios. The scenarios differ by the assumed initial generational account balances $GA_{0,t}$. For better intuition, they are calibrated in terms of tax rates required for non-pension expenditures (τ_{other}). By construction, one feasible policy in each scenario is to provide no pension and to impose only τ_{other}. However, optimal policy in all cases calls for pension benefits financed by additional taxes.

In the low-tax scenario, optimal pensions start earlier and are more generous than in the baseline, starting with 0.68 at age 70. The high-tax case has lower benefits that start later. Note that the average tax rates τ_{av} in Table 2 include τ_{other}, so payroll taxes decline from 0.18 percent in the low-tax case to 0.09 percent in the high-tax case. By comparison, benefits in a fixed-parameter system are much more sensitive to other taxes. In the low-tax case, a fixed-parameter system would offer comparable benefits (0.64) starting earlier (age 66); in the high-tax case, benefits would be drastically lower (0.18).

The column marked Δc in Table 3 shows welfare differences between optimal and fixed-parameter system in percent of lifetime consumption. The welfare gains for $\tau_{other} = 0.2$ are similar to $\tau_{other} = 0.1$. Welfare gains are smaller for $\tau_{other} = 0$ because optimal pensions then start fairly early and the fixed-parameter pensions are relatively generous, so they are closer to the optimal pension than in the other scenarios.

Note that these comparisons assume that cohorts can adjust their private savings. Unexpected policy changes in mid-life would have quite different, more adverse effects. That is, the analysis implicitly assumes that pension changes are pre-announced or phased in slowly.

In summary, an optimal pension system responds to changes in budget conditions by significantly varying the pension age, which allows the government to keep the per-period benefit amount much more stable that it could afford under a fixed pension age.

5 Conclusions

The paper shows that an optimal public pension system should have maximally deferred benefits. For any given present value, payments should start when funding is sufficient to cover all consumption for all remaining periods of life. Private savings pay for consumption between retirement and the pension age and are no longer needed once pensions start. The optimal pension age is generally higher than the retirement age.

The economic intuition is that private savings have recurrent costs for asset management and/or annuitization, whereas public pensions incur a one-time excess burden. The excess burden is only worth incurring if the interval between taxes and benefits is long enough. Hence public pensions have a comparative advantage at long horizons, whereas private savings are more efficient in early years of retirement.

These results are shown to be efficiency properties that hold regardless of redistributional preferences. They hold for earnings-linked and regular (uncommitted) pensions. With cross-sectional heterogeneity, optimal pension ages generally differ by income (or other characteristics). A progressive system would set higher pension ages for higher-income retirees. Numerical calculations suggest significant welfare gains from replacing a fixed-parameter system by an optimal pension.

An important implication in the context of pension reforms is that a gradual, pre-announced increase in the pension age, perhaps differentiated by income levels to maintain progressivity, would be more efficient than changes that reduce replacement rates.

The idea of pensions starting late in life is not radical in a historical context. When public pensions were first introduced in Germany in the 1890s, the pension age of 65 was close to adult life expectancy (i.e., conditional on survival to adulthood). When the U.S. social security system was created in 1935, the pension age of 65 was close to life expectancy at birth. Remarkably, the pension age has stayed virtually constant even as life expectancy has increased significantly over time. If pension systems had been adjusted in line with life expectancy, most developed countries would now have a pension age of about 75.

References

[1] Auerbach, Alan, Laurence Kotlikoff, and Willi Leibfritz. 1999. *Generational Accounting around the World*. Chicago: University of Chicago Press.

[2] Bohn, Henning. 1992. "Budget Deficits and Government Accounting." *Carnegie-Rochester Conference Series on Public Policy*, 37: 1–84.

[3] Bohn, Henning. 2006. "Who Bears What Risk? An Intergenerational Perspective." In: *Restructuring Retirement Risks*, ed. David Blitzstein, Olivia Mitchell and Stephen Utkus, 10–36. Oxford University Press.

[4] Brugiavini, Agar. 1993. "Uncertainty Resolution and the Timing of Annuity Purchases." *Journal of Public Economics*, 50(1): 31–62.

[5] Fenge, Robert, Silke Uebelmesser, and Martin Werding. 2006. "On the Optimal Timing of Implicit Social Security Taxes Over the Life Cycle." *FinanzArchiv / Public Finance Analysis*, 62(1): 68–107.

[6] Gamble, Keith Jacks, Patricia Boyle, Lei Yu, and David Bennett. 2015. "How Does Aging Affect Financial Decision Making?" Boston College, Center for Retirement Research.

[7] Gruber, Jonathan, and David Wise. 2008. *Social Security and Retirement around the World*. Chicago: University of Chicago Press.

[8] Mitchell, Olivia, James Poterba, Mark Warshawsky, and Jeffrey Brown. 2001. "New Evidence on the Money's Worth of Individual Annuities." In: *The Role of Annuity Markets in Financing Retirement*, ed. Olivia Mitchell, James Poterba, Mark Warshawsky and Jeffrey Brown. MIT Press.

[9] Poterba, James. 2014. "Retirement Security in an Aging Population." *American Economic Review: Papers & Proceedings* 2014, 104(5): 1–30.

[10] Rupert, Peter, and Giulio Zanella. 2012. "Revisiting wage, earnings, and hours profiles." *Journal of Monetary Economics*, 72(C): 114-130.

[11] Yaari, Menahem E. 1965. "Uncertain Lifetime, Life Insurance, and the Theory of the Consumer." *Review of Economic Studies*, 32(2): 137–150.

A Appendix

A.1 Proofs of the Propositions

Proof 1. The overall problem is to maximize W_{t_0} subject to (6), (7), and (8) by choice of $\{c_{i,t+i}, x_{i,t+i}, \tau_{i,t+i}, B_{i,t+i}\}_{i\geq 0, t\geq t_0}$ for given $D_{t_0} = 0$, $\{G_t\}_{t\geq t_0}$, $A_{i,t_0} = 0$ for all $i \geq 0$ at $t = t_0$, and $A_{0,t} = 0$ for $t > t_0$. The additional incentive constraints (3)-(5) are moot since they are be satisfied at the solution obtained without imposing them. The welfare problem subdivides straightforwardly because the welfare function W_{t_0} and the constraint (8) are both additively separable and the variables entering into $\{GA_{i_0,t_0}\}_{0\leq i_0\leq i_{\max}}$ and $\{GA_{0,t}\}_{t>t_0}$ are cohort-specific. ■

Proof 2. If $B_j > 0$ for any age $j \in [i_R, i_{\max})$, then $\Lambda_{B_j} = 0$ in (9), hence $\Lambda_{GA} = \beta^j R^j u'(c_j)$. Moreover, $\Lambda_{B_{j+1}} \geq 0$ in (9) implies $\beta^{j+1}R^{j+1}u'(c_{j+1}) \leq \Lambda_{GA}$, so $\beta R u'(c_{j+1}) \leq u'(c_j)$. Then in (3), $\kappa_i^a > 0$ implies $\Lambda_{x_j^a} = u'(c_i) - R_{j+1}^a \cdot \Pi_j \beta u'(c_{j+1}) > 0$, so $x_j^a = 0$. Since $u'(0) = \infty$, finite Λ_{GA} in (9) implies $c_i > 0$. Given $x_j \leq 0$, $c_i > 0$ requires $B_k > 0$ for at least one period $k > i$. Given $B_k > 0$, the above argument for j applies to k, and by finite recursion, the argument for $B_k > 0$ applies to $k = i_{\max}$. Given $B_{i_{\max}} > 0$, $\Lambda_{GA} = \beta^j R^j u'(c_j) = \beta^{i_{\max}} R^{i_{\max}} u'(c_{i_{\max}})$ implies $\Lambda_{GA} = \beta^i R^i u'(c_i)$ for all $i \in [j, i_{\max}]$ and hence $\beta R u'(c_{i+1}) = u'(c_i)$ for all $i \in [j, i_{\max} - 1]$. In (3-5), $\beta R u'(c_{i+1}) = u'(c_i)$ implies $x_j^a = d_j = 0$, hence $c_i = B_i$. ■

Proof 3. Suppose $B_k = 0$ for all $k \geq i_R$. Since $u'(0) = \infty$, individual must set $x_i > 0$ for $i \geq i_R - 1$ to ensure $c_i > 0$. Hence $K_{[i,i_{\max}]}(R\beta)^{i_{\max}} u'(c_t^{i_{\max}}) = (R\beta)^i u'(c_t^i)$ for $i = i_R - 1$ applies with $K_{i+1}(x_i) = 1 - \kappa_i^a < 1$. The same applies for $i < i_R - 1$ if $x_j > 0$ for $j \in [i, i_R - 1]$. From (10) $(R\beta)^i u'(c_i) = \Lambda_{GA}\varphi(\tau_i)$, so $K_{[i,i_{\max}]}(R\beta)^{i_{\max}} u'(c_t^{i_{\max}}) = \Lambda_{GA}\varphi(\tau_i)$. By assumption, $\varphi(\tau_i)/K_{[i,i_{\max}]} > 1$, which implies $(R\beta)^{i_{\max}} u'(c_t^{i_{\max}}) > \Lambda_{GA}$, contradicting (9). Hence $B_k = 0$ for all $k \geq i_R$ is inefficient. Hence setting $B_k > 0$ for some $k \geq i_R$ is a necessary condition for efficiency. ■

Proof 4. From (10), $(R\beta)^{i_R-1} u'(c_{i_R-1}) = \Lambda_{GA}\varphi(\tau_{i_R-1})$, hence $\beta^j R^j u'(c_i) = \Lambda_{GA}\frac{\varphi(\tau_{i_R-1})}{K_{[i_R-1,j]}}$ for all $j \geq i_R$. If $\varphi(\tau_{i_R-1}) < K_{[i_R-1,i]}$ for some $i \geq i_R$, then $\varphi(\tau_{i_R-1}) < K_{[i_R-1,i]} \leq K_{[i_R-1,j]}$ for all $j \leq i$, which implies $(R\beta)^i u'(c_j) < \Lambda_{GA}$, so $\Lambda_{B_j} > 0$ and hence $B_j = 0$. ■

Proof 5. For $\tau_{i-1} > 0$ and $\tau_i > 0$, (10) and (11) imply $\varphi(\tau_{i-1}) = \beta^{i-1}R^{i-1} \cdot u'(c_{i-1})/\Lambda_{GA} = K_i(x_{i-1}) \cdot \beta^i R^i u'(c_i)/\Lambda_{GA} = K_i(x_{i-1}) \cdot \varphi(\tau_i)$. ■

Proof 6. The argument was provided in the text. An easy formal proof is by contradiction: Suppose benefits were paid in some period i_1 while private assets are positive. Then private asset must be used to fund consumption in some period $i_2 > i_1$, which implies $K_{k+1}(x_k) < 1$ for $i_1 \leq k < i_2$, which in turn implies that v_{i_R} could be increased by a marginal reduction in $B_{i_1}^e$ and a marginal increase in $B^e_{i_2}$ that keeps the sum $b_{i_1} + b_{i_2}$ unchanged, contradicting optimality. ■

Proof 7. By assumption, an allocation that maximizes W_{t_0} exists, it specifies retirement benefits $\{B_{i,t+i}(\eta)\}_{i \geq i_R, t \geq t_0}$ for each cohort and type. Since $GB_{i_R}(\eta) > 0$, $B_i(\eta) > 0$ for some $i \geq i_R$, so $i_P = \min\{i : B_i(\eta) > 0\}$ exists. Since benefits are additively separable in the government's budget constraint, optimal benefits for each type and for each cohort $t \geq t_0 - i_R$ must maximize $E_t[U^R_t(\eta)]$ subject to given $GB_{i_R}(\eta)$ and $A_{i_R}(\eta)$ by choice of $B_{j,t+j}(\eta)$ and $c_{i,t+j}(\eta)$. The text claims that for each type η pension benefits must be structured so that $B_j(\eta) = c_j(\eta)$ for all $j > i_P$. (This is as in Prop.2 and Prop.6. The additional claim $x_{i_P}(\eta) = 0$ follows as corollary.) The proof is by contradiction, in two parts: ■

(a) Proof for regular benefits: Suppose for contradiction that optimal policies satisfies $B_i(\eta) < c_i(\eta)$ for some period $i_+ > i_P$. Then $A_i(\eta) \geq c_{i_+}(\eta) - B_{i_+}(\eta) > 0$ implies non-zero assets for $[i_P, i_+]$, hence $K_{[i_P,i_+]} < 1$. Consider a marginal increase in $\Delta B_{i_+}(\eta) > 0$ at age i_+ combined a marginal reduction $\Delta B_{i_P}(\eta) = -\Delta B_{i_x}(\eta)\Pi_{[i_P,i_+]}/R^{i_x - i_P}$ at age i_P that leaves $GB_{i_R}(\eta)$ unchanged. Since $A_i(\eta) > 0$ for $i \in [i_P, i_+]$, individuals can reduce $A_i(\eta)$ marginally by $\Delta A_i(\eta) = -\Delta B_{i_+}(\eta)\Pi_{[i,i_+]}/R^{i-i_P}/K_{[i,i_+]}$ to maintain unchanged $c_i(\eta)$ for $i \in (i_{P+1}, i_+]$. This variation raises consumption in period i_P by $\Delta c_{i_P}(\eta) = -\Delta x_{i_P}(\eta) - \Delta B_{i_P}(\eta) = (1/K_{[i_P,i_+]} - 1)|\Delta B_{i_P}(\eta)| > 0$, which shows that the variation increases utility, contradicting optimality.

(b) Proof for earnings-linked benefits: Suppose for contradiction that $B_i(\eta) < c_i(\eta)$ for some period $i_+ > i_P$. Then marginal increase in $b_{i+}(\eta)$ combined with $\Delta b_{i_P}(\eta) = -\Delta b_{i+}(\eta)$ would increase $v_{i_R}(\eta)$ by $\Delta v_{i_R}(\eta) = \Delta b_{i_P}(\eta) + \Delta b_{i+}(\eta)/K_{[i_P,i_+]} > 0$. Hence the original policy did not minimize tax distortions, in addition to not maximizing utility, again contradicting optimality.

Proof 8. Follows from the marginal arguments in the Proof of Prop.7. ■

A.2 The Welfare Problem with Uniform Taxes

Consider the problem of Section 1 with the additional constraint of uniform, age-independent taxes. As discussed in Section 1 (Prop.1), the welfare problem of maximizing W_{t_0} subject to (8), (7), and (6) subdivides straightforwardly if taxes $\tau_{i,t+i}$ can vary across cohorts. With uniform taxes, the problem can be subdivided differently, namely into (A) maximizing W_{t_0} over retirement-age choices $\Omega_R = \{c_{i,t+i}, x_{i,t+i}, B_{i,t+i}\}_{i \geq i_R, t \geq t_0}$ conditional on working-age choices $\Omega_w = \{c_{i,t+i}, x_{i,t+i}, \tau_{i,t+i}, B_{i,t+i}\}_{i \geq 0, i < i_R, t \geq t_0}$; and (B) maximizing over the conditioning variables Ω_w.

To do this, first divide utility and welfare into working-age and retirement components, $U_t = U^{work}_t + U^R_t$, where $U^{work}_t = \sum_{i=0}^{i_R - 1} \beta^i \sigma_{i,t+i} u(c_{i,t+i})$ and $U^R_t = \sum_{i=i_R}^{i_{\max}} \beta^i \sigma_{i,t+i} u(c_{i,t+i})$, and $W_{t_0} = W^{work}_{t_0} + W^R_{t_0}$, where $W^{work}_{t_0} = \sum_{t=t_0 - i_{\max}}^{\infty} \beta^{t-t_0} \omega_t E_{t_0}[U^{work}_t]$ and $W^R_{t_0} = \sum_{t=t_0 - i_{\max}}^{\infty} \beta^{t-t_0} \omega_t E_{t_0}[U^R_t]$.

Second, note that the generational accounts divide similarly: $GA_{0,t} = GA_{work,t} + \rho^{i_R} GA_{i_R, t+i_R}$ for generic $t \geq t_0$, was defined in the text; for $t < t_0 - i_R$, the same division applies with truncated work and/or retirement periods.

Third, note that Ω_w implies values $GA_{i_R,t+i_R}$ and $A_{i_R,t+i_R}$ for generations $t \geq t_0 - (i_R - 1)$, and values GA_{t_0-t,t_0} and A_{t_0-t,t_0} for generations $t \leq t_0 - i_R$, and that the conditioning variables are relevant for the retirement period only through GA and A. Hence part (A) of the problem reduces to maximizing $W_{t_0}^R$ subject GA- and A-values at retirement. (For generations $t \geq t_0 - (i_R - 1)$, the relevant variables are $GA_{i_R,t+i_R}$ and $A_{i_R,t+i_R}$; for generations $t \leq t_0 - i_R$, they are GA_{t_0-t,t_0} and A_{t_0-t,t_0}.)

Moreover, since $\{c_{i,t+i}, x_{i,t+i}, B_{i,t+i}\}_{i \geq i_R}$ for any t enters into U_t^R and $GA_{i,t+i}$ for only one generation, the problem of maximizing $W_{t_0}^R$ separates into generation-specific problems of $E_{t_0}[U_t^R]$. Specifically:

(A1) For cohorts $t \geq t_0 - (i_R - 1)$, maximize $E_t U_t^R$ subject to (7) and (6) by choice of $\{c_{i,t+i}, x_{i,t+i}, B_{i,t+i}\}_{i \geq i_R}$ for given $GA_{i_R,t+i_R}$ and $A_{i_R,t+i_R}$. The solutions define indirect utility functions

$$V_{i_R,t}^R(GA_{i_R,t+i_R}, A_{i_R,t+i_R}) = \max\{E_{t_0} \sum_{i=i_R}^{i_{\max}} \beta^i \sigma_{i,t+i} u(c_{i,t+i}) | GA_{i_R,t+i_R}, A_{i_R,t+i_R}\}.$$

(A2) For cohorts $t \in [t_0 - i_{\max}, t_0 - i_R]$, the analogous truncated problems starting at age $i = t_0 - t$ define indirect utility functions $V_{t_0-t,t}^R(GA_{t_0-t,t_0}, A_{t_0-t,t_0})$.

Given solutions to (A1) and (A2), the remaining welfare problem (part B) is to maximize

$$W_{t_0} = W_{t_0}^{work} + \sum_{t=t_0-i_R}^{\infty} \beta^{t-t_0} \omega_t V_{i_R,t}^R + \sum_{t=t_0-i_{\max}}^{t_0-i_R} \beta^{t-t_0} \omega_t V_{t_0-t,t}^R.$$

by choice of Ω_w and of generational account and asset balances at retirement.

By construction, problems (A1) and (A2) do not depend on working age. Hence optimal pensions (if nonzero) have the properties described in Prop.7.