Optimal Public Pensions: Full Coverage for End-of-Life Consumption

Henning Bohn*

June 5, 2015

Abstract

Public pensions and private saving are commonly considered complementary funding sources from which the elderly should draw simultaneously throughout their retirement. This paper shows that sequential use of funding sources would be significantly more efficient—a period of entirely private funding followed by a period of consumption fully financed by public pensions. Thus the standard multi-pillar approach to retirement requires rethinking. Implied is a sharp distinction between the age of retirement—a matter of personal preferences and resources—and the age of eligibility for public pensions. The analysis also implies that changes in longevity and in funding levels are optimally absorbed by gradual adjustments in the eligibility age, leaving annual payments largely unchanged.

*Department of Economics, NH 2127, UCSB, Santa Barbara, CA 93106. Email: henning.bohn@ucsb.edu. Home page: 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 (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 funding structure would be more efficient. Optimal retirement financing starts with a period in which retirement consumption is entirely privately funded followed by a period with entirely public funding. Once public pensions start, the optimal benefit is high enough to finance all desired consumption. Retirees optimally spend down all their private assets by the time their pension starts.

The basic argument is that public pensions and private saving have comparative advantages at different horizons. For private retirement saving, key challenges are longevity risk and managing assets. Longevity risk motivates annuitization. Annuitization requires that retirement wealth is turned over to outside managers, which incurs intermediation costs, creates moral hazards, and causes illiquidity. Default risk is also a concern because annuitants are creditors of the insurance company. These costs and contracting problems tend to be cumulative over the term of a contract. For public retirement funding, the main challenge is the excess burden of taxation, which may be substantial but is not intrinsically cumulative. This suggests that the use of public funds—any given amount—should be concentrated late in life when survival probabilities are low, private funding would require extremely long contracts, and excess burden is amortized over many decades. Private saving is relatively more efficient for the period soon after retirement.

The idea of decoupling public pension eligibility from retirement is consistent with recommendations to increase the retirement age in response to rising longevity, but not the same. The retirement age here is an individual choice – an optimal labor supply decision – whereas pension eligibility is a policy choice. Instead of working longer when the age of pension eligibility is raised, individual may save privately to bridge the gap between retirement and pension eligibility.

In the U.S., social security makes actuarial adjustments when individuals start drawing benefits before or after their "full" retirement age, but only within the age range 62-70. My analysis is consistent with the common financial advice that retirees should defer claiming social security. However, the reasoning is based on welfare analysis, not on "gaming" the system; and the quantitative analysis suggest that age 70 is still too low.

The main contribution of the paper is to present sequential funding – private then public – as a general principle for the design of public pension systems. The principle also applies to non-cash benefits. For example, Medicaid coverage for nursing homes can viewed as retiree benefit that is appropriated focused on very old age.

To study the optimal interaction of private saving and public pensions, one must avoid assumptions that would produce extreme solutions. If private financial markets were perfect, even small tax distortions would make public funding
inefficient. If governments could credibly promise to return addition "contributions" with interest, suitable tax-transfer schemes could replace all private retirement saving. To avoid these extremes, the model assumes intermediation costs for private annuities and it restricts labor taxes to be linear. A combination of public and private retirement funding is then efficient, and the focus is on how to combine them optimally.

Several related issues are not modeled explicitly but should be noted. 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)). 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 (see 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. These issues strengthen the case for sequential funding.

There is a related literature on why annuities are not widely used, contrary to theoretical predictions of full annuitization (Yaari (1965)). Brugiavini (1993) is a classic. Numerous empirical studies have examined the money’s worth of annuities, e.g. Mitchell et al. (2001). Forman and Chen (2008) discuss the optimal retirement age in the context of increasing longevity. There is also a literature on the optimal timing of payroll taxes, e.g., Fenge, Uebelmesser and Werding (2006), which is relevant here for determining optimal levels of public versus private retirement funding.

The paper is organized as follows. Section 1 sets up an overlapping generations model with stochastic mortality and describes conditions for Pareto-efficient policies. The basic model allows age- and cohort-dependent taxes and pension benefits but abstracts from cross-sectional heterogeneity. Age-dependence is allowed to avoid a priori restrictions on structure and timing of taxes and transfers; heterogeneity is omitted because cross-sectional redistribution would distract from the intertemporal issues. Section 2 derives the main analytical results about optimal pensions. Section 3 provides a numerical analysis showing sizeable welfare gains from sequential pension funding. Section 4 comments on special cases and examines extensions; this includes comments

---

1To illustrate the challenge of retirement planning, consider a person who wants to retire at age 65, expects to cover half of retirement consumption from public pensions and the other half from accumulated private saving. Suppose life expectancy at retirement is 80 and the maximum age is 120. Funding 50 percent of annual consumption privately means that the person must either convert all private assets into life annuities or leave substantial accidental bequests. Buying annuities means finding a company that can be trusted to make payments reliably for 55 years from 65 to 120. If public pensions were delayed instead and then covered all consumption, the individual planning horizon would be much reduced. For example, suppose consumption starting at age 80 has about the same present value as 50 percent of consumption starting at age 65. Then the government should be indifferent between paying 50 percent starting at 65 or 100 percent starting at 80. The latter would reduce the individual planning horizon from 55 years to 15 years and much reduce the cost of annuitization and/or accidental bequests.
on uniform (age-independent) taxes, pension commitments, and cross-sectional wage differences. Section 5 concludes. Proofs are in an appendix.

1 A Basic Life-Cycle Model

1.1 Population and Preferences

Let time $t$ be discrete. Cohort $t$ is born at time $t$, starts working at age $i_W$ and has maximum life span $i_{\text{max}}$. Let $\Pi_{[0,j],t} = E_t[\sigma_{i,t+i}]$ denote unconditional survival probabilities to age $i$, where $\sigma_{i,t+i}$ is a 0-1 indicator for being alive (1) or not (0) 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 for general $j > i$, and $\Pi_{i,t+i} = \Pi_{[0,i+1],t}/\Pi_{[0,i],t}$ denotes one-period survival at time $t+i$. Assume $\Pi_{i,t+i} \in (0,1)$ for all $i < i_{\text{max}}$, $\Pi_{i_{\text{max}},t+1} = 0$, and $\Pi_{i,t+i}$ decreases with age.

Individuals have preferences over consumption $c_{i,t+i}$ at ages $i \geq i_W$ defined by

$$U_t = \sum_{i=i_W}^{i_{\text{max}}} \beta^i \sigma_{i,t+i} u(c_{i,t+i})$$

Childhood consumption is omitted (subsumed into parental consumption); for reference, normalize $c_{i,t+i} = 0$ for $i < i_W$. Utility $u$ is increasing and concave with $u'(0) = 1$, and $\beta > 0$.

The cohort size $N_t$ is exogenous and members of each cohort are identical. Though cross-sectional heterogeneity is empirically important for pensions policy, it is a distraction here and hence deferred to Section 4.

For convenience, let time-subscripts be omitted when dealing with single generic cohort $t$. Let all variables equal zero during childhood ages $i < i_W$ (e.g., $c_{i,t+i} = 0$ for $i < i_W$), so equations below are defined for all $i$, avoiding case distinctions. Also, let lagged terms in $U_t$ be omitted when expected utility $E_{t+i}U_t$ is evaluated at ages $i > i_W$.

1.2 Labor Supply and Economic Retirement

Assumption on labor supply are designed to keep the supply side simple. For any cohort $t$, assume overall work effort $L_i \in \{0,1\}$ at ages $i \geq i_W$ be indivisible and have productivity $w_i \geq 0$. Effort is either supplied to the market, denoted $l_i \in [0,L_i]$, or to home production $H$. Assume there is an exogenous cost of working $\mu_i \geq 0$. Market labor is subject to a wage tax at rate $\tau_i$. After-tax labor income is then

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

where $I$ is the indicator function and $H$ is increasing, concave, and satisfies $H(1) < 1$ and $H'(0) = \infty$. 

4
Maximizing $y_i$ with respect to labor implies $l(\tau_i) \equiv 1 - (H')^{-1}(1 - \tau_i)$, which is decreasing in $\tau_i$. Define $Y(\tau) \equiv (1 - \tau)l(\tau) + H(1 - l(\tau))$ and $T(\tau) \equiv \tau l(\tau)$. Then labor income at age $i$ is $y_i = w_i Y(\tau_i) - \mu_i$.

Let $\hat{\tau} = \arg \max \{\tau_i | l(\tau_i)\}$ be the revenue-maximizing tax rate (Laffer curve peak), let $\hat{Y} = Y(\hat{\tau})$ be the corresponding income, and let $\varphi(\tau) = \frac{T'(\tau)}{-Y'(\tau)} \in (0, 1)$ for $\tau \in (0, \hat{\tau})$ denote tax receipts per unit tax-induced reduction in income, so $\frac{1}{\varphi(\tau)} - 1 > 0$ is the excess burden.\footnote{Note that $Y'(\tau) = -l(\tau) + (1 - \tau - H')l'(\tau) = -l(\tau) < 0$ and $T'(\tau) = l(\tau) - \tau l'(\tau) \in (0, -Y'(\tau))$ for all $\tau \in (0, \hat{\tau})$, so $\frac{T'(\tau)}{-Y'(\tau)}$ is well-defined and within $(0, 1)$.}

Retirement is modeled by assumptions on $\mu_i$. For now, assume $\mu_i$ is a step function that jumps from $\mu_i = 0$ for $i < i_R$ to $\mu_i = \bar{\mu} > \hat{Y} \cdot \max_{i \geq i_R} \{w_i\}$ for $i \geq i_R$ at some age $i_R \in (0, i_{\max})$. Then $i_R$ is a well-defined retirement age in the sense that $L_i = 0$ is optimal for all $i \geq i_R$.

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

### 1.3 Private Saving with Financial Frictions

Suppose private financial markets offer a constant return $r$ on saving, except that annuitization has fixed cost $\kappa_a > 0$ and borrowing has an intermediation cost $\kappa_d > 0$.

The constancy of market returns is for simplicity. Time variation and uncertainty would needlessly clutter the notation. The two frictions are empirically relevant and they yield tractable deviations from the perfect-markets paradigm.

The annuitization cost exemplifies a saving friction that becomes more severe with age, which is central to the analysis. One might suspect that saving are subject to other frictions that increase with age, e.g., due to medical or mental impairments, but one friction is sufficient for the analysis.

The intermediation cost is a technical assumption to discourage leverage, notably pension-related borrowing. The possibility of costless borrowing to pay pension contributions and borrowing against future pensions would add misleading degrees of freedom to pension policy, and it would distract from the analysis.

Saving frictions are conceptually important here because public pensions involve a round-trip of funds — from participants to the government during working age, and from government to participants in retirement. For any given amount of net taxes paid or net transfers received, such round-tripping could not be efficient if private savings were entirely frictionless.

Consider a generic cohort $t$ at age $i$. With the above assumptions, the per-period gross return on regular assets is $R = 1 + r$, for all $i$. Conditional on survival to age $i + 1$, the return on annuities is $R_{i+1}^{\kappa} (1 - \kappa_a)/\Pi_\cdot R$. Private debts $d_i$ are annuitized (or equivalently, not collectible upon death) and accumulate at rate $R_{i+1}^{d} (1 + \kappa_d)/\Pi_\cdot R$. Let net saving be $y_i + B_i - c_i$, which is after-tax labor income $y_i$ plus government benefits $B_i \geq 0$ minus consumption...
$c_i$; let $A_i$ denote initial net assets; and let $(x_i^n, x_i^a, d_i)$ be the vector of end-of-period regular assets, annuities, and debt. Every period, net saving and initial assets are allocated to new investments:

$$x_i^n + x_i^a - d_i = x_i \equiv A_i + y_i + B_i - c_i,$$

where $A_0 = 0$. Net assets at the start of the next period are

$$A_{i+1} = R_{i+1}^a x_i^n + R x_i^a - R_{i+1}^d d_i.$$

The first-order conditions for optimal investments and borrowing are

$$R_{i+1}^a \cdot \Pi_j \beta' (c_{i+1}) - u' (c_i) + \Lambda x_i^d = 0 \text{ for } x_i^d \geq 0$$

$$R \cdot \Pi_i \beta' (c_{i+1}) - u' (c_i) + \Lambda x_i^n = 0 \text{ for } x_i^n \geq 0$$

$$R_{i+1}^d \cdot \Pi_j \beta' (c_{i+1}) - u' (c_i) - \Lambda d_i = 0 \text{ for } d_i \geq 0$$

where $\Lambda_\xi$ denotes the Kuhn-Tucker multiplier for a generic condition $\xi \geq 0$. From (4-6), assets are annuitized if and only if $\Pi_i \leq 1 - \kappa_a$. Since mortality is increasing with age, annuitization starts at some critical age $i_a$, so $x_i^a = 0$ for $i < i_a$ and $x_i^a = 0$ for $i \geq i_a$. Moreover, debt and assets are mutually exclusive. Hence $x_i = x_i^1$ if $x_i \geq 0$ and $i < i_a$, $x_i = x_i^a$ if $x_i \geq 0$ and $i \geq i_a$, and $x_i = -d_i$ if $x_i < 0$.

To streamline the exposition below, define the return on net assets by

$$R_{i+1}(x_i) = K_{i+1}(x_i) \cdot \frac{R}{\Pi_i},$$

where $K_{i+i}(x_i) = \begin{cases} 1 - \kappa_a & \text{if } x_i > 0 \text{ and } i \geq i_a \\ \Pi_i & \text{if } x_i > 0 \text{ and } i < i_a \\ 1 + \kappa_d & \text{if } x_i = -d_i < 0 \\ 1 & \text{if } x_i = 0 \end{cases}$

summarizes investment-management and borrowing costs. The normalization $K_{i+i}(0) = 1$ will be convenient for corner solutions below. Also, let $K_{i+i} = \max \{ \Pi_{i+i}, 1 - \kappa_a \} < 1$ denote the cumulative shrinkage in total returns from $i_1$ to $i_2$ when $x_j > 0$ for $i_1 \leq j < i_2$, i.e., for savers.

Then asset dynamics are $A_{i+1} = R_{i+1}(x_i)x_i$. End-of-period net assets, conditional on survival, follow the recursion

$$x_i = y_i + B_i - c_i + R_s(x_{i-1})x_{i-1}.$$

For $i < i_R$, $y_i = w_i Y(\tau_i)$ depends on taxes and the $B_i$ would be lump-sum rebates of distortionary taxes. (They are inefficient when $\tau_i > 0$ but included here to cover cases with $\tau_i = 0$). For $i \geq i_R$, $B_i$ represents retirement benefits and $y_i = 0$.

Individuals maximize $E_t[U]$ by choice of $\{c_i, x_i\}$ subject to (7) and $A_0 = 0$ for given (rational) expectations about future policy.

### 1.4 Fiscal Policy: Taxes and Transfers

The objective is to characterize optimal policy towards a generic cohort $t$. However, the policy problem is presented as optimal plan in a specific initial period
to ensure that alternative policies respect the same initial conditions. 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 benefits \( B_{i,t} \geq 0 \) applied to cohorts \( t \geq t_0 - i_{\text{max}} \) at ages \( i \geq \max\{i_{W}, t_0 - t\} \). (For children, set \( \tau = B = 0 \).) Both may vary over time and by age and are treated as discretionary. (Uniform taxes and pension commitments are discussed in Section 4.) Though individual survival 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

\[
P_{B,t} = \sum_{i\in[0,i_{R}+1]} n_{t-i} \Pi_{[0,i], t-i} w_{i,t} T(\tau_{i,t}) - \sum_{i\in[i_{\text{max}}, t]} n_{t-i} \Pi_{[0,i], t-i} B_{i,t} - G_t;
\]

where revenues are summed over working-age cohorts (born between \( t \) and \( t - i_{R} + 1 \)) and benefits are summed over all living cohorts. Over time, debt accumulates at the interest rate \( r \), so debt at the start of period \( t + 1 \) is \( D_{t+1} = R \cdot (P_{B,t} + D_t) \). The government’s intertemporal budget constraint is

\[
D_{t_0} = \sum_{i=0}^{\infty} \rho^{t-0} P_{B,t}, \text{ where } \rho = 1/R.
\]

Fiscal policy combines efficiency and redistributional issues. They can be disentangled by sorting taxes and benefits into generational accounts. Let

\[
G_{A_{i_0},t+i_0} = \sum_{i=i_0}^{i_{\text{max}}} \rho^{i-i_0} \Pi_{[i_0,i], t+i_0} [T(\tau_{i,t+i}) w_{i,t+i} - B_{i,t+i}]
\]

denote the generational account of cohort \( t \) at age \( i_0 \), per unit population, which is the present value of taxes minus transfers at ages \( i \geq i_0 \) discounted to age \( i_0 \). The government’s intertemporal budget constraint (IBC) at time \( t_0 \) can then be written as linear combination of generational components:

\[
D_{t_0} + \sum_{t=t_0}^{\infty} \rho^{t-t_0} G_t = \sum_{i=0}^{i_{\text{max}}} n_{t-i} G_{A_{i},t_0} + \sum_{t=t_0+1}^{\infty} \rho^{t-t_0} n_i G_{A_0,t}.
\]

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

Implicit in this budget accounting is a distinction between debt payments and transfers. Debt is a commitment whereas transfers are generally discretionary. Commitments matters here because taxes are distortionary and obligations may imply future taxes. To highlight the commitment issue, debt payments are not included in the generational account.\(^3\)

\(^3\)This differs from generational accounting models that assume lump sum taxes (e.g., Auerbach, Kotlikoff and Leibfritz (1999)). See Bohn (1992) for a discussion of government accounting and the role of tax distortions.

To avoid the Samaritan’s dilemma, also assume a commitment to equal treatment within each cohort, so an individual who saves less than others will not receive additional benefits.
Pareto efficiency requires that policy maximizes a welfare function of the form
\[ W_{t_0} = \sum_{t=t_0}^{\infty} \beta^{t-t_0} \omega_t E_{t_0}[U_t] \]
for some sequence of weights \( \omega_t > 0 \).

Note that with tax distortions captured by \( Y \) and \( T \), policy has no incentive to distort saving. Hence \( \{c_{i,t+i}, x_{i, t+i}\} \) can also be treated as policy choices. Moreover, with cohort-dependent taxes and benefits, each policy variable enters into exactly one of the generational accounts in (9). Hence the welfare problem of maximizing \( W_{t_0} \) divides into three separate problems:

(i) For each cohort \( t > t_0 \), optimal policy maximizes \( E_{t_0}[U_t] \) for given \( GA_{0,t} \), which is the generational account at birth.

(ii) For each cohort alive at time \( t_0 \), optimal policy maximizes \( E_{t_0}[U_t] \) for given \( GA_{i,t_0} \), which is the generational account at age \( i = t - t_0 \) at time \( t_0 \).

(iii) Optimal policy maximizes \( W_{t_0} \) by choice of initial generational account balances \( F_{GA_{i,t_0}} \) and \( F_{GA_{0,t}} \), taking as given how \( E_{t_0}[U_t] \) depends on generational accounts.

Since only part (iii) depends on welfare weights, parts (i) and (ii) are efficiency conditions. The main results will be based on these efficiency conditions.

2 Properties of Efficient Fiscal Policies

2.1 Life-cycle Analysis: Optimal Taxes and Pensions

Consider the government’s problem of maximizing a generic cohort’s utility at age \( i_0 \), given the generational account \( GA_{i_0} \).

Since the government maximizes utility, individual rationality is satisfied trivially. Optimal benefits and taxes must satisfy the first order conditions
\[
\begin{align*}
\beta^i u'(c_i) \Pi_{[0,i]} - \Lambda_{GA} \cdot \rho \Pi_{[0,i]} + \Lambda_{B_i} &= 0 \\
\beta^i u'(c_i) Y'(\tau_i) w_i + \Lambda_{GA} \rho \Pi_{[0,i]} T'(\tau_i) w_i + \Lambda_{\tau_i} &= 0
\end{align*}
\]
for all \( i \geq i_0 \), where \( \Lambda_{GA} \) is the shadow value of per-capita resources in (8), and where \( \Lambda_{B_i} \geq 0 \) and \( \Lambda_{\tau_i} \geq 0 \) are Kuhn-Tucker multipliers. Equivalently:
\[
\begin{align*}
\beta^i R^i u'(c_i) &= \Lambda_{GA} - \Lambda_{B_i} R^i / \Pi_{[0,i]} \leq \Lambda_{GA} \quad \text{and} \\
\beta^i R^i u'(c_i) &= \Lambda_{GA} \cdot \varphi(\tau_i) + \Lambda_{\tau_i} \cdot \psi_i
\end{align*}
\]
where \( \varphi(\tau_i) \) captures tax distortions and \( \psi_i = \frac{(1+r)^i}{\varphi(\tau_i) w_i \Pi_{[0,i]}} > 0 \) is a scale factor. Several results follow.

\[ \text{For } t > t_0 \text{ and } i_0 = 0, \text{ this is part (i) of the welfare problem above. For } t = t_0 \text{ and } i_0 \geq 0, \text{ this is part (ii) of the welfare problem above. To avoid case distinction, the problem is presented for generic } t \text{ and } i_0. \]
Proposition 1 If optimal public pension benefits are nonzero at any age $i_+ \geq i_R$, then public pensions pay for all consumption in all subsequent periods. Private wealth optimally declines to zero in the period when pension benefits start.

The intuition that for any generational account balance, a marginal reduction in benefits at age $i$ allows benefits at age $i + 1$ to increase by $R_i = \frac{\kappa_i}{1 - \kappa_i}$, which exceeds the return on private assets, both on regular assets (because $\Pi_i < 1$) and on annuities (because $\alpha > 0$). Technically, the period with non-zero benefits has $B_{i_R} = 0$, so $\beta^{i+} R^i u'(c_{i+1}) = \Lambda_{GA}$ links marginal utility to the generational account. In all subsequent periods, private savings with returns less than $R_i$ would violate $\beta^i R^i u'(c_i) \geq \Lambda_{GA}$. Hence a policy that provides benefits while private assets are positive must be inferior to a policy that defers benefits until private assets are zero. (See appendix for detailed proof.)

Prop. 1 is sharply at odds with the conventional wisdom that retirees should tap a mixture of private and public funds throughout their retirement. Optimal pensions jump from zero to 100% of consumption within at most one transition period. In the transition period, pensions start and all remaining private assets are exhausted.

For reference, define the eligibility age $i_e = \min\{i \geq i_R : B_i > 0\}$ as the age at which pensions start. Then from Prop. 1, optimal policy implies $B_i = c_i$ and $A_i = 0$ for all $i > i_e$.

Proposition 1 applies if benefits are non-zero. The following propositions provide conditions for non-zero benefits and suggest that there should be a time gap between the end of contributions and the start of pensions.

Proposition 2 Suppose there is an age $i \leq i_R - 1$ such that savers not receiving any pension would set $x_j > 0$ for all $j \in [i, i_R]$. If $\varphi(i) > K_{[i, i_{max}]}$, then optimal policy necessarily implies $B_i > 0$ for some $i \geq i_R$.

Prop. 2 shows that financial frictions provide a rationale for public pensions. The intuition is that pure private saving – zero public pensions – cannot be optimal if the cumulative loss of investment returns (measured by $K_{[i, i_{max}]}$) as compared to costless annuities is greater than the excess burden of taxes. The relevant time period is the period with assets $x_j > 0$. Note that without public pensions, non-zero consumption would require $x_j > 0$ for at least the period $[i_R - 1, i_{max}]$; hence $\varphi(i_{i_R-1}) > K_{[i_{i_R-1}, i_{max}]}$ is a sufficient condition for $B_i > 0$.

The model implicitly assumes that the government can costlessly operate a tax-transfer scheme that in effect substitutes for private saving. If public pensions were pre-funded with real assets, it would indeed be unreasonable to ignore management costs. 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, and possibly a cost-reduction from reduced government activity in bond markets. Pay-as-you-go financing is a comparative advantage in this context.

Proposition 3 If tax rates are high enough in the period before retirement that $\varphi(i_{i_R-1}) < K_{[i_{i_R-1}, i]}$ for some $i \geq i_R$, then $B_j = 0$ is optimal for all $j \in [i_R, i]$.
Under the conditions of Prop. 3, there is a time gap between retirement and the start of pensions. The initial years of retirement fully financed by private saving. Conversely, if $B_i > 0$ is optimal, then optimal taxes in period $i_R - 1$ must be low enough that the excess burden is small (specifically, $\varphi(\tau_{i_R - 1}) \geq K_{i_R - 1}$, which means $\frac{1}{\varphi} \leq \min(\kappa_a, 1 - \Pi_{i_R - 1})$).

**Proposition 4** Optimal tax rates satisfy $\varphi(\tau_{i-1})/\varphi(\tau_i) \in [\Pi_{i-1}, 1 + \kappa_d]$.

Note that $[\Pi_{i-1}, 1 + \kappa_d]$ is an interval around one, so optimal taxes are smooth over time. The details depend on the functional form of $\varphi$ and on the path of savings or borrowing. With $\varphi$ decreasing in $\tau$, optimal tax rates are declining over periods of asset accumulation (for $x_{i-1} > 0$, one can show $\varphi(\tau_{i-1})/\varphi(\tau_i) = \max[\Pi_{i-1}, 1 - \kappa_d] < 1$). This is broadly consistent with Fenge, Uebelmesser and Werding (2006). The main point here is that if $B_i > 0$, so $\tau_{i_R - 1}$ small from Prop. 3, then taxes must be relatively low for several years prior to retirement. In combination, Prop. 3 and 4 imply that there should be a time gap between taxes and benefits – either a waiting period after retirement, or low taxes before retirement, or both.

More specific implications of optimal policy are best examined quantitatively, which is in Section 3 below.

### 2.2 A Finance Interpretation

This section provides a finance-style interpretation. Because policy maximizes cohort utility, the choice could be delegated to a representative member of the cohort who maximizes $E_t U_t$ subject to given $GA_0$ by choice of $\{c_i, B_i, \tau_i, A_i\}$.

This problem has a recursive representation that will be convenient for numerical analysis. The generational account can be interpreted as a liability, which is reduced by tax payments and increased by benefit payouts. To formalize this, define $\alpha_i = -GA_i/\Pi_{i\alpha_0}$ as the “generational net assets” of a member cohort $t$ alive at age $i$. From (8),

$$\alpha_{i+1} = \frac{R}{\Pi_i} \cdot [\alpha_i + T(\tau_i)w_i - B_i].$$

(12)

accumulates without management or annuitization costs.

The cohort’s problem is then analogous to a saving problem with two separate investment funds, henceforth labeled fund $A$ and fund $a$. The fund balances $(A_i, \alpha_i)$ serves as state variables. The Bellman equation with planning horizon $j = t_{\max} - i \geq 1$ is

$$V_{j+1}(A_i, \alpha_i) = \max\{u(c_i) + \beta \Pi_i V_j(A_{i+1}, \alpha_{i+1})\}$$

subject to (12), $A_{i+1} = R_{i+1}(x_i)\alpha_i$, $x_i = A_i + B_i - c_i + (Y(\tau_i)w_i - \mu_i)$, and the inequalities $c_i \geq 0$, $B_i \geq 0$, $\tau_i \geq 0$, and $w_i Y(\tau_i) \geq \mu_i$. All assets are consumed at $t_{\max}$, so $B_{t_{\max}} = \alpha_{t_{\max}}$ and $V_1(A_{t_{\max}}, \alpha_{t_{\max}}) = u(A_{t_{\max}} + \alpha_{t_{\max}})$.

It is important that 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 $\alpha$ are invoiced through taxes. To clarify this, let $z_i = T(\tau_i)w_i$ denote such "deposits". If $z_i = 0$, individuals would have income $w_iY(0)$. If $z_i > 0$, income is $w_iY\{T^{-1}(z_i/w_i)\}$. Hence deposits into fund $\alpha$ have an opportunity cost $\chi(z_i) = w_i[Y(0) - Y(\tau_i)] = w_iY(0) - w_iY\{T^{-1}(z_i/w_i)\}$.

in the sense that $\chi(z_i)$ enters the budget equation for consumption, $c_i = w_iY(0) + A_i - x_i + B_i - \chi(z_i)$.

Note that $\chi'(z_i) = -Y'(\tau_i)/T'(\tau_i) = 1/\varphi(\tau_i)$, so $\chi'(z_i) - 1 > 0$ is the marginal excess burden of taxes. The excess burden acts like a one-time purchase fee for fund $\alpha$. Thus the choice between funds $A$ and $\alpha$ is equivalent to choice between a no-load mutual fund with annual fees and a fund with initial charge and zero annual cost.

This provides a finance intuition for why optimal retirement planning draws down fund $A$ before tapping into fund $\alpha$ (Prop.1), why the tradeoff involves a comparison between one-time excess burden and annual costs (Prop.2), and why there is time gap between deposits (taxes) and withdrawals in fund $\alpha$ (Prop.3-4).

### 2.3 Optimal Retirement, regardless of financing

Proposition 1 relies only on (10) whereas Propositions 2-4 depend on the excess burden in (11). This suggests that Proposition 1 applies regardless of the tax system, provided only that retirement is financed by a non-trivial combination of private assets and public pensions.

**Proposition 5** Regardless of tax system, optimal policy maximizes expected retiree utility $U_t^{R} = \sum_{i=t+1}^{\tau_{\text{max}}} \beta^i \pi_{s,t+i}u(c_{s,t+i})$ subject to given $A_{i,t} > 0$ and $\alpha_{i,t} > 0$. If $A_{i,t} > 0$ and $\alpha_{i,t} > 0$, there is at most one transitional period in which consumption is financed from a mix of public and private funds. Before the transition, public benefits are zero and consumption is funded privately. In the transitional period, private wealth declines to zero. Public funds pay for all subsequent consumption.

Prop.1 and Prop.5 express essentially the same result in complementary ways. Prop.5 requires no assumptions about taxes but is silent about conditions for $\alpha_{i,t} > 0$. Prop.1 invokes specific assumption on taxes, which are useful to show that such conditions exists (Prop.2), but assumptions about taxes necessarily restrict the scope of the result.

To illustrate that Prop.5 applies more widely than Prop.1, suppose taxes are constrained to be uniform across age, i.e., $\tau_{i,t} = \tau_{s,w,t}$ for all $i$ (for this section only). This is relevant because age- and cohort-specific payroll taxes assumed above are uncommon in practice. Note that generational accounts in retirement, $GA_{i,t+i} = -\sum_{j=i}^{\tau_{\text{max}}} \rho^{j-i} \Pi_1^{[i,j]} B_{j,t+j}$ for $i \geq i_R$, do not depend on taxes, and that generational accounts are additively separable between working age and retirement. Since each value $B_{i,t+i}$ for $i \geq i_R$ enters exactly one generation-specific components of (9), optimal policy requires that for any $GA_{i_R,t+i_R}$, retirement
benefits for each cohort maximize $E_t[U_{t+1}^{R}]$. (See appendix for how maximizing $E_t[U_{t+1}^{R}]$ fits into the overall welfare problem. Section 4 will provide additional applications of Prop.5.)

3 Numerical Analysis

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. The focus is on comparisons across scenarios and their implications for policy responses to long-run changes in the economic and demographic environment.

3.1 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 ($i_e = i_R$). 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 fixed-parameters that pay a constant benefit $B_{i,t} = \bar{B}$ for all $i \geq i_R$ and impose a constant tax rate $\tau_{i,t} = \bar{\tau}$ for all $i < i_R$. (The critique below concerns mainly the eligibility age, not the “smoothness” of benefits once they start.)

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.

3.2 Baseline Assumptions

Survival rates are taken from the U.S. Social Security Administration for the 2010 cohort, averaged over males and females. While normalizing $i_W = 0$ streamlines the theoretical analysis, natural ages are more instructive here. Hence assume working age starts at $i_W = 20$, retirement at $i_R = 66$, and the life span is $i_{max} = 120$. Population is constant, $N_t = 1$. Bonds pay $r = 0.02$ and the rate of time preference equals the bond rate ($\beta = \rho$), so consumption would be constant without management costs. Assume $\kappa_a = 0.01$ for annuitization costs, which implies annuitization starting at $i_a = 60$. Assume negligible borrowing costs, $\kappa_d = 0.0001$, non-zero only to preclude borrowing against pensions. Wages $w_i$ are proxied by an age-earning profile taken from Rupert and Zanella (2012), normalized to one at peak earnings (age 50). Work costs are $\mu_i = 0$ for $i < i_R$.

---

5This 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).
Utility has power form, \( u(c) = \frac{1}{1-\gamma} c^{1-\gamma} \) with \( \gamma = 2 \). Home production is \( H(1-l) = h - hl^{1+\varepsilon} \) with parameters \( h = 0.5 \) and \( \varepsilon = 2 \). This implies a labor supply elasticity of \( 1/\varepsilon = 0.5 \) and \( \hat{\tau} = \frac{\varepsilon}{1+\varepsilon} = 2/3 \). All calculations are done with one period consisting of two calendar years (i.e., true ages 20-21 are period 1 in the calculations, ages 22-23 is period 2, etc.). Since peak earnings are normalized to one, all real values should be interpreted as fractions or multiples of two-year peak earnings.

The generational account is calibrated so that the initial balance per-capita is equal to the revenue from a constant tax rate \( \tau_{other} = 10\% \). Non-pension expenditures are important because they determine the excess burden of pension contributions. Hence a range of values will be considered.

### 3.3 Baseline Results

Figures 1-6 show the time paths of the main variables in the baseline scenario. Optimal pensions (figure 1) are zero until age 70 and jump to 0.61 at age 74. With fixed parameters, in contrast the pension benefit is constant at 0.44, which is about 70\% of the post-74 optimal value.

The key results here are the jump in benefits and the significant interval of retirement with zero benefits. The exact timing of the jump will depend on model parameters (see sensitivity analysis below), but a specific scenario is instructive for understanding the implications.

The transitional value of 0.04 at age 72 is best interpreted as full benefit for a fractional period. That is, 0.04 for two years is equivalent to zero for 14/15 of the 2-year period and 0.61 for the remainder, so optimal benefits are best interpret as zero until age 73.9, and 0.61 thereafter. This fractional-period interpretation is used throughout the analysis below.

In Figures 2-4, private net assets (fund A), generational net assets (fund \( \alpha \)), and total net assets are similar in the optimal and fixed-parameter cases until retirement. In retirement, the optimal plan spends private assets first and depletes them by age 74. Until then, generational 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 kicks in. 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 private assets accumulate and retirement approaches. The cost of annuitization favors early-life pension contributions. In this scenario (though not in general), fixed pensions have higher tax rates, which is due to the cost of paying benefits for a longer period. For comparisons below, the overall level of optimal taxes is usefully summarized by an average tax rate \( \tau_{\text{av}} \), defined as constant rate that would yield the same lifetime revenue; here \( \tau_{\text{av}} = 23.8\% \).

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

Another perspective on welfare is to consider an unexpected reform at the time of retirement. Shifting from fixed to optimal benefits at retirement (given the asset positions under fixed parameters) would provide welfare gains equal
Table 1: Sensitivity Analysis

<table>
<thead>
<tr>
<th>Setting</th>
<th>$t_e$</th>
<th>$\tau_{aw}$</th>
<th>$B_{i&gt;t_e}$</th>
<th>$\tau$</th>
<th>$B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>73.9</td>
<td>.24</td>
<td>.61</td>
<td>.27</td>
<td>.44</td>
</tr>
<tr>
<td>$\kappa_a = 0.5%$</td>
<td>77.5</td>
<td>.20</td>
<td>.62</td>
<td>.20</td>
<td>.28</td>
</tr>
<tr>
<td>$\kappa_a = 5%$</td>
<td>73.9</td>
<td>.24</td>
<td>.60</td>
<td>.31</td>
<td>.53</td>
</tr>
<tr>
<td>$\varepsilon = 1$</td>
<td>79.8</td>
<td>.17</td>
<td>.56</td>
<td>.16</td>
<td>.18</td>
</tr>
<tr>
<td>$\varepsilon = 3$</td>
<td>71.2</td>
<td>.28</td>
<td>.66</td>
<td>.33</td>
<td>.60</td>
</tr>
<tr>
<td>$\gamma = 4$</td>
<td>73.2</td>
<td>.25</td>
<td>.63</td>
<td>.27</td>
<td>.44</td>
</tr>
</tbody>
</table>

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 eligibility age is chosen optimally, variations in “size” and in taxes translate into earlier or later eligibility age. The benefit amounts vary only to the extent that optimal consumption varies. For all scenarios considered here, the optimal age of pension eligibility appears to be in the early- or mid-70s, which is higher than 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.

### 3.5 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% reduction in mortality, which means life expectancy at age 20 increases by about 3.3 years. One scenario assumes an unchanged retirement age $i_{R} = 66$, the other assumes an increase to $i_{R} = 68$ (so $i_{e} = 0$ for $i < 68$).

If these changes occur unexpectedly, intergenerational risk sharing suggests that young and future cohorts should at least partially insure older cohorts (see Bohn (2001), 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 eligibility age, the average tax rate in the optimal system, optimal benefits for $i > i_{e}$, 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 eligibility 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, eligibility at 66 implies a cut in benefits from .44 to .39.

In the scenario with retirement at 68, the optimal pension system offers slightly greater benefits are an even higher eligibility age. The intuition is that two more years with labor income raise the optimal consumption profile at all ages, though only slightly. The optimal eligibility age rises because the

---

#### Table 2: Optimal pensions when mortality declines

<table>
<thead>
<tr>
<th>Mortality</th>
<th>$i_{R}$</th>
<th>$i_{e}$</th>
<th>$\tau_{av}$</th>
<th>$B_{i &gt; i_{e}}$</th>
<th>$\bar{\tau}$</th>
<th>$\bar{B}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>66</td>
<td>73.9</td>
<td>.238</td>
<td>.610</td>
<td>.270</td>
<td>.44</td>
</tr>
<tr>
<td>25% less</td>
<td>66</td>
<td>75.9</td>
<td>.239</td>
<td>.604</td>
<td>.267</td>
<td>.39</td>
</tr>
<tr>
<td>25% less</td>
<td>68</td>
<td>76.1</td>
<td>.235</td>
<td>.607</td>
<td>.273</td>
<td>.45</td>
</tr>
</tbody>
</table>
benefits to support higher consumption can only be paid for a somewhat shorter period. In a fixed-parameter system, the eligibility age would simply rise to 68 and benefits would be similar to the baseline; thus benefits are sensitive to the retirement 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%. If benefits are kept unchanged, the present value of optimal benefits would increase by 14.6% and cover about 96% of the funding needs, keeping consumption essentially unchanged. The intuition is that 96% 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%, roughly in proportion to the present value of consumption. Per-period consumption would have to decline by about 2%.

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.

3.6 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 (9), 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 \( GA_{0,t} > 0 \)), small cohort sizes would also tighten (9). Low birth rates are then a budgetary problem and also 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 convenience, they are calibrated in terms of tax rates required for non-pension expenditures (\( \tau_{other} \)), and stated as balances \( GA_{iw,t} \) at the start of work. By construction, one feasible policy in each scenario is to provide no pension and to impose only \( \tau_{other} \). 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 .68 benefits at age 70. The high-tax case has lower benefits that start later. Note that the average tax rates in table 2 include \( \tau_{other} \), so taxes for pensions decline from 18% in the low-tax case to 9% 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
Table 3: Optimal pensions in relation to other taxes

<table>
<thead>
<tr>
<th>Setting</th>
<th>$\tau_{other}$</th>
<th>$GA_{i,w}$</th>
<th>$t_r$</th>
<th>$\tau_{av}$</th>
<th>$B_{i&gt;1}$</th>
<th>$\tau$</th>
<th>$B$</th>
<th>$\Delta c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-tax</td>
<td>0%</td>
<td>0</td>
<td>70.0</td>
<td>18%</td>
<td>.68</td>
<td>22%</td>
<td>.64</td>
<td>0.2%</td>
</tr>
<tr>
<td>Baseline</td>
<td>10%</td>
<td>0.98</td>
<td>73.9</td>
<td>24%</td>
<td>.51</td>
<td>27%</td>
<td>.44</td>
<td>0.6%</td>
</tr>
<tr>
<td>High-tax</td>
<td>20%</td>
<td>1.85</td>
<td>79.3</td>
<td>29%</td>
<td>.54</td>
<td>28%</td>
<td>.18</td>
<td>0.7%</td>
</tr>
</tbody>
</table>

would offer comparable benefits (.64) starting earlier (age 66); in the high-tax case, benefits would be drastically lower (.18).

The column marked $\Delta c$ in Table 3 shows welfare differences between optimal and fixed-parameter system in terms of lifetime consumption. The welfare difference are similar for $\tau_{other}$ at 10% and 20%. Welfare differences are smaller when $\tau_{other} = 0$ because optimal pensions then start fairly early and fixed-parameter pensions are relatively generous, so they are more similar 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 long-term changes in budget conditions by significantly varying the eligibility age, which allows the government to keep the per-period benefit amount much more stable that it could afford under a fixed eligibility age.

4 Special Cases and Extensions

This section examines several special cases and extensions. The special cases are settings in which retirement funding is either entirely private ($\alpha_i = 0$), assuming frictionless private markets, or entirely public (so $A_i = 0$), assuming non-distortionary public pensions. The special cases suggest that the investment frictions and tax assumed in the basic model are necessary for modeling a setting with a mix of public and private retirement funding. The extensions examine partial commitment to public pensions and cross sectional heterogeneity.

4.1 Frictionless Capital Markets: No Public Pensions

For this section, consider the limiting case $\kappa_a = \kappa_d = 0$. Then (4) and (6) imply $\beta Ru'(c_{i+1}) = u'(c_i)$ for all $(t, i)$. Moreover, $\Pi_i < 1$ in (5) implies $x^v_t = 0$. This confirms that unless there are financial frictions, Yaari’s full-annuitization result applies (Yaari (1965)). Moreover, iterating on asset accumulation (3) implies an intertemporal budget constraint

$$\sum_{i=1}^{i_{max}} \rho^{t-i_{o}} \Pi_{[i_{o}, i]} c_i = A_{i_{o}} + \sum_{i=1}^{i_{max}} \rho^{t-i_{o}} \Pi_{[i_{o}, i]} (y_i + B_i) \equiv PVLR_{i_{o}}.$$
maximizing \( W \) and let \( GA \). Planning horizon \( t \), saving through \( x \), subject to the same constraints as in Section II.2, except that \( GA \neq 0 \). If \( GA \leq 0 \), efficiency requires \( \tau_i = 0 \) for all \( i \), and the timing of benefits is arbitrary. If \( GA > 0 \), efficiency requires \( B_i = 0 \) for all \( i \). Optimal taxes conditional on work must minimize excess burden, so \( \tau_i = \bar{\tau} \) is constant for \( i < i_R \).

In summary, optimal policy will either impose taxes on a cohort or provide transfers, but never both. A pension system, in the traditional sense of a system that provides benefits financed by distortionary taxes, would necessarily be inefficient.\(^7\) Hence some financial frictions are essential for modeling public pensions.

### 4.2 Frictionless Government Pensions: No Private Saving

For this section, suppose the government could accept voluntary pension contributions and commit to repay them as an annuity stream with the same present value, using \( r \) as discount rate.

Let voluntary contributions be denoted \( \hat{T}_i \) and tracked as individual account \( \hat{\alpha}_i \), which accumulates according to

\[
\hat{\alpha}_{i+1} = \frac{R}{\Pi} \cdot \left[ \hat{\alpha}_i + \hat{T}_i - \hat{B}_i \right],
\]

where \( \hat{B}_i \) denote payouts. Adding this option, the cohort’s problem becomes a savings problem with three investment funds. The Bellman equation with planning horizon \( j = i_{\text{max}} - i \geq 1 \) is

\[
V_{j+1}(A_i, \alpha_i, \hat{\alpha}_i) = \max \{ u(c_i) + \beta \Pi \cdot V_j(A_{i+1}, \alpha_{i+1}, \hat{\alpha}_{i+1}) \}
\]

subject to the same constraints as in Section II.2, except that \( x_i = A_i + B_i - c_i + (Y(\tau_i)w_i - \mu_i) + \hat{B}_i - \hat{T}_i \), and the boundary condition is \( V_1(A_{i_{\text{max}}}, \alpha_{i_{\text{max}}}, \hat{\alpha}_{i_{\text{max}}}) = u(A_{i_{\text{max}}} + \alpha_{i_{\text{max}}} + \hat{\alpha}_{i_{\text{max}}}) \). Since \( R_{i+1}(x_i) < R/\Pi \) for all \( x_i > 0 \), saving through \( \hat{\alpha} \)

\(^7\)To complete the optimal policy argument, let \( U_t(GA_{i_{t+1}}) \) for \( GA_{i_{t+1}} \leq 0 \) and \( U_t(\tau_i) \) for \( GA_{i_{t+1}} \geq 0 \) denote the maximum utility levels of generation \( t \) as function of benefits or taxes and let \( \Lambda_{IBC} \) denote the shadow value of resources in (9). Then the first-order conditions for maximizing \( W_t \) imply that benefits are paid if \( \beta^{t-\omega} \omega_1 \frac{\partial U_t}{\partial A_{t_{t+1}}} > \beta^{t-\omega} \omega_1 \Lambda_{IBC} \) would apply at zero benefits; and taxes are imposed if \( \beta^{t-\omega} \omega_1 \frac{\partial U_t}{\partial \tau_t} < \beta^{t-\omega} \omega_1 \Lambda_{IBC} \) would apply at zero taxes. This suggests that a wide range of taxes and generational account values are consistent with efficiency, in the sense that they could be attributed to high or low welfare weights. However, since revenues are needed for non-pension expenditures, one may suspect that optimal policy will require positive taxes unless welfare weights are very uneven.
return-dominates private saving; since $\chi'(z_i) > 0$, tax-financed public pensions are more costly than a system with individual accounts. Hence all retirement saving is optimally done through individual accounts, so $A_i \leq 0$ at all times.

A credible promise to repay voluntary contributions is critical for this result and arguably unrealistic. If the government retains discretion over other policies, one must suspect that such promises are not time-consistent. For example, suppose $B_i > 0$ and suppose members of the same cohort had different account balances $\hat{\alpha}_{iR}$ at retirement. Then on the margin, the government would optimally reduce discretionary benefits $B_i$ to the individual with higher $\hat{\alpha}_{iR}$.

In contrast, the policy in Section II is time-consistent. That is, if the government re-optimized at time $t_0 + 1$, optimality conditions for variables dated $t \geq t_0 + 1$ are the same as at time $t_0$. This feature motivates in part why the basic model in Section 2 assumes discretion.

4.3 Limited Commitment: Earnings-linked pensions

Public pensions in many countries are linked to individual earnings. This suggests that governments have to some extent solved the time-consistency problem noted in the previous section. Hence it is worth documenting that the principal results of the basic model generalize to a setting with earnings-linked pensions. The critical assumption is that the earnings-linkage provides less than a full repayment of contributions, an assumption satisfied by construction for all pensions systems that operate under pay-go or are less than fully funded.

Let $\phi_i$ 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. Throughout, assume limited credibility imposes an upper bound $\phi_i \leq \phi < 1$ low enough to preclude voluntary taxes.

A payroll tax of $\tau_i w_i l_i$ can then be divided into an individual-account contribution of $\hat{T}_i = \phi_i \tau_i w_i l_i$ and a contribution of $(1 - \phi_i) \tau_i w_i l_i$ to the cohort’s generational account. That is, earnings-credits are treated as individual account balances that accumulate as in (14) and promise annuitized benefits.

A credible earnings-linkage influences optimal labor supply, because on the margin, an increase in $l_i$ implies a higher pension. The valuation of pension promises is non-trivial, however, because due to costly individual annuitization, the value of an annuitized pension to individuals generally differs from the cost to the government.

For the accounting, assume earnings-credits earned in different years are allocated proportionally to pension payments made during retirement; this

---

8 One exception is the commitment to honor debt, which points to an important distinction between debt and individual pensions. Debt is generally marketable and widely held, which makes selective default practically impossible, whereas annuitized pensions are necessarily individualized.

9 Linkages $\phi_i$ 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 indexed to, say, aggregate wages (U.S. case), $\phi_i$ varies whenever the index grows differs from the discount rate.
is common practice and without loss of generality since pensions will be optimized below. Let \( b_j \in [0, 1] \) denote the present-value shares of earnings-linked pensions per unit of \( \hat{\alpha}_{iR} \), so \( \hat{B}_j = b_j \hat{\alpha}_{iR} R^{j-iR} \) is paid at age \( j \) and \( \sum_{j=iR}^{i_{\text{max}}} \hat{b}_j = 1 \). The present value per unit of \( \hat{\alpha}_{iR} \) to individuals is then \( v_{iR} = \sum_{j=iR}^{i_{\text{max}}} \hat{b}_j \prod_{k=i}^{j} \frac{R_k}{\Pi_k} \), which account for credit-accumulation in (14) at rate \( R/\Pi_k \) and discounting at rate \( R_{k+1}(x_k) \). By definition of \( R_{k+1}(x_k) \), this case be written in terms of cumulative costs as \( v_{iR} = \sum_{j=iR}^{i_{\text{max}}} \hat{b}_j / K_{iR} \). Discounting similarly to the various years of contribution, the earnings-link at age \( i \) is valued at

\[
v_i = \sum_{j=iR}^{i_{\text{max}}} \frac{\hat{b}_j}{\prod_{k=i}^{j} K_{k+1}(x_k)} \tag{15}
\]

If \( x_k \geq 0 \), the products reduce to \( K_{iR,j} \). Then \( K_{iR,j} < 1 \) implies \( v_i > 1 \), so annuities are valued higher than their cost. The assumption of no voluntary contributions can then be formalized as \( \phi < 1/\max_i \{ v_i \} \).

Given \( v_i \), labor supply maximizes \( y_i + v_i \phi_i \tau_i \hat{w}_i \hat{d}_i \), which implies \( 1 - \tau_i - H'(1-l_i) + v_i \phi_i \tau_i = 0 \) or \( l_i = 1 - (H')^{-1} [1 - (1 - v_i \phi_i) \tau_i] \). This is identical to labor supply in the basic model at tax rate \( (1 - v_i \phi_i) \tau_i \). Thus the earnings-linkage reduces tax distortions proportionally by \( v_i \phi_i \).

Policy is again a welfare-maximization problem as in Section 2, but with \( \phi_i \) and \( b_j \) as additional policy choices. The focus here is on the optimal structure of earnings-linked pensions. (A full analysis would distract from the main focus.) Minimizing tax distortions is equivalent to maximizing \( v_i \phi_i \). For any given \( \phi_i \), maximizing \( v_i \) for any \( i \) is equivalent to maximizing \( v_i \phi_i \) in (15) by choice of \( b_j \in [0, 1] \). Since \( K_{k+1}(x_k) < 1 \) for \( x_k > 0 \), the optimal solution is to provide sequential funding with some transition period \( i_e \) : \( b_j = 0 \) for all \( j < i_e \), \( b_j > 0 \) for all \( j > i_e \) calibrated \( \hat{b}_j \hat{\alpha}_{iR} R^{j-iR} = B_j = c_j \) to fully fund consumption. In the transition \( i_e \), \( \hat{b}_i = 1 - \sum_{j>i_e}^{i_{\text{max}}} \hat{b}_j \) provides partial funding \( (0 < B_j \leq c_j) \). To summarize:

**Proposition 6** In a pensions system with commitment to pay an earnings-linked pension, there is at most one transitional period in which consumption is financed from a mix of promised public and other funds (private saving or discretionary benefits). Before the transitional period, promised benefits are zero. After the transitional period, promised benefits pay for all consumption.

Thus the structure of pensions is identical to discretionary pensions. Since the arguments are entirely about retirement, they apply analogously for alternative assumptions about taxes, e.g., if taxes were restricted to be uniform across ages. Minimizing the excess burden of any given tax requires maximizing \( v_i \phi_i \), and hence Proposition 6 applies.

---

10The polar case \( \phi_i = \phi = 1 \) would yield the frictionless government pensions discussed above. Cases with \( \phi_i < 1 \) and \( v_i \phi_i > 1 \) are excluded because they would require tedious case distinctions.
To clarify, Prop.6 applies even if a cohort receives both earnings-linked and discretionary benefits. To maximize labor supply incentives, earnings-linked benefits are optimally paid at the end (ages \([i_{e}, i_{\max}]\)). Discretionary benefits would cover an interval \([i_{e}, i_{\max}]\) ending when earnings-linked benefits start; and benefits in Prop.5 should be interpreted as total benefits \(B_{i} + \tilde{B}_{i}\).

### 4.4 Cross-Sectional Heterogeneity

Now suppose each cohort is a continuum of members with a distribution of labor productivities. Let \(w_{i,t}\) now denote a cohort’s average wage and let relative wages \(\eta\) have a distribution \(F\) with unit mean \(\int \eta dF(\eta) = 1\). For simplicity, \(F\) is assumed time-invariant and relative earnings are fixed over the life-cycle.\(^{11}\)

Individuals maximize utility \(E_{t}\left[U_{t}(\eta)\right]\), earn income \(y_{i,t+i}(\eta)\) as in (1), and accumulate assets according to

\[
\begin{align*}
    x_{i,t+i}(\eta) &= A_{i,t+i}(\eta) + B_{i,t+i}(\eta) + \tilde{B}_{i,t+i}(\eta) - c_{i,t+i}(\eta) + y_{i,t+i}(\eta) \\
    A_{i,t+i}(\eta) &= R_{t+i}[x_{i-1,t+i-1}(\eta) - x_{i-1,t+i-1}(\eta)]
\end{align*}
\]

Assume a welfare function

\[
W_{t_{0}} = \sum_{t=t_{0}}^{\infty} \beta^{t-t_{0}} \int \omega_{t}(\eta)E_{t_{0}}[U_{t}(\eta)]dF(\eta).
\]

with weights that \(\omega_{t}(\eta)\) may depend on \(\eta\) and on time.

Let taxes be linear – either age-specific or uniform – and assume promised earnings-linked \(\phi_{i}(\eta)\) (if any) are constant over time. Discretionary benefits \(B_{i,t+i}(\eta)\) may depend on age, time, and the individual’s relative wage. Promised benefits are again defined by share parameters \(b_{i}(\eta)\), so \(\tilde{B}_{i,t+i}(\eta) = \tilde{b}_{i} \tilde{\alpha}_{i}(\eta) R_{t+i-1}^{1-R_{i}}\), where \(\tilde{\alpha}_{i}(\eta)\) is accumulated as in (14) from credits \(\tilde{T}_{i,t+i}(\eta) = \eta \phi_{i}(\eta) T_{i,t+i} w_{i,t+i}(\eta)\).

It is straightforward to show that, provided an optimal policy exists, retirement benefits for each \(\eta\) must have the same structure as in the homogenous agent case. (See appendix.) That is, policy maximizes \(E_{t_{0}}U_{t}^{R}(\eta)\) at each productivity level \(\eta\), and retirement consumption is financed privately up to a transition period \(i.e_{t}(\eta)\). The eligibility dates \(i.e_{t}(\eta)\) generally differ cross-sectionally.

But once the pension benefit starts for group \(\eta\), it covers all optimal consumption in all subsequent periods. If a group receives both earnings-linked and discretionary benefits, optimal discretionary benefits again cover the period before earnings-linked benefits start.

### 4.5 Endogenous Retirement Age

Suppose cost of working increases gradually so \(\tilde{Y} < \mu_{i}/w_{i} \leq Y(0)\) applies for some age range \([i_{R1}, i_{R2}]\). Then individually rational retirement depends on

\(^{11}\)One-dimensional heterogeneity should suffice here, as this section is meant to sketch how the basic model generalizes, not to prove the general results.
taxes. Because retirees pay no taxes, optimal taxes must not trigger premature retirement and hence satisfy $\tau_i \leq \bar{\tau}_i = Y^{-1}(\mu_i/w_i)$. Subject to these additional conditions, optimal policy would determined as in the main model (with $i_R = i_{R2}$).

In practice, these additional constraints appear to be violated, especially in pension systems that provide benefits at retirement and conditional on not working. (see Gruber and Wise (2008)). One reason may be rigidity: due to medical improvements, $\mu_i$ is likely to decline over time, so the age range $[i_{R1}, i_{R2}]$ shifts upward. A pension system with substantial gap between $i_R$ and $i_e$ would automatically avoid these problems.

5 Conclusions

The paper shows that optimal retirement financing is sequential – entirely private until a transition period and then entirely public. For any given present value of public pensions, generous coverage of consumption needs starting late in life is more efficient than a smaller benefit starting at retirement. Hence the optimal eligibility age for pensions is generally higher than the retirement age.

These results is shown formally in a model in which private savings have recurrent costs whereas public pensions incur a one-time excess burden. In this setting, pensions have a comparative advantage at long horizons, whereas private saving are more efficient at short horizons.

More broadly, financial planning for an open-ended lifetime poses daunting problems even for sophisticated and well-informed savers. Realistically, many savers will make costly mistakes. If the government covered the end of life, individuals could limit their attention to the period until benefits start, which would much simplify their planning.

The idea of late-life benefits is not radical in a historical context. When public pensions were first introduced in Germany in the 1890s, the eligibility age of 65 was close to adult life expectancy, conditional on survival to adulthood. When the U.S. social security system was created in 1935, the eligibility age of 65 was comparable to life expectancy at birth. Remarkably, the eligibility age has stayed constant in many countries even as life expectancy has increased over time (or at most rising slightly to 67). If eligibility ages had been adjusted to keep up with life expectancy, pension eligibility in most developed countries would now be around 75.

References


A Appendix

A.1 Proofs of Propositions1-5

Proof 1. If $B_j > 0$ for any age $j \in [i_R, i_{\text{max}})$, then $\Lambda_{B_j} = 0$ in (10), hence $\Lambda_{GA} = \beta R^j u'(c_j)$. Moreover, $\Lambda_{B_{j+1}} \geq 0$ in (10) implies $\beta R^{j+1} u'(c_{j+1}) \leq \Lambda_{GA}$, so $\beta R u'(c_{j+1}) \leq u'(c_j)$. Then in (4), $\kappa_a > 0$ implies $\Lambda_{x_j} = u'(c_i) -$
Proof 2. Assume for contradiction that $B_i = 0$ for all $i$. Since $u'(0) = \infty$, $c_i > 0$ for all $i \geq i_R$, which implies recursively that $x_i = (c_{i+1} + x_{i+1}) / R_i(x) > 0$ for all $i \geq i_R - 1$, which implies $R_i(x_i) = \frac{1 + r}{1 + \Pi_i - \kappa_i}$. Hence $R_i(x_i)u'(c_{i+1}) = u'(c_i)$ is the first-order condition for optimal consumption for all $i \geq i_R - 1$. Using the definition of $K_{i,i_R}$, $K_{i,i_{1,R}}[(1 + r)\beta]u'(c_{i_{1,R}}) = u'(c_i)$ for $i = i_R - 1$; the same applies analogously for $i < i_R - 1$ if $x_j > 0$ for $j \in [i, i_R - 1]$. From (11), $u'(c_i) \geq \frac{\Delta_{GA}^i}{\beta^i(1+r)} \varphi(\tau_i)$ and $u'(c_{i_{1,R}}) \leq \frac{\Delta_{GA}^i}{\beta^i(1+r)} \varphi(\tau_i)$, which contradicts the assumption $K_{i,R_{i_{1,R}}} < \varphi(\tau_i)$. If benefits are paid at any age $i < i_{1,R}$, then $\Delta_{GA} = (\beta(1+r))^i u'(c_{i_{1,R}})$ and $\beta(1+r)u'(c_{i_{1,R}}) \leq u'(c_i)$, which in turn implies $x_j^* = x_j^* = 0$.

Proof 3. Since $\tau_{i_{1,R}} > 0$, (11) implies $u'(c_{i_{1,R}}) = \frac{\Delta_{GA}^i \varphi(\tau_{i_{1,R}})}{\beta^i(1+r)}$ and hence $\beta^i(1+r)u'(c_i) \leq \Delta_{GA}^i \varphi(\tau_{i_{1,R}}) / K_{i,R_{i_{1,R}}}$. When $\varphi(\tau_{i_{1,R}}) < K_{i,R_{i_{1,R}}}$, this implies $\beta^i(1+r)u'(c_i) \leq \Delta_{GA}^i$, so $\Delta_{B_i} > 0$ and hence $B_i = 0$.

Proof 4. Follows directly from (11) when $\tau_{i_{1,R}} > 0$.

Proof 5. By assumption, $T(\tau_{j_{1,i+j}})$ in (9) replaced by another tax system, $y_i = w_iY(\tau_i)$ is replaced by the implied values of market income, and the resulting welfare problem has a solution. Note that the generational accounts in (9) separate additively into $GA_{0,t} = GA_{work,t} + \rho^{i_R}GA_{i_{1,R},t+i_{1,R}}$, where $GA_{work,t} = \sum_{j=0}^{i_{1,R}} \rho^{-i} \Pi_{j,i_R}(T(\tau_{j+i_R})u_{j+t+j} - B_{j+t+j})$. Note that $GA_{i_{1,R}} = -\sum_{j=1}^{i_{1,R}} \rho^{i-j} \Pi_{i_{1,R}}(B_{j+i_R})$ for $i \geq i_{1,R}$ does not depend on taxes and that each value $B_{j+i_R}$ for $i \geq i_{1,R}$ enters exactly one generation-specific component of (9). Hence optimal policy requires that retirement benefits $\{B_{j+i_{1,R}}\}_{j=0}^{i_{1,R}}$ for each cohort $t \geq t_0 - i_{1,R}$ maximize $E_{t+i_{1,R}}[U_t]$ subject to given $GA_{i_{1,R},t+i_{1,R}}$, where $U_t = \sum_{j=1}^{i_{1,R}} \beta^j \sigma_{j+i_R}u_{j+i_R}$ denotes utility in retirement. For $t < t_0 - i_{1,R}$, analogous reasoning implies that policy maximizes $E_{t_0}[U_{t_0}]$ at time $t_0$, subject to given $GA_{i_{1,R},i_R}$, $i = t_0 - t$ at that time. The claims about transition are non-trivial if $\alpha_{i_{1,R}} > 0$ and $\alpha_{i_R} > 0$. Since $\alpha_{i_R} > 0$ implies $B_i > 0$ for some $i \geq i_{1,R}$, the claims follow from Prop. 1.

(Prop.6 was proved in the text.)
A.2 The Welfare Problem with Uniform Taxes

The overall problem is to maximize $W_0$ subject to (9), (8), and (7) by choice of \{ci,t+i, xi,t+i, τi,t+i, Bi,t+i\} ≥ 0, t ≥ t0 for given $D_0 = 0$, \{Ga\} ≥ t0 at $t = t_0$, and $A_{0,t} = 0$ for $t > t_0$. The additional incentive constraints (4)-(6) are moot since they will be satisfied at the solution.

If the $τ_{i,t+i}$ are unrestricted, the welfare problem subdivides straightforwardly as discussed in Section 2. With uniform taxes, the problem can be divided instead into (A) maximizing over retirement choices \{ci,t+i, xi,t+i, Bi,t+i\} ≥ 0, t ≥ t0 conditional on working-age choice \{ci,t+i, xi,t+i, τi,t+i, Bi,t+i\} ≥ 0, i < tR, t ≥ t0; and (B) maximizing over the conditioning variables \{ci,t+i, xi,t+i, τi,t+i, Bi,t+i\} ≥ 0, i < tR, t ≥ t0.

To do this, first separate divide utility and welfare into working-age and retirement components, $U_t = U^\text{work}_t + U^R_t$, where $U^\text{work}_t = \sum_{i=0}^{\text{max}} \beta^i \sigma_{i,t+i} u(c_{i,t+i})$ and $U^R_t = \sum_{t=t_0}^{\text{max}} \beta^t \sigma_{t+i} u(c_{i,t+i})$, and $W_0 = W^\text{work}_0 + W^R_0$ where $W^\text{work}_0 = \sum_{t=t_0}^{t_0} \beta^t \omega_t E_t[U^\text{work}_t]$ and $W^R_0 = \sum_{t=t_0}^{\text{max}} \beta^t \omega_t E_t[U^R_t]$.

Second, note that the generational accounts divide similarly: for generic $t ≥ t_0$, $GA_{0,t} = GA_{\text{work},t} + \rho^t GA_{\text{ret},t+iR}$ was defined in the text; for $t ∈ [t_0 - iR + 1, t_0 - 1]$, the same division applies with truncated work period; and for $t < t_0 - iR$, the retirement period is truncated and $GA_{\text{work},t} = 0$ is moot.

Third, note that \{ci,t+i, xi,t+i, τi,t+i, Bi,t+i\} ≥ 0, i < tR, t ≥ t0 imply values of $GA_{iR,t+iR}$ and $A_{iR,t+iR}$ for generations $t ≥ t_0 - (iR - 1)$, and values $GA_{t_0-t_0}^R$ and $A_{t_0-t_0}$ for generations $t ≤ t_0 - iR$, 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 $W^R_0$ subject $GA$- and $A$-values at retirement. For generations $t ≥ t_0 - (iR - 1)$, the relevant variables are $GA_{iR,t+iR}$ and $A_{iR,t+iR}$; for generations $t ≤ t_0 - iR$, they are $GA_{t_0-t_0}$ and $A_{t_0-t_0}$.

Moreover, since \{ci,t+i, xi,t+i, Bi,t+i\} ≥ 0R for any $t$ enters only into $U^R_t$ and $GA_{iR,t+iR}$ for one generation, the problem of maximizing $W^R_0$ separates into generation specific problems of maximizing $E_0U^R_t$. Specifically:

(A1) For $t ≥ t_0 - (iR - 1)$, maximize $E_0U^R_t$ subject to (8) and (7) by choice of \{ci,t+i, xi,t+i, Bi,t+i\} ≥ 0R for given $GA_{iR,t+iR}$ and $A_{iR,t+iR}$. The solutions define indirect utility functions

$$V^R_{iR,t}(GA_{iR,t+iR}, A_{iR,t+iR}) = \max \{E_0 \sum_{i=iR}^{\text{max}} \beta^i \sigma_{i,t+i} u(c_{i,t+i})|GA_{iR,t+iR}, A_{iR,t+iR}\}.$$

(A2) For $t ∈ [t_0 - \max, t_0 - iR]$, maximize $E_0U^R_{t_0}$ subject to (8) and (7) by choice of \{ci,t+i, xi,t+i, Bi,t+i\} ≥ 0R for given $GA_{t_0}$ and $A_{t_0}$. The solutions define indirect utility functions

$$V^R_{t_0-t,t}(GA_{t_0-t_0}, A_{t_0-t_0}) = \max \{E_0 \sum_{i=t_0-t}^{\text{max}} \beta^i \sigma_{i,t+i} u(c_{i,t+i})|GA_{t_0-t_0}, A_{t_0-t_0}\}.$$

Given solutions to (A1) and (A2), the remaining welfare problem is to max-
imize

\[ W_{t_0} = W_{t_0}^{work} + \sum_{t=1_0-i}^{\infty} \beta^{t-t_0} \omega_t V_{s_0-t}^R (G_{j, t+1} A_{j, t+1} A_{i, t+1} A_{J, t+1}) \]
\[ + \sum_{t=1_0-i}^{\infty} \beta^{t-t_0} \omega_t V_{s_0-t}^R (G_{j, t-1} A_{j, t-1} A_{i, t-1} A_{J, t-1} A_{J, t-1}) \]

Since taxes matter only at this stage, results about retirement implied by (A1) and (A2) do not involve assumptions about taxes.

A.3 Results with Heterogenous Productivity

By assumption, an allocation that maximizes \( W_{t_0} \) exists. The text claims that retirement benefits must be structured as in Prop.5 and Prop.6.

Proof for earnings-linked benefits: If the sequence \( \{ b_i(\eta) \}_{i \geq i} \) failed to satisfy Prop.6, one obtains a contradiction to optimality as follows: Define \( i_\epsilon = \min \{ i : b_i(\eta) > 0 \} \) as the first period with non-zero payments. Suppose \( B_{i, t+i}(\eta) \neq c_{i, t+i}(\eta) \) for any \( i > i_\epsilon \), which would violate Prop.6. Then also \( B_{i, t+i}(\eta) = c_{i, t+i}(\eta) \) for some \( i > i_\epsilon \), since \( B_{i, t+i}(\eta) - c_{i, t+i}(\eta) > 0 \) implies private assets that must be spent subsequently. The consider a marginal increase in \( b_{i, t+i}(\eta) \) at \( i = i_\epsilon \), combined with \( \Delta b_{i, t+i}(\eta) = -\Delta b_{i, t+i}(\eta) \) at \( i = i_\epsilon \) and \( \Delta A_i(\eta) = -\Delta A_i(\eta) \) for \( i \in [i_\epsilon + 1, i_\epsilon] \); note that \( A_i(\eta) > 0 \) is implied by \( B_{i, t+i}(\eta) < c_{i, t+i}(\eta) \), so \( \Delta A_i(\eta) < 0 \) is feasible and \( K^{[i, i-]} < 1 \). The change \( \Delta b_{i, t+i}(\eta) \) keeps \( \alpha_i(\eta) \) unchanged, so promises are kept. The change \( \Delta A_i(\eta) \) implies \( c_{i, t+i}(\eta) \) for all \( i \geq i_\epsilon \) remains feasible. Since \( K^{[i, i-]} < 1 \), \( v_{i, t+i}(\eta) = \sum_{j=i_\epsilon}^{i_\epsilon} b_j(\eta) / K^{[i, i-]} \) changes by \( \Delta V_{i, t+i}(\eta) = \Delta b_{i, t+i}(\eta) + \Delta b_{i, t+i}(\eta) / K^{[i, i-]} > 0 \). The increase implies a shift to market labor and higher tax revenues that can be used for additional transfers that raise \( W_{t_0} \), contradicting the optimality of the original allocation.

QED.

Proof for discretionary benefits: Let total benefits be \( B_{i, t+i}(\eta) = B_{i, t+i}(\eta) + B_{i, t+i}(\eta) \), where \( B_{i, t+i}(\eta) = c_{i, t+i}(\eta) \) for \( i > i_\epsilon \) and \( B_{i, t+i}(\eta) < c_{i, t+i}(\eta) \) for \( i \leq i_\epsilon \) as above, and assume \( B_{i, t+i}(\eta) > 0 \) for some \( i \). If \( B_{i, t+i}(\eta) \) failed to satisfy Prop.5, one obtains a contradiction to optimality as follows: Define \( i_\epsilon = \min \{ i : B_{i, t+i}(\eta) > 0 \} \). Suppose \( B_{i, t+i}(\eta) \neq c_{i, t+i}(\eta) \) for any \( i > i_\epsilon \), which would violate Prop.5. Then \( i_\epsilon = \max \{ i > i_\epsilon : B_{i, t+i}(\eta) \neq c_{i, t+i}(\eta) \} \) exists. Moreover, \( B_{i, t+i}(\eta) < c_{i, t+i}(\eta) \) because \( B_{i, t+i}(\eta) - c_{i, t+i}(\eta) > 0 \) would imply positive private assets and hence \( c_{i, t+i}(\eta) > B_{i, t+i}(\eta) \) in a subsequent period. Also, \( i_\epsilon \leq i_\epsilon \) because of Prop.6. Hence \( B_{i, t+i}(\eta) = B_{i, t+i}(\eta) \) for all \( i < i_\epsilon \), and hence \( B_{i, t+i}(\eta) > 0 \). This implies a deferral in pensions: a marginal increase \( \Delta B_{i, t+i}(\eta) > 0 \) combined with \( \Delta B_{i, t+i}(\eta) = -\Delta B_{i, t+i}(\eta) + R^{t+i} - \epsilon < 0 \) is feasible and leaves \( G_{i, t+i}(\eta, \eta) \) unchanged. Moreover, \( B_{i, t+i}(\eta) < c_{i, t+i}(\eta) \) implies \( A_i(\eta) > 0 \) for all \( i \in (i_\epsilon, i_\epsilon) \), which means the sequence \( A_i(\eta) \) can be reduced marginally to keep \( c_i(\eta) \) unchanged.
for all \( i \in (i_e, i_\bar{e}) \). Since \( K^{[i_e, i_\bar{e}]} < 1 \), the implied \( \Delta x_{i_e}(\eta) \) is smaller than \( \Delta B_{i_e, t+i_e}(\eta) \). Hence \( \Delta c_{i_e}(\eta) = -\Delta x_{i_e}(\eta) + \Delta B_{i_e, t+i_e}(\eta) > 0 \) and \( \Delta U^R_t > 0 \), contradicting optimality. QED.

Final remark: The central argument for both types of benefits is that a marginal deferral, holding constant the present value to government, will increase the value to individuals. This argument applies until benefits are deferred to the maximum extent possible, which provides the intuition for why benefits should cover all consumption in the final periods of life. Note that the argument does not depend on redistribution, which explains why the basic model abstracts from cross-sectional heterogeneity.