Mortality Regressivity and Pension Design*

Youngsoo Jang[†] Svetlana Pashchenko [‡] Ponpoje Porapakkarm[§]

July 15, 2023

Abstract

How should we compare welfare across pension systems in presence of differential mortality? A commonly used standard utilitarian criterion implicitly favors the long-lived over the short-lived. We investigate under what conditions this ranking is reversed. We clearly distinguish between the redistribution along mortality and income dimensions, and thus between mortality and income progressivity. We show that when mortality is independent of income, mortality progressivity can be optimal only when (i) there is more aversion to inequality in lifetime utilities compared to aversion to consumption inequality, (ii) life is valuable. When the short-lived tend to have lower income, mortality progressivity can be also optimal when income redistribution tools are limited. In this case, mortality progressivity is used to substitute for income progressivity.

Keywords: Mortality-related redistribution, Pensions, Social Security, Annuities, Life-Cycle Model

JEL Classification Codes: G22, H21, H55, I38

*Pashchenko acknowledges financial support from the Terry-Sanford Award.

[†]University of Queensland; Email: youngsoo.jang@uq.edu.au

[‡]University of Georgia; Email: svetlanap.econ@gmail.com

[§]National Graduate Institute for Policy Studies; Email: p-porapakkarm@grips.ac.jp

1 Introduction

How should we compare welfare across different pension designs when people have unequal lifespans? A common approach is to use the standard utilitarian welfare theory. One important implication of this theory is that optimal pension benefits are independent of expected longevity. This means that, everything else equal, people who live longer will receive more in total payments, implying the redistribution from the short-lived to the long-lived.

One obvious objection to this type of redistribution is that due to the negative income-mortality correlation it is income-regressive. A broader question, however, is whether the redistribution from the short-lived to the long-lived is desirable even if income and mortality were not correlated.

The problem of how to account for contribution of non-economic factors such as lifespan to general welfare has been discussed since at least Atkinson and Bourguignon (1982), but the standard utilitarian welfare function is still widely used for policy analysis. Recently, a growing number of structural studies have been incorporating differential mortality into the quantitative analysis of pension systems (e.g., Jones and Li, 2022; Laun et al., 2019; Sanchez-Romero, 2019). However, the question of how the implicit bias towards the long-lived embedded in the standard utilitarianism affects welfare measurements has largely been left out of this discussion.

In this paper, we develop a general framework to separately analyze the progressivity/regressivity along income and mortality dimensions. We use our framework to characterize conditions under which a regressive redistribution along mortality dimension, i.e., from the short-lived to the long-lived, is not optimal. To develop our approach, we need to take a stand on the following three points: (i) how far we are willing to deviate from the standard utilitarian approach, (ii) why longevity matters for the welfare assessment, (iii) how we assess the mortality-related redistribution in isolation from income redistribution.

In our approach to social welfare, we do not specifically incorporate a bias towards the short-lived by, for example, increasing their pareto weight. Instead, our sole deviation from the standard utilitarian theory is to introduce aversion to inequality in lifetime utilities. This is done by assuming social welfare is a concave function of individual welfare, while a linear function is assumed in the standard utilitarian approach (see the discussion in Atkinson and Stiglitz, 1970). Since individual welfare is a concave function of consumption, both welfare criteria imply aversion to consumption inequality. However, the aversion to inequality in lifetime utilities is present only when individual welfare is aggregated non-linearly.

In the environment we consider, longevity matters for welfare because life is valuable. We assume that being alive brings additional non-pecuniary benefits, following a long tradition in the value of life literature (e.g., Hall and Jones, 2007; Rosen, 1988). In this approach, the long-lived have an utility advantage over the short-lived even when their consumption is the same.

Differentiating between income- and mortality-related redistribution is an important part of our analysis. To conceptualize ideas, we consider a framework where individuals differ in initial wealth/endowment and mortality. We then examine an average tax imposed by a particular redistributive scheme. If, everything else equal, the average tax is higher (lower) for the short-lived compared to the long-lived, we refer to this as mortality regressivity (progressivity). This conceptual framework allows us to examine the redistribution along mortality dimension for a given degree of income redistribution.

Using our theoretical framework, we derive several interesting results. We start by considering a simple case when people differ in their mortality but not in endowments. We show that in this environment, there is a tension between the aversion to two types of inequality: in consumption and in individual welfare. When aversion to consumption inequality exceeds aversion to inequality in individual welfare, mortality regressivity is always optimal. Moreover, when the aversion to these two types of inequality is the same, mortality regressivity is still optimal. In order for the redistribution from the long-lived to the short-lived to be optimal, aversion to lifetime inequality must be *stronger* than aversion to consumption inequality. Importantly, the push towards mortality progressivity is stronger when life is more valuable. This is because in this case, the variation in individual welfare due to lifespan variation is higher, calling for larger compensation to the short-lived.

We next examine a more general framework where there is heterogeneity in both endowments and mortality. We show that the optimality of mortality regressivity in this environment is determined by the same factors as described above as long as social planner can freely redistribute endowments. We then consider the case when endowments cannot be directly redistributed, and show that mortality progressivity can be optimal even if there is no aversion to inequality in lifetime utilities. This result arises when there is a positive correlation between life expectancy and income. Moreover, the stronger is the income-mortality correlation, the more mortality-progressive is the optimal allocation. This is because it is optimal to redistribute towards low-income people, but since this option is not available, the next best thing is to redistribute towards people with high mortality as they are more likely to be poor. Put differently, mortality progressivity becomes optimal as a substitute for income progressivity which cannot be directly achieved.

This conceptual framework easily lends itself to the study of pension systems' design, and this represents the next step of our analysis. To do this, we modify our approach by introducing working and retirement stages of the life-cycle. People pay contributions during their working life that depend on their labor productivity, and receive pension benefits when retired. The lifelong contributions to the pension system can be thought of as one's endowment, and pension benefits represent the annuitized value of this endowment.

We show that in absence of concern for lifetime inequality, optimal pensions are equalized across individuals, while with aversion to lifetime inequality, optimal pensions become positively linked to mortality and negatively to labor productivity. The mortality-pension link is used to compensate high-mortality people for their short life, and income-pension link - to compensate low-income people for their low consumption during working life.

In the final part of our analysis, we provide a quantitative illustration using a life-cycle model where people differ in their labor income and mortality. We estimate labor income and survival probabilities from the Panel Study of Income Dynamics (PSID) and the Health and Retirement Study (HRS), respectively. We use our quantitative model for the comparative welfare analysis of different spending-neutral pension systems. These systems differ in their mortality progressivity but are restricted in their ability to directly redistribute income. We show that when we shut down the heterogeneity in labor income, the welfare gains from mortality progressivity are relatively low even when aversion to lifetime inequality is high. In contrast, with heterogeneity in labor income, the welfare gains of moving to more mortality-progressive pensions can exceed 2% of annual consumption. This happens because, given the strong estimated correlation between labor income and mortality, it is optimal to use mortality progressivity to increase consumption of low-income people when direct income redistribution is restricted.

Our results thus emphasize that there are two distinct reasons making mortality progressivity optimal. First is the desire to compensate high-mortality people for their short life. This effect is only present when life is valuable and there is strong aversion to inequality in lifetime utilities. Second reason is the desire to redistribute towards low-income people. This effect is present when income and mortality are negatively correlated and when there are limited instruments to directly redistribute income. Importantly, while in both cases, mortality progressivity produces welfare gains, the underlying cause differs. In the first case, optimality of mortality progressivity is driven by the concern for the short-lived, while in the second - by the concern for the poor.

Our conclusions have important implications for the comparative pension policy analysis in presence of unequal lifespans. A typical pension system usually involves regressive redistribution along mortality dimension and progressive redistribution along income dimension. We suggest a framework that can be used to analyze the two types of redistribution separately. This allows us to do a more detailed comparison of different policies, and to separate the sources of welfare changes between the increase in welfare of the poor and in that of the short-lived. We show that in this analysis, three modeling assumptions are of crucial importance: the nature of income-mortality correlation, the value of life, and the assumed degree of aversion to lifetime inequality.

The rest of this paper is organized as follows. Section 2 reviews the related literature. Section 3 introduces the model and derives the results for the case when there is only mortality heterogeneity. Section 4 studies the case with both endowment and mortality heterogeneity, and Section 5 extends our framework to analyze different pension designs. Section 6 describes the quantitative illustration, and Section 7 concludes.

2 Related literature

Our paper belongs to several strands of literature. First are theoretical studies that investigate optimal pension/tax systems in presence of differential mortality. When evaluating the effects of different reforms, two approaches are typically used. The first approach is to focus on a set of Pareto improving reforms, i.e., the reforms that do not make any individual worse off compared to the status-quo (Adema et al., 2016; Hosseini and Shourideh, 2019). The second approach is to use ex-ante welfare comparison, commonly relying on the standard utilitarian welfare theory. This approach favors the redistribution from the short-lived to the long-lived, and this theoretical result was labeled by Leroux and Ponthiere (2013) "the paradox of the double penalization of the short-lived". The short-lived are penalized by nature by being endowed with low life expectancy and, in addition, they are penalized by lower lifetime consumption. One important implication of this result is the optimality of pension system with mortality-independent benefits, i.e., pooled annuitization (Cremer et al., 2010).

To avoid the double penalization paradox when considering pension arrangements, several approaches were used. Leroux and Ponthiere (2013) suggest imputing the consumption equivalence of longer life and include it as a part of social planner problem, while Pestieau and Racionero (2016) incorporate higher social weight on the utility of the short-lived. Bommier et al. (2011a, 2011b) use a standard utilitarian welfare function but allow individuals to have temporal risk aversion, i.e., aversion with respect to life duration. We contribute to this line of research by characterizing a set of conditions when mortality progressivity optimally arises in a relatively standard setting, i.e., we only deviate from the standard utilitarianism by introducing aversion to inequality in individual welfare. Importantly, we illustrate the importance of value of life, the nature of income-mortality correlation, and available income-redistributive tools for this analysis.

The second strand of literature we belong to investigates the normative and positive aspects of pension systems using a quantitative framework, an active line of research starting from Auerbach and Kotlikoff (1987). Many of the earlier studies abstract from differential mortality by assuming that people have fixed identical lifespans (Huggett and Parra, 2010; Ndiaye, 2020) or that survival probability does not differ across agents once age is controlled for (Conesa and Krueger, 1999; Kitao, 2014; Nishiyama and Smetters, 2007). With more detailed empirical documentation of the large inequality in mortality (Chetty et al., 2016), a growing number of studies start incorporating differential mortality in their analysis of pension systems. An important application of the quantitative models with differential mortality is the study of welfare consequences of various pension reforms such as reorganization of benefits and/or financing approaches (see an extensive review in Jones and Li, 2022).

This development raises an important question of how to model differential mortality and its relationship with other economic variables, and three approaches have been considered. The first and second approaches make opposite assumptions about the correlation of socio-economic variable and mortality: there either a one-to-one link between mortality and income (Bagchi, 2019; Hosseini and Shourideh, 2019; Sanchez-Romero, 2019; Sheshinski and Caliendo, 2021) or there is no association between the two (Bagchi and Jung, 2020; Imrohoroglu and Kitao, 2012). The third approach allows for a richer modeling of the interaction between mortality and economic variables by introducing a third stochastic variable, health, which evolution can be correlated with socio-economic variables and which affects life expectancy. At the same time, socio-economic variables such as education may also directly affect survival (Jones and Li, 2022; Laun et al., 2019) or the interaction between survival and economics variables can be entirely intermediated by health (Pashchenko and Porapakkarm, 2021). In the third approach, people with higher income are more likely to be healthy and thus are more likely to live longer, but the correlation between income and mortality is not perfect.

We contribute to this line of research by suggesting a framework to better understand the distinct role of redistribution along income and mortality dimensions in the comparative analysis of pension policies. Importantly, we show that how the relationship between income and mortality is modeled is very important for this type of welfare assessments.

The third strand of literature we relate to studies the issues related to the value of life. The constraint that continuation utility of being alive exceeds that of being dead is not commonly enforced in studies with individual optimization problems since, in most cases, it does not affect results. However, in certain applications this constraint plays a crucial role. Among them are valuation of changes in life expectancy or health (De Nardi et al., 2022; Hall and Jones, 2007; Murphy and Topel, 2006) or saving and portfolio choice with non-additive preferences (Pashchenko and Porapakkarm, 2022). We contribute to this literature by showing that the value of life can play an important role in welfare evaluations of pension reforms. This happens when these evaluations are done using a more general social welfare criteria that take non-pecuniary factors into account.

3 Only mortality heterogeneity

We start by constructing a framework where individuals only differ in their mortality. We consider the case with heterogeneous endowments in Section 4.

3.1 Environment

Consider a representative cohort whose initial mass is one. Individuals enter the model at age $t = 0$ with the same endowment A and receive no additional income over their lifetime. Individuals live to the maximum age of T. We denote the probability to survive from age t to $t + 1$ as θ_i. We assume θ_i does not vary with age but differs across individuals with $\theta_i \sim G(\theta_i)$ on the interval $[\theta_{min}, \theta_{max}]$, and the average survival is defined as

$$\overline{\theta} = \int_{\theta_{min}}^{\theta_{max}} \theta_i \, dG\left(\theta_i\right)$$

Individuals discount the future at the rate β. Resources can be transferred intertemporally at the rate r. We assume $\beta(1 + r) = 1$.

We consider two versions of the model. First is the laissez-faire situation when each individual, on entering the model at $t = 0$, converts his endowment A into an annuity based on an actuarially fair price.[1] We assume no consumption takes place at the starting period $t = 0$, and between $t = 1$ and $t = T$, consumption is equal to income coming from the annuitized endowment. In the second version of the model, consumption each period (for $t > 0$) is decided upon by social planner who optimizes ex-ante welfare subject to the aggregate resource constraint.

We denote the actuarially fair price of a unit of annuity income for an individual i as q_i, where:

$$q_i = \sum_{t=1}^{T} \left(\frac{\theta_i}{1 + r} \right)^t$$

Note that $q_i \in [q_{min}, q_{max}]$, where $q_{min} = q(\theta_{min})$ and $q_{max} = q(\theta_{max})$. We denote the average

[1] The one-time annuitization in the first period is optimal when there is no uncertainty except that in survival and annuities are actuarially fair. See Pashchenko (2013) for a formal proof.

annuity price as $\overline{q}$, where

$$\overline{q} = \int_{q_{min}}^{q_{max}} q_i \, dG\left(\theta_i\right).$$

It is worth noting that there is a one-to-one relationship between annuity price and individual survival probability, so in the subsequent analysis, we will consider individuals along the dimension of θ_i and q_i interchangeably.

Individuals derive utility from consumption c_{it} based on utility function $u(c_{it})$, where $u(c_{it})$ is strictly increasing, weakly concave and twice continuously differentiable. In addition, individuals derive non-pecuniary benefits from the fact that they are alive, which are captured by a positive constant b. Thus, total utility per period of an individual i, v_{it}, can be represented as follows:

$$v_{it} = u(c_{it}) + b$$

The utility in the state of death is normalized to zero.[2] Thus life is valued more than death when $v_{it} > 0$. The ex-ante lifetime utility of an individual i (or lifetime utility from the perspective of age $t = 0$) is

$$V_i = \sum_{t=1}^{T} (\beta\, \theta_i)^t \left(u(c_{it}) + b\right) = \sum_{t=1}^{T} (\beta\, \theta_i)^t \, v_{it} \tag{1}$$

Given that every agent in the economy gets endowment A, and the total initial mass of people in a representative cohort is one, the aggregate resource constraint for this economy can be represented as follows (to make notations less cluttered, we drop the limits of integration):

$$\int \sum_{t=1}^{T} \left(\frac{\theta_i}{1+r}\right)^t c_{it} \, dG\left(\theta_i\right) = A \tag{2}$$

In what follows, we will focus on age-invariant consumption allocations such that $c_{it} = c_i \;\; \forall \; t$, i.e., each individual's consumption is constant over time. This is the case in the laissez-faire economy, and we will later confirm that age-independent consumption is also optimal.

We next introduce two definitions.

Definition 1 We call the consumption allocation $\{c_i\}$ *feasible* if it satisfies the aggregate resource constraint in Eq (2).

A convenient way to describe the relationship between consumption and mortality is by using the concept of elasticity. We introduce the relevant elasticity in the next definition.

[2] This approach is common in the value of life literature, see, for example, Hall and Jones (2007) and Murphy and Topel (2006). An alternative approach is to re-normalize disutility from being dead instead of assuming extra utility from being alive. Rosen (1988) shows that these two approaches are equivalent.

Definition 2 Consider a feasible consumption allocation $\{c_i\}$. The elasticity of consumption c_i to actuarially fair annuity price q_i is defined as follows:

$$\varepsilon_{q_i}^{c} \;=\; \frac{dc_i}{dq_i} \cdot \frac{q_i}{c_i}$$

Note that $\varepsilon_{q_i}^{c}$ shows how consumption changes with individual's survival probability (or individual's annuity price).

3.2 Defining mortality regressivity

Intuitively, mortality regressivity implies that there is regressive redistribution along the dimension of mortality, i.e., form the short-lived to the long-lived. To understand whether a particular allocation is regressive or progressive, we first need to establish a reference point. We use as a reference point a laissez-faire allocation, when each individual annuitizes his endowment at the actuarially-fair price, and thus there is no redistribution across mortality types. To emphasize the absence of redistribution, we will refer to this as neutral allocation.

Definition 3 A feasible consumption allocation $\left\{c_i^N\right\}$ is *neutral* if

$$c_i^N = \frac{A}{q_i} \quad \forall \; i$$

It is worth noting that this corresponds to the situation when the elasticity of consumption to mortality is equal to -1 , i.e. , $\varepsilon_i^{cq} = -1 \quad \forall \; i$.

Consider next some feasible allocation $\{c_i\}$. To understand whether it is mortality-regressive or progressive, we compare it with the neutral (or laissez-faire) allocation $\{c_i^N\}$. We can think of $\{c_i^N\}$ as the allocation before the redistribution takes place, and $\{c_i\}$ as that after the redistribution. We can think of the difference $c_i^N - c_i$ as a tax (possibly negative), and we define the average tax AT_i as follows:

$$AT_i = \frac{c_i^N - c_i}{c_i^N} \;=\; 1 - \frac{c_i}{c_i^N}$$

We define mortality regressivity/progressivity based on how the average tax changes with mortality.

Definition 4 A feasible consumption allocation $\{c_i\}$ is *mortality-regressive* (*mortality-*

progressive) if the average tax decreases (increases) with life expectancy or q_i:

$$\frac{\partial AT_i}{\partial q_i} < (>)\ 0.$$

We next are going to formulate the proposition linking mortality regressivity and the concept of elasticity introduced in Definition 2.

Proposition 1 Consider a feasible allocation $\{c_i\}$. If this allocation is mortality-regressive (mortality-progressive) then the following is true:

$$\varepsilon^c_{q_i} > (<) -1 \quad \forall \ \ i$$

Proof We will do the proof for the mortality-regressive case. That for the mortality-progressive case is analogous. Using the definition of the neutral allocation $\{c_i^N\}$, we can rewrite the average tax as follows:

$$AT_i = 1 - \frac{c_i q_i}{A}$$

The derivative of the average tax with respect to q_i is

$$\frac{\partial AT_i}{\partial q_i} = -\frac{1}{A}\frac{\partial (c_i q_i)}{\partial q_i} \ = \ -\frac{c_i}{A}\,(\varepsilon^c_{q_i} + 1),$$

where the last equality follows from the definition of elasticity. Since in the mortality-regressive case $\frac{\partial AT_i}{\partial q_i} < 0$, it follows that $\varepsilon_i^{cq} > -1$. This finishes the proof of the proposition.

One important example to consider is the feasible allocation that is the same for all agents, $\overline{c} = \frac{A}{\overline{q}}$. This allocation is mortality-regressive as the average tax is decreasing in q_i:

$$AT_i = 1 - \frac{q_i}{\overline{q}} \tag{3}$$

Note that the elasticity is equal to zero, $\varepsilon^c_{q_i} = 0$.

3.3 Optimality of mortality regressivity

To understand under what conditions mortality progressivity can optimally arise, in this section, we consider the social planner problem for this economy. In our formulation of this problem, we deviate from the standard utilitarian approach in that we allow for the aversion to inequality in lifetime utilities. In the standard utilitarian approach, this inequality does not matter because the total welfare is a linear sum of individual lifetime utilities V_i, and thus V_i of different individuals are perfect substitutes. In our case, total welfare is a sum of a concave function of V_i, which relaxes the assumption of perfect substitution (see Atkinson and Stiglitz, 1970).

Social planner maximizes welfare of a representative cohort subject to the aggregate resource constraint:

$$\max_{\{c_{it}\}} \quad \int \Psi(V_i)\, dG\,(\theta_i) \tag{4}$$

$$\text{s.t.} \qquad \int \sum_{t=1}^{T} \left(\frac{\theta_i}{1+r} \right)^t c_{it}\, dG\,(\theta_i) = A \tag{5}$$

Here $\Psi(\cdot)$ represents planner's attitude towards inequality in lifetime utilities V_i , where V_i is defined in Eq (1). Note that in a standard utilitarian welfare case, $\Psi(\bullet)$ is linear. We assume that $\Psi(\bullet)$ is strictly increasing, weakly concave and twice continuously differentiable. Eq (5) is the aggregate resource constraint.

Denoting the Lagrange multiplier on the resource constraint as λ, we can write the first-order conditions as follows:

$$\frac{\partial\, \Psi(V_i)}{\partial\, V_i}\, \beta^t\, \theta_i^t\, \frac{\partial\, u(c_{it})}{\partial\, c_{it}} = \left(\frac{\theta_i}{1+r} \right)^t \lambda$$

Given that $\beta(1+r) = 1$, this can be simplified as follows:

$$\frac{\partial\, \Psi(V_i)}{\partial\, V_i} \cdot \frac{\partial\, u(c_{it})}{\partial\, c_{it}} = \lambda \qquad \forall \;\; i\,, t \tag{6}$$

Two important properties of the optimal allocation follow from Eq (6):

1. $c_{it} = c_i \quad \forall \;\; i$, i.e., it is optimal to give every individual a constant consumption stream.

2. $\dfrac{dc_i}{dq_i} \leqslant 0$, i.e., optimal consumption is non-increasing in longevity. This follows because for any two individuals i and j, given weak concavity of $\Psi(\cdot)$, Eq (6) implies that if $c_i \geq c_j$ then $V_i \leq V_j$. Since for a given level of consumption, lifetime utility is increasing in q_i ($V_i = q_i(u(c_i) + b)$), this can only be true when $\dfrac{dc_i}{dq_i} \leq 0$. This implies that $\varepsilon^c_{q_i} \leq 0 \quad \forall \;\; i$. In the subsequent analysis, we focus on $\mid \varepsilon^c_{q_i} \mid$.

When consumption is age-invariant, the per-period utility v_i is also the same at each age. We can use this fact together with the definition of the annuity price and the assumption

$\beta(1+r)=1$, to rewrite the lifetime utility V_i in Eq (1) as follows:

$$V_i = v_i q_i$$

Before proceeding, we are going to introduce the following notations:

$$\frac{\partial\, \Psi(V)}{\partial\, V} \equiv \Psi_V\,, \qquad \frac{\partial\, u(c)}{\partial\, c} \equiv u_c$$

$$\frac{\partial^2\, \Psi(V)}{\partial\, V^2} \equiv \Psi_{VV}\,, \qquad \frac{\partial^2\, u(c)}{\partial\, c^2} \equiv u_{cc}$$

Below, we summarize the key assumptions we use for our analysis.

Assumption 1. Both $u(\cdot)$ and $\Psi(\cdot)$ are strictly increasing, weakly concave and twice continuously differentiable

Assumption 2. $\beta(1+r)=1$

Assumption 3. Life is valuable for all individuals: per-period utility of being alive v_i exceeds zero (utility at death), $v_i = u(c_i)+b>0$ for all i.

The following proposition formalizes a criteria to determine whether an optimal allocation is mortality-regressive or -progressive.

Proposition 2 Consider the consumption allocation $\{c_i\}$ that represents the solution to the social planner problem described in Eqs (4)-(5). Under Assumptions 1-3, whether this allocation is mortality-regressive/progressive can be determined as follows:

1. If $\Psi(\cdot)$ is linear, $\{c_i\}$ is the same for all individuals, $c_i=\overline{c}$ for all i, and is mortality-regressive.

2. If $\Psi(\cdot)$ is strictly concave, $\{c_i\}$ is mortality-regressive (mortality-progressive) if

$$\frac{u_{c_i} c_i}{v_i} + \frac{R_{ui}}{R_{\Psi i}} > (<)\ 1 \qquad \forall \ \ i \tag{7}$$

where $R_{ui} = -\dfrac{u_{cc_i}}{u_{c_i}}\, c_i$ and $R_{\Psi i} = -\dfrac{\Psi_{VV_i}}{\Psi_{V_i}}\, V_i$ are the coefficients of relative risk aversion of functions $u(\cdot)$ and $\Psi(\cdot)$, respectively.

Proof In the subsequent discussion, we will drop an individual's subscript i except in cases when we want to emphasize the difference between individuals.

Using the simplified notation introduced above, we can rewrite the FOC in Eq (6) as follows:

$$\Psi_V\, u_c = \lambda \tag{8}$$

Taking the full differential of this equation around the optimal allocation, we get:

$$\Psi_{VV}\, v\, u_c\, dq + (\Psi_{VV}\, u_c^2\, q + \Psi_V\, u_{cc})\, dc = 0$$

Hence

$$\frac{dc}{dq} = -\frac{\Psi_{VV}\, v\, u_c}{\Psi_{VV}\, u_c^2\, q + \Psi_V\, u_{cc}} = -v\, \frac{\dfrac{\Psi_{VV}}{\Psi_V}\, V}{\dfrac{\Psi_{VV}}{\Psi_V}\, V\, u_c\, q + \dfrac{u_{cc}}{u_c} c \dfrac{V}{c}}$$

Using the definitions of the coefficients of the relative risk aversion and the fact that $V = vq$, we can transform this as follows:

$$\frac{dc}{dq} = -\frac{v}{q}\, \frac{R_\Psi}{R_\Psi\, u_c + R_u\, \dfrac{v}{c}}$$

Rewriting this expression in terms of elasticity, we have:

$$|\, \varepsilon^c_{q_i} \,| = \frac{R_\Psi}{R_\Psi\, \dfrac{u_c\, c}{v} + R_u}$$

It directly follows that when $R_\Psi = 0$ ($\Psi(\cdot)$ is linear), $|\, \varepsilon^c_{q_i} \,| = 0$, and optimal consumption is the same for all agents. This proves part 1 of the proposition. When $R_\Psi \neq 0$, we can rewrite the elasticity as follows:

$$|\, \varepsilon^c_{q_i} \,| = \frac{1}{\dfrac{u_c\, c}{v} + \dfrac{R_u}{R_\Psi}} \tag{9}$$

The consumption allocation is mortality-regressive (-progressive) when $|\, \varepsilon^c_{q_i} \,| < (>)\, 1$ by Proposition 1. This finishes the proof of the proposition.

Intuition To better understand the intuition, consider the key expression of Proposition 2 in Eq (7), which contains two terms. The first term represents the elasticity of per-period utility v to consumption, $\frac{u_c\, c}{v} = \frac{dv}{dc} \cdot \frac{c}{v}$. It determines how sensitive is per-period utility to the marginal change in consumption. This sensitivity depends on whether consumption is all that matters, or whether there are also non-pecuniary factors that affect utility.

The second term $\frac{R_u}{R_\Psi}$ measures the relative concavity of the functions $u(\cdot)$ and $\Psi(\cdot)$, and thus determines whether social planner is more concerned about inequality in lifetime utilities or in consumption.

When non-pecuniary benefits of being alive (b) are high, the first term is relatively low, and if it is combined with strong aversion to lifetime inequality (the second term is low), this represents a push for mortality progressivity. This happens because it is optimal to compensate the short-lived for their low lifetime utility with higher consumption. In the opposite case, when consumption is the dominant component in utility, and the planner is less concerned about inequality in lifetime welfare, it is more likely that the optimal allocation is mortality-regressive.

Additional results We next are going to show that when we introduce an additional assumption that both $u(\cdot)$ and $\Psi(\cdot)$ are constant relative risk aversion functions, Proposition 2 gives raise to two corollaries.

Assumption 4. The functions $u(\cdot)$ and $\Psi(\cdot)$ have constant relative risk aversion, i.e., $R_{u_i} = R_u$ and $R_{\Psi_i} = R_\Psi$ for all i.

Corollary 1 Suppose Assumptions 1-4 hold. Then the optimal consumption allocation is always mortality-regressive if $R_u \geq R_\Psi$.

Proof: When $R_\Psi = 0$, this follows from part 1 of Proposition 2. When $R_\Psi > 0$, this follows from Eq (7) and the fact that $v > 0$ and $u_c > 0$.

An important implication of Corollary 1 is that even if $u(\cdot)$ and $\Psi(\cdot)$ have the same degree of concavity ($R_u = R_\Psi$), mortality regressivity is still optimal. In other words, in order for mortality progresivity to be optimal, it is not enough to introduce aversion to inequality in lifetime utilities, it is essential that the concern for lifetime inequality is *stronger* than the concern for consumption inequality, i.e., $\Psi(\cdot)$ should be more concave than $u(\cdot)$.

Corollary 2 Suppose Assumptions 1-4 hold. In addition, assume that $u(\cdot)$ and $\Psi(\cdot)$ are strictly concave. Then mortality progressivity of the optimal consumption allocation $\{c_i\}$ is more likely to arise when per-period utility of being alive b is larger.

Proof: Consider the FOC in Eq (8) for two agents, i and j, with $q_i < q_j$. We can combine the FOCs as follows:

$$\frac{\Psi_{V_i}}{\Psi_{V_j}} \frac{u_{c_i}}{u_{c_j}} = \left(\frac{q_i}{q_j}\right)^{-R_\Psi} \left(\frac{u(c_i)+b}{u(c_j)+b}\right)^{-R_\Psi} \frac{u_{c_i}}{u_{c_j}} = 1,$$

Here we used the fact that $V_i = q(u(c_i)+b)$ and the assumption that $\Psi(\cdot)$ is the CRRA function with risk aversion R_ψ. Consider a perturbation of this equation around the optimal allocation when we change b. Since $q_i < q_j$, we have $c_i > c_j$ and the ratio $\left(\dfrac{u(c_i)+b}{u(c_j)+b}\right)^{-R_\Psi}$

increases in response to the marginal change in b (holding allocation fixed at the optimal level). Hence, we now have $\Psi_{V_i} u_{c_i} > \Psi_{V_j} u_{c_j}$, so it is optimal to rearrange the allocation in a way that c_i increases and c_j decreases. In other words, resources are reallocated from low-mortality to high-mortality agents, which represents a move towards mortality progressivity.

Intuitively, Corollary 2 implies that when the value of being alive increases, unequal lifespans create larger dispersion in lifetime utilities. Social planner who is averse to this type of inequality reduces it by compensating high-mortality people for their short life with higher consumption, thus moving towards mortality progressivity.

3.4 The aversion to inequality in consumption versus in lifetime utilities

The key implication of Proposition 2 is that the relative concavity of functions $u(\cdot)$ and $\Psi(\cdot)$ is important for the optimality of mortality progressivity/regressivity. To better illustrate the intuition of this result, we consider two opposite cases in terms of whether $u(\cdot)$ or $\Psi(\cdot)$ is more concave. We are going to assume that one of these functions is linear and the other is strictly concave, in which case the optimal consumption allocations can be derived analytically.

Only consumption inequality matters: $\Psi(\cdot)$ is linear and $u(\cdot)$ is concave In this case, social planner is concerned only about consumption inequality and is neutral to inequality in lifetime utilities. It is optimal to equalize consumption, $c_i = \overline{c}$. The optimal consumption $\overline{c}$ can be found from the aggregate resource constraint:

$$\int q_i \, \overline{c} \, dG(\theta_i) = \overline{c} \int q_i \, dG(\theta_i) = \overline{c} \, \overline{q} = A,$$

implying $\overline{c} = \dfrac{A}{\overline{q}}$. This means that social planner makes individuals annuitize their endowment at the same pooled price. We will refer to this consumption allocation as the *utmost mortality-regressive.*

Note that the resulting lifetime utility is linearly increasing in q_i:

$$V_i = q_i \left(u(\overline{c}) + b \right)$$

The average tax for this case is given in Eq (3). The elasticity of consumption to annuity price is zero, $|\, \varepsilon^c_{q_i} \,| = 0$ for $\forall \;\; i$.

Only inequality in lifetime utility matters: $\Psi(\cdot)$ is concave and $u(\cdot)$ is linear In this case, social planner aims to equalize lifetime utilities, while being indifferent to inequality in consumption. Assuming $u(c) = c$, we can solve for the optimal allocation as follows. From the first-order condition in Eq (8) we have:

$$\Psi_{V_i} = \lambda \quad \forall \;\; i,$$

implying $V_i = \overline{V} \;\; \forall \;\; i$.

Since $q_i\,(u(\,c_i) + b) = q_i\,c_i + q_i\,b = \overline{V}$, we can transform the aggregate resource constraint:

$$\int q_i\,c_i\,dG\,(\theta_i) = \int \big(\overline{V} - q_i\,b\big) dG\,(\theta_i) = \overline{V} - \overline{q}\,b = A$$

Thus $\overline{V} = A + \overline{q}\,b$ and

$$c_i \;=\; \underbrace{\frac{A}{q_i}}_{\text{neutral}} \;+\; \underbrace{b\,\frac{\overline{q} - q_i}{q_i}}_{\text{compensation for short life}} \tag{10}$$

When $b = 0$, this allocation is neutral (or laissez-faire). When $b > 0$, people with low life expectancy ($q_i < \overline{q}$) receive transfers financed by reduction in consumption of people with high life expectancy ($q_i > \overline{q}$). We will refer to this consumption allocation as the *utmost mortality-progressive*.

We can write the average tax as:

$$AT_i = 1 - \frac{A + b(\overline{q} - q_i)}{A} \;=\; b\,\frac{q_i - \overline{q}}{A},$$

and it is increasing in q_i implying mortality progressivity.

The absolute value of the elasticity of consumption to annuity price is greater than one as long as $b > 0$:

$$|\,\varepsilon^c_{q_i}\,| = \frac{A + b\,\overline{q}}{A + b\,(\overline{q} - q_i)} > 1 \quad \forall \;\; i$$

Numerical illustration The important difference between the utmost mortality-regressive ($\Psi(\cdot)$ is linear) and the utmost mortality-progressive ($u(\cdot)$ is linear) allocations is how consumption changes with longevity type. In the utmost mortality-regressive case, $\dfrac{dc_i}{dq_i} = 0$, while $\dfrac{dc_i}{dq_i} < 0$ for the utmost mortality-progressive case. Overall, the mortality progressivity is determined by the speed at which consumption declines with longevity.

This can be best illustrated with the following numerical example. We consider the

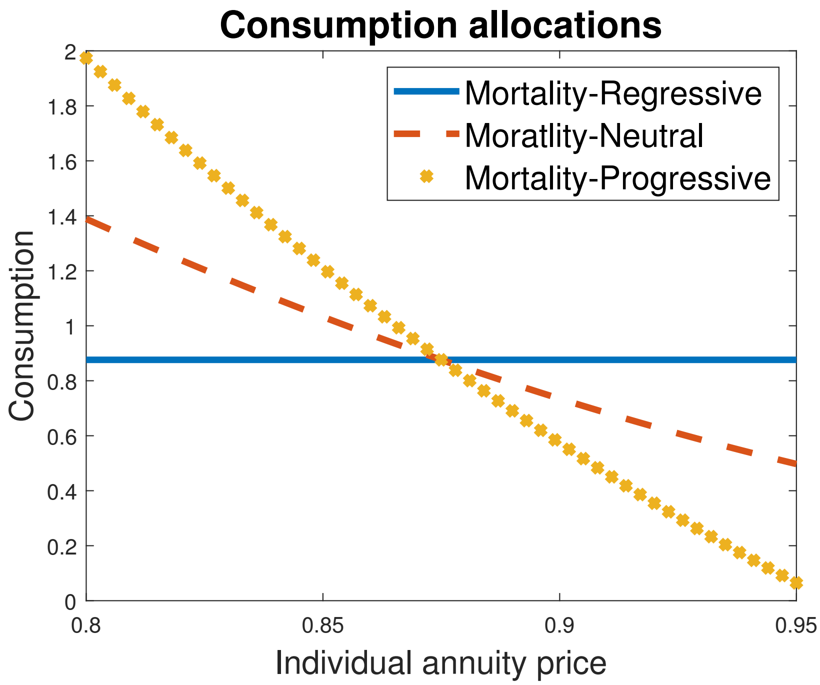

Figure 1: Consumption allocations differing in their mortality progressivity.

example where $A = 5$, $r = 2\%$, $b = 1$, maximum lifespan $T = 20$ and the survival probability being uniformly distributed in the interval $[0.8, 0.95]$. In Figure 1 we plot three allocations: mortality-neutral, the utmost mortality-regressive, and the utmost mortality-progressive. The figure shows that consumption declines much quicker as longevity increases in case of the mortality-progressive case compared to the mortality-neutral case.

4 Mortality and endowment heterogeneity

In this section, we are going to relax the assumption of equal endowments. Instead, we examine the situation when individuals differ in both mortality and endowments, with the two possibly being correlated.

4.1 Environment

We denote an endowment of an individual i as a_i with $a_i \sim F(a_i)$ on the interval $[a_{min}, a_{max}]$, and with average endowment $\int a_i \, dF(a_i) = \overline{A}$. We denote the joint distribution of a_i and q_i as $H(a, q)$, and the endowment-weighted average annuity price as $\overline{q}^a$:

$$\overline{q}^a = \frac{\int\limits_q \int\limits_a q_i a_i \; H(a_i, q_i) \, da_i \, dq_i}{\overline{A}},$$

As before, we focus on age-invariant consumption allocations (c_i is the same for all t). Consumption does not vary with age in the laissez-faire case, and it is also the property of the optimal allocation, which can be shown in the same way as in Section 3. We can thus write the aggregate resource constraint as

$$\int_q \int_a \sum_{t=1}^{T} \left(\frac{\theta_i}{1+r} \right)^t c_i \, H(a_i, q_i) \, da_i \, dq_i = \overline{A} \tag{11}$$

We next need to define the reference allocation that can be used to determine mortality regressivity/progressivity. For this, we modify our definition of the neutral or laissez-faire consumption allocation (Definition 3 from Section 3) as follows.

Definition 3.1 A feasible consumption allocation $\{c_i^N\}$ is *neutral* if

$$c_i^N = \frac{a_i}{q_i} \quad \forall \ i$$

This corresponds to the situation when each individual converts his endowment a_i into an annuity based on his actuarially-fair annuity price q_i.

There may be redistribution along both mortality and endowment dimensions, and we wish to disentangle the two, i.e., to analyze redistribution along the mortality dimension for a given endowment distribution. For this, we are going to modify our Definition 4 from Section 3.

As before, we are going to start by comparing a feasible allocation $\{c_i\}$ with the neutral allocation $\{c_i^N\}$, constructing a tax $c_i^N - c_i$, with the average tax AT_i taking the same form as before:

$$AT_i = 1 - \frac{c_i}{c_i^N}$$

Our modified definition of mortality progressivity/regressivity is stated as follows.

Definition 4.1 A feasible consumption allocation $\{c_i\}$ is mortality-regressive/-neutral/-progressive when the average tax decreases/does not change/increases with life expectancy or q_i, *given the endowment* a_i:

$$\left. \frac{\partial AT_i}{\partial q_i} \right|_{a_i} < (=) > \ 0.$$

It is worth noting that this definition also includes a concept of *mortality-neutrality* differing from that of neutrality more generally. When endowments do not differ across agents, these two concepts are the same. With heterogeneous endowments, some feasible allocations can be not neutral (not laissez-faire), while still being mortality-neutral. One example of such allocation is the following: $c_i = \frac{\overline{A}}{q_i}$. The average tax in this case is $AT_i = 1 - \frac{\overline{A}}{a_i}$. This tax does not change with longevity or q_i for a given endowment a_i.

In order to understand whether a particular allocation is mortality-progressive/regressive, we can still apply Proposition 1 but in a slightly modified form.

Proposition 1.1 Consider a feasible allocation $\{c_i\}$. If this allocation is mortality-regressive (mortality-progressive) then the following is true:

$$\varepsilon^c_{q_i\,|\,a_i} > (<) -1 \quad \forall \;\; i,$$

where $\varepsilon^c_{q_i\,|\,a_i}$ is the partial elasticity of consumption to mortality, i.e., the elasticity for a given level of endowment a_i:

$$\varepsilon^c_{q_i\,|\,a_i} = \left.\frac{dc_i}{dq_i}\right|_{a_i} \cdot \frac{q_i}{c_i}.$$

Proof See Appendix A.

4.2 Optimality of mortality regressivity

The social planner's problem in the environment with both mortality and endowment heterogeneity can be formulated as follows:

$$\max_{\{c_i\}} \quad \int \Psi(V_i)\, H(a_i, q_i)\, da_i\, dq_i \tag{12}$$

subject to the aggregate resource constraint in Eq (11). The following proposition compares optimal consumption allocations in this case with the case of homogeneous endowments considered in Section 3.

Proposition 3 Consider the consumption allocation $\{c_i\}$ that represents the solution to the social planner problem described in Eqs (12) and (11). Suppose Assumptions 1-3 hold, and $\int a_i\, dF\,(a_i) = \overline{A} = A$, i.e., the total endowment in the economies with heterogeneous and homogeneous endowments are the same. Then the optimal consumption allocation in the two economies is the same.

Proof Since $\overline{A} = A$, the aggregate resource constraint is the same in the two economies. Denoting the Lagrange multiplier on the aggregate resource constraint as λ, we can write the FOC as follows:

$$\Psi_{V_i}\, u_{c_i} = \lambda \tag{13}$$

This is equivalent to the FOC in Eq (8) for the social planner problem in Eqs (4)-(5). The two social planner problems have the same solution, which finishes the proof of the proposition.

It is worth noting that the FOC in Eq (13) implies that consumption allocation $\{c_i\}$ does not depend on individual endowments a_i but can depend on mortality q_i. Proposition 3 shows that when social planner can freely redistribute endowments, the characterization of the optimal allocation does not change when we allow for heterogeneous endowments. This also means that the conditions for mortality regressivity are the same as described in Proposition 2.

We next examine the two extreme cases considered above, when either function $\Psi(\cdot)$ or $u(\cdot)$ is linear. This results in either the utmost mortality-regressive allocation (when only consumption inequality matters), or the utmost mortality-progressive allocation (when only lifetime inequality matters). Using the same steps as in Section 3.4, we can find the utmost mortality-regressive consumption allocation:

$$c_i = \frac{\overline{A}}{\overline{q}} \tag{14}$$

The utmost mortality-progressive allocation takes the following form:

$$c_i = \frac{\overline{A}}{q_i} + b\ \frac{\overline{q} - q_i}{q_i} \tag{15}$$

In both cases, individual endowments are pooled together and equally distributed across agents (everyone gets $\overline{A}$). The pooled endowment is then annuitized at the average annuity price in the utmost mortality-regressive case. In the utmost mortality-progressive case, pooled endowments are annuitized at the individual actuarially-fair prices with a compensation for short life added when $b > 0$.

4.3 Restricted social planner problem

We next consider a more interesting case when social planner cannot directly choose consumption allocations, but instead has an access to a limited set of instruments. It is worth start by noting that in the unrestricted social planner problem, optimal consumption allocations vary between the utmost mortality-regressive (Eq 14) and the utmost mortality-progressive (Eq 15) cases. We can approximate consumption allocations in between these two extreme cases with the following parametric form:

$$c_i = (1-\alpha_1)\,\frac{\overline{A}}{\overline{q}} + \alpha_1\,\frac{\overline{A}}{q_i} + \alpha_2\, b\,\frac{\overline{q}-q_i}{q_i}, \tag{16}$$

where $\alpha_1 \in [0,1]$ and

$$\begin{cases} \alpha_2 = 0 & ;\ \text{if } \alpha_1 < 1 \\ \alpha_2 \geq 0 & ;\ \text{if } \alpha_1 = 1 \end{cases}$$

When $\alpha_1 = 0$, the allocation is the utmost mortality-regressive. As α_1 increases, the allocation moves away from the utmost-regressive and towards the mortality-neutral case. Once $\alpha_1 = 1$, the allocation is mortality-neutral. Increasing α_2 above zero introduces mortality progressivity. The aggregate resource constraint is met for all described combinations of α_1 and α_2.

Importantly, in the unrestricted social planner problem, the correlation between endowment and mortality does not matter since individual endowments are pooled together. We next consider a situation when the ability of social planner to redistribute endowments is limited. Specifically, we assume that individual endowments a_i cannot be changed. This case is interesting because it allows us to focus on the redistribution along mortality dimension, and to understand the role of the mortality-endowment correlation.

We thus modify the consumption allocation rule in Eq (16) as follows:

$$c_i = (1-\alpha_1)\,\frac{a_i}{\overline{q}^a} + \alpha_1\,\frac{a_i}{q_i} + \alpha_2\, b\,\frac{\overline{q}-q_i}{q_i} \tag{17}$$

Compared to Eq (16), the first term in this equation is divided by $\overline{q}^a$ as opposed to $\overline{q}$ in order to meet the aggregate resource constraint. This way, varying α_1 and α_2 does not change the total spending on consumption allocations.

We can now define the constrained social planner problem: social planner maximizes social welfare by choosing α_1 and α_2:

$$\max_{\alpha_1,\alpha_2} \int_q \int_a \Psi(V_i)\ H(A_i, q_i)\, da_i\, dq_i \tag{18}$$

where $V_i = q_i(u(c_i) + b)$ and c_i is given in Eq (17).

Our goal is to understand how the mortality-endowment correlation affects the optimality of mortality progressivity. In this analysis, we focus on the case which is least favorable to mortality progressivity based on our analysis in the previous section: we assume that social planner is indifferent to inequality in lifetime utilities and has aversion to consumption inequality ($\Psi(\cdot)$ is linear and $u(\cdot)$ is strictly concave).

Our key results of this section are summarized in Proposition 4.

Proposition 4 Consider the constrained social planner problem described in Eq (18), and suppose Assumptions 1-4 hold. In addition, assume $R_\Psi = 0$ and $R_u > 1$. The optimal choice of α_1 and α_2 can be summarized as follows:

(1) If $cov(q, a) = 0$, then at the optimum $\alpha_1 = 0$,

(2) If $cov(q, a) > 0$, then

(i) At the optimum $\alpha_1 > 0$,

(ii) If $cov(q, \frac{a}{q}) < 0$, then at the optimum $\alpha_2 = 0$,

(iii) If $cov(q, \frac{a}{q}) > 0$, then at the optimum $\alpha_2 > 0$.

Proof See Appendix B.

The intuition for the part (i) of Proposition 4 follows from our analysis in the previous section: when social planner's only concern is consumption inequality (as $\Psi(\cdot)$ is assumed to be linear), it is optimal to equalize consumption. The closest social planner can get to equalizing consumption when he cannot directly redistribute endowments is by annuitizing an endowment of each agent at the same pooled price.

Once we introduce the correlation between mortality and endowment, optimal allocations change. Specifically, when short-lived people tend to have lower endowments, there is a push for mortality progressivity, which is more pronounced when the mortality-endowment link is stronger, $cov(q, \frac{a}{q}) > 0$. We formally show that the latter condition is stronger than $cov(q, a) > 0$ in the Auxiliary proposition in Appendix B. Intuitively, the condition $cov(q, \frac{a}{q}) > 0$ implies that even though people with longer life expectancy face higher actuarially fair annuity prices q, their endowments a tend to be so high that they still can obtain higher annuity income, $\frac{a}{q}$.

It is important to point out that the push towards mortality-progressivity in this case arises not because social planner wants to equalize lifetime utility and to compensate high-mortality agents for their short life, but because mortality progressivity is used as a substitute for income progressivity in absence of direct instruments to redistribute endowments.

5 Mortality regressivity and pension design

In this section, we extend our theoretical framework to analyze pension design. We modify the setup of Sections 3 and 4 in two ways. First, instead of a representative cohort, we consider the overlapping generations model. Second, we introduce two stages of the life-cycle: working and retirement periods. During working period, each individual receives labor income and pays contributions to the pension system, during retirement period, he receives pension benefits. In this environment, we can think of total contributions to the system as one's endowment, and of pension benefits as the annuitized value of the endowment. People may differ both in endowments and mortality, and pension system may thus feature redistribution along both dimensions. We focus on a set of revenue-neutral policies, i.e., policies that are financed by the same tax revenue.

5.1 Environment

We consider an overlapping generations model where each individual lives for T periods: for the first R periods, an agent receives labor income ϵ_i, between periods $R+1$ and T, an agent is retired. The population grows at the rate n.

In our analysis, we maintain Assumptions 1-4 from Section 3. We also assume the economy is dynamically efficient, so that the population growth rate is equal to the interest rate, $n = r$. In addition, we assume agents have inelastic labor supply and cannot save. In our quantitative model in the next section, we relax the assumption of no savings.

Agents differ in mortality and labor productivity ϵ_i, with $\epsilon_i \sim F(\epsilon_i)$, and we denote the average productivity $\overline{\epsilon} = \int \epsilon \, dF(\epsilon)$. Agents survive to period R with probability one, and after that, the probability to survive from age t to $t+1$ is θ_i with $\theta_i \sim G(\theta_i)$ and $\overline{\theta} = \int \theta \, dG(\theta)$. Labor productivity and mortality can be correlated and we denote their joint distribution as $H(\epsilon, \theta)$.

The actuarially fair price q_i of a unit of lifelong annuity income for an individual i acquired before retirement (at age R) is:

$$q_i = \sum_{t=1}^{T-R} \left(\frac{\theta_i}{1+r} \right)^t,$$

with the average annuity price denoted as $\overline{q}$. As before, we use actuarially fair annuity price q_i and mortality θ_i interchangeably in the subsequent discussion.

During the working stage of the life-cycle, each agent pays proportional tax τ on his labor income. After retirement, each agent receives benefits ssb_i. Given our assumption that agents cannot save, each period's consumption is equal to income (either labor income or pension income), and the lifetime utility can be represented as follows:

$$
\begin{aligned}
V_i = \sum_{t=1}^{R} \ \beta^{t-1}(u(\epsilon_i(1-\tau))+b) \ + \ \beta^{R-1}\sum_{t=1}^{T-R} \ (\beta\theta)^t(u(ssb_i)+b) \\
= \sum_{t=1}^{R} \ \beta^{t-1}(u(\epsilon_i(1-\tau))+b) \ + \ \beta^{R-1}q_i \ (u(ssb_i)+b)
\end{aligned}
\tag{19}
$$

The last equality follows from the assumption $\beta(1+r) = 1$ (Assumption 2) and the definition of the annuity price. Denoting the lifetime utility during working period as $V_i^W = \sum_{t=1}^{R} \ \beta^{t-1}(u(\epsilon_i(1-\tau))+b)$, and during the retirement period as $V_i^R = q_i \ (u(ssb_i)+b)$, we can also write:

$$
V_i = V_i^W + \beta^{R-1}V_i^R
$$

5.2 Pension system

There are two ways to set up a pension system in this environment, either on a fully funded or on a pay-as-you-go basis. In the first case, an individual pays contributions to his pension account that is later annuitized. In the second case, working-age people finance pensions of retirees. In this section, we show that under the assumptions of dynamic efficiency and inelastic labor supply, the two systems are equivalent. This result is convenient because it allows us to think of pension contributions in both cases as individuals' notional balances or endowments.

In the fully-funded pension system, an individual i has a notional balance IC_i, which represents his lifelong contributions:

$$
IC_i = \tau\epsilon_i \ \sum_{t=1}^{R} \ (1+r)^{t-1}
$$

Upon retirement, the notional balance IC_i is converted into an annuity. Thus, for each individual, $ssb_i \, q_i = IC_i$. Since this is true for every individual, it is true for the whole cohort of pre-retirement age ($t = R$), and we can write the balance equation for the fully-funded pension system as follows:

$$
\int_q \int_\epsilon ssb \ q \ H(\epsilon, q) \, d\epsilon \, dq = \int_\epsilon IC dF(\epsilon)
\tag{20}
$$

In the pay-as-you-go system, individuals' contributions are used to finance pensions of the old. Denoting as N the initial size of the cohort who is currently of age T (the oldest cohort), we can write the pension system balance equation as follows:

$$N \int\limits_q \int\limits_\epsilon ssb \sum_{t=1}^{T-R} \theta^t (1+n)^{T-R-t} H(\epsilon, q)\, d\epsilon\, dq = N \int\limits_\epsilon \tau\epsilon \sum_{t=1}^{R} (1+n)^{T-t} dF(\epsilon)$$

Dividing both sides by $N(1+n)^{T-R}$, which is the size of the cohort who is about to retire, we have

$$\int\limits_q \int\limits_\epsilon ssb \sum_{t=1}^{T-R} \left(\frac{\theta}{1+n}\right)^t H(\epsilon, q)\, d\epsilon\, dq = \int\limits_\epsilon \tau\epsilon \sum_{t=1}^{R} (1+n)^{t-1} dF(\epsilon) \tag{21}$$

We can use our assumption of dynamic efficiency ($n = r$) and the definition of the actuarially-fair annuity price to show that the balance equations for the fully-funded and pay-as-you-go pension systems in Eqs (20) and (21) are equivalent. Moreover, we can think of the right-hand side of Eq (21) as the average of notional balances IC_i, since when $n = r$ the following is true:

$$IC_i = \tau\epsilon_i \sum_{t=1}^{R} (1+n)^{t-1}$$

In this stylized framework, we will treat IC_i as one's endowment, and pension benefits ssb_i as the annuitized value of this endowment.

It is worth noting that since we focus on a set of revenue-neutral policies, we fix the tax rate τ. Given the assumption of inelastic labor supply, this means the average contributions $\overline{IC} \equiv \int\limits_\epsilon ICdF(\epsilon)$ do not vary across policies.

5.3 Definitions

We next modify our key definitions from Section 4, starting from that of feasibility.

Definition 1.2 We call pension benefits $\{ssb_i\}$ *feasible* if they satisfy the pension system balance equation in Eqs (20) or (21).

Since in our framework, pension benefits can be thought of as annuitized value of one's endowment, this leads us to the following definition of neutral pension benefits.

Definition 3.2 Feasible pension benefits $\{ssb_i^N\}$ are *neutral* if

$$ssb_i^N = \frac{IC_i}{q_i} \quad \forall \; i$$

This corresponds to the situation when each individual converts his lifelong contributions to the pension system IC_i into an annuity based on his actuarially-fair annuity price q_i. Thus, neutral benefits involve no redistribution along neither mortality nor endowment dimensions.

A pension system may involve redistribution along both income and mortality dimensions, and we wish to analyze these two separately. For this, we are going to modify our definition of mortality regressivity (Definition 4.1 from Section 4) and add the definition of endowment regressivity/progressivity.

To do this, we start by comparing feasible pension benefits $\{ssb_i\}$ with neutral benefits $\{ssb_i^N\}$, constructing a tax $ssb_i^N - ssb_i$, with the average tax AT_i taking the following form:

$$AT_i = 1 - \frac{ssb_i}{ssb_i^N}$$

Our modified definition of mortality progressivity/regressivity is stated as follows.

Definition 4.2 Feasible pension benefits $\{ssb_i\}$ are mortality-regressive/-neutral/-progressive when the average tax decreases/does not change/increases with life expectancy or q_i, *given the endowment (lifetime pension contributions)* IC_i:

$$\left.\frac{\partial AT_i}{\partial q_i}\right|_{IC_i} < (=) > \ 0.$$

In a similar fashion, we can define the endowment regressivity/progressivity in our environment:

Definition 6 Feasible pension benefits $\{ssb_i\}$ are endowment-regressive/-neutral/-progressive when the average tax decreases/does not change/increases with endowment IC_i, *given the life expectancy or* q_i:

$$\left.\frac{\partial AT_i}{\partial IC_i}\right|_{q_i} < (=) > \ 0.$$

5.4 Optimal pension system

Consider the social planner problem in this environment. Since the tax rate τ and the size of the pension program are fixed, social planner only chooses how to set pension benefits subject to the pension system balance equation:

$$\max_{\{ssb_i\}} \int_q \int_\epsilon \Psi(V_i)\ H(\epsilon_i, q_i)\, d\epsilon_i\, dq_i \tag{22}$$

s.t.

$$\int_q \int_\epsilon ssb_i \; q_i \; H(\epsilon_i, q_i) \, d\epsilon_i dq_i = \overline{IC}, \tag{23}$$

where V_i is given in Eq (19).

Denoting the Lagrange multiplier on the constraint as $\beta^{R-1}\lambda$, we can write the FOC as follows:

$$\Psi_{V_i} \, u_{ssb_i} = \lambda, \tag{24}$$

where

$$\frac{\partial \, \Psi(V_i)}{\partial \, V_i} \equiv \Psi_{V_i} \,, \qquad \frac{\partial \, u(ssb_i)}{\partial \, ssb_i} \equiv u_{ssb_i}$$

The FOC in Eq (24) implies the following property of the optimal pension benefits. Since V_i is increasing in both q_i and ϵ_i, when $\Psi(\cdot)$ is strictly concave, ssb_i is decreasing in q_i (given ϵ_i) and in ϵ_i (given q_i). The later effect is new compared to the environment studied in Sections 3 and 4, where an individual's endowment does not affect optimal consumption. This effect arises because social planner can only change consumption after retirement, while concern for lifetime inequality also makes the utility over working period matter for optimal pensions.

We can now formulate the modified version of Proposition 2.

Proposition 2.1 Consider pension benefits $\{ssb_i\}$ that represent the solution to the social planner problem described in Eqs (22)-(23). Under Assumptions 1-4, whether these benefits are mortality- and endowment-regressive/progressive can be determined as follows.

1. If $\Psi(\cdot)$ is linear, $\{ssb_i\}$ are mortality-regressive.

2. If $\Psi(\cdot)$ is strictly concave, then $\{ssb_i\}$ are mortality-regressive (-progressive) if

$$\frac{u_{ssb_i} ssb_i}{v_i^R} + \frac{R_u}{R_\Psi} \, \frac{V_i}{\beta^{R-1} V_i^R} \; > (<) \; 1 \qquad \forall \;\; i \tag{25}$$

   where $v_i^R = u(ssb_i) + b$ is the flow utility per period after retirement.

3. $\{ssb_i\}$ are always endowment-progressive.

Proof: See Appendix C.

Intuition Consider the condition for mortality regressivity/progressivity in Eq (25). Similarly to Eq (7) in Section 3, it contains two parts. The first term represents the elasticity of per-period utility during retirement period v_i^R to pension benefits ssb_i, i.e., $\frac{u_{ssb_i} ssb_i}{v_i^R} = \frac{dv_i^R}{dssb_i} \cdot \frac{ssb_i}{v_i^R}$. It describes how sensitive is per period utility after retirement to the marginal change in pension benefits (or consumption).

The second term measures the relative concavity of the functions $u(\cdot)$ and $\Psi(\cdot)$, and thus captures how important is concern for consumption inequality compared to concern for inequality in lifetime utilities. Unlike in Eq (7), here the measure of relative concavity $\frac{R_u}{R_\Psi}$ is multiplied by the term $\frac{V_i}{\beta^{R-1}V_i^R}$, which is the ratio of lifetime utility V_i to the discounted lifetime utility during the retirement period, $\beta^{R-1}V_i^R$. Since the lifetime utility is the sum of utilities during working and retirement periods, this expression is larger than one. This multiplicative term makes the condition for mortality progressivity harder to meet, as it scales up the ratio $\frac{R_u}{R_\Psi}$, thus putting more weight on consumption inequality. Moreover, the smaller is V_i^R relative to V_i, the more pronounced this effect is. This is because inequality in mortality only plays a role after retirement. Thus, the less important is retirement period for total welfare, the less important is mortality inequality.

It is also worth making a quick note about the relationship between pensions and endowments. In the neutral case (when each individual receives pension benefits based on his own lifetime contribution IC_i), pension benefits are higher for people with high productivity ϵ_i. In contrast, optimal pensions are the same for all productivity types in the standard utilitarian welfare settings ($\Psi(\cdot)$ is linear), implying the redistribution form high- to low-productivity types. Introducing aversion to lifetime inequality ($\Psi(\cdot)$ is concave) makes pension benefits *decreasing* in productivity, thus involving even more pronounced progressive redistribution along the endowment dimension. As mentioned earlier, this is to compensate low-productive people for their low consumption during working years: as social planner does not control consumption at young age, more unequal pensions are used to reduce dispersion in lifetime utilities.

5.5 Two examples

To better illustrate the intuition, we once again consider two extreme cases differing in whether inequality in consumption or in lifetime utilities matters for welfare.

Case 1: Only aversion to consumption inequality matters ($\Psi(\cdot)$ is linear and $u(\cdot)$ is concave). In this case, optimal social security benefits are the same for all individuals and take the following form:

$$ssb_i = \frac{\overline{IC}}{\overline{q}} \quad \text{for } \forall\ i$$

Thus, social planner equalizes consumption after retirement by pooling together all individual contributions (or endowments), and by annuitizing equalized endowments at the pooled annuity price $\overline{q}$.

We can write down the average tax in this case as:

$$AT_i = 1 - \frac{\overline{IC}}{IC_i}\frac{q_i}{\overline{q}}$$

The tax increases in endowment IC_i but decreases in life expectancy or q_i, implying endowment progressivity but mortality regressivity.

Case 2: Only aversion to lifetime inequality matters ($\Psi(\cdot)$ is concave and $u(\cdot)$ is linear). Assuming $u(c) = c$ and using the assumption $\beta(1+r) = 1$, we can solve for ssb_i with the following result:

$$ssb_i = \underbrace{\frac{\overline{IC}}{q_i}}_{\text{part 1}} + b\underbrace{\frac{\overline{q} - q_i}{q_i}}_{\text{part 2}} + \underbrace{\frac{1-\tau}{\tau}\frac{\overline{IC} - IC_i}{q_i}}_{\text{part 3}}$$

The expression for optimal benefits contains three parts. The first part is the mortality-neutral allocation with fully-redistributed pension contributions, i.e., contributions are pooled together, divided equally among agents, and then converted into an annuity based on the individual actuarially-fair price. The second part is the compensation for short life which is positive (negative) for people with low (high) life expectancy. It is worth noting that these two parts are also present in the expression for optimal consumption when only lifetime inequality matters considered in the previous sections (see Eq (10) in Section 3 and Eq (15) in Section 4).

The third part is new: this is an additional redistributive component that social planner uses to compensate low-income people for having low consumption during their working years. This term arises because social planner can only affect consumption after retirement, while aiming to reduce inequality in lifetime utilities.

We can express the average tax in this case as:

$$AT_i = \frac{1}{\tau}\frac{IC_i - \overline{IC}}{IC_i} + b\frac{q_i - \overline{q}}{IC_i}$$

The average tax increases in life expectancy or q_i, and in endowment or IC_i, implying both mortality- and endowment progressivity.[3]

[3] It can be shown that AT_i increases in IC_i when $\frac{1}{\tau}\overline{IC} + b(\overline{q} - q_i) > 0$. The latter inequality always hold, otherwise ssb_i is negative.

6 Quantitative illustration

In this section, we aim to understand the relative quantitative importance of the theoretical mechanisms described above. To do this, we quantitatively solve a life-cycle model with two stages, working and retirement periods, as described in Section 5. We then use the model to assess the welfare effects of different pension designs that vary in the degree of mortality regressivity.

6.1 Model description

Individuals A model period is one year. Individuals enter the model at age $t = 25$. Up to age R, individuals receive labor income, after age R, individuals retire and receive pensions.

Individuals are ex-ante different in their type j, $j \in [1, ..., J]$ which is fixed throughout their life. Type affects individual's survival probability and labor productivity. As in Section 5, we assume agents survive with probability one till age R. For age $t > R$, we denote the type-dependent probability to survive from age t to $t+1$ as θ_t^j. The earnings of an individual are equal to λ_t^j, the idiosyncratic productivity that depends on age (t) and type (j).

During the working period ($t = 1...R$), an individual pays tax τ on his labor earnings, after retirement ($t = R+1...T$), he receives pension benefits ssb.

The state variables of an individual are assets (k_t), age (t), and type (j). The optimization problem of an individual can be represented as follows:

$$V_t(k_t, j) = \max_{c_t, k_{t+1}} \left\{ u(c_t) + b + \theta_t^j V_{t+1}(k_{t+1}, j) \right\} \tag{26}$$

subject to

$$k_t\,(1+r) + inc_t = k_{t+1} + c_t, \tag{27}$$

where

$$inc_t = \begin{cases} \lambda_t^j (1-\tau) & ;\text{ if } \; t \leq R \\ ssb & ;\text{ if } \; t > R \end{cases} \tag{28}$$

Social Security There is a pension system that collects contributions from the young and pays out benefits to the old. As in Section 5, we assume the economy is dynamically efficient, $n = r$. Denoting the distribution of agents over states as $\mathcal{M}(\cdot)$, we can write down the pension system balance equation as follows:

$$\int_{t \leq R} \tau \lambda_t^j \; \mathcal{M}(k, j, t) = \int_{t > R} ssb \; \mathcal{M}(k, j, t)$$

Note that due to the dynamic efficiency assumption, the balance equation is the same for the pay-as-you-go and fully funded systems (as discussed in Section 5.2). Since all agents survive with probability one till age R, the left-hand side of the pension balance equation represents the average of individuals' pension contributions IC_j, where

$$IC_j = \tau \sum_{t=1}^{R} (1+n)^{t-1} \lambda_t^j = \tau \sum_{t=1}^{R} (1+r)^{t-1} \lambda_t^j$$

The last equality follows from the dynamic efficiency assumption, and it allows us to think of IC_j as a notional pension balance or individual endowment.

As before, we consider pension benefits as the annuitized value of endowments. Our goal is to examine how varying degree of mortality progressivity of the pension system affects welfare. To do this, as in Section 4, we consider the environment where endowments IC_j cannot be changed. In other words, we restrict the ability of the pension system to directly redistribute individual contributions/endowments.

Welfare We aggregate the lifetime utilities of different individuals into the ex-ante measure of welfare as follows:

$$EW = \int \Psi \left(V_1 \left(k_1, j \right) \right) \mathcal{M} \left(k_1, j \right) \tag{29}$$

We assume that both the utility function $u(\cdot)$ and the function used to aggregate individual welfare $\Psi(\cdot)$ are of the CRRA type with risk aversion parameters $R_u = \sigma$ and $R_\Psi = \gamma$:

$$u(c) = \frac{c^{1-\sigma}}{1-\sigma}$$

and

$$\Psi(V_1) = \frac{V_1^{1-\gamma}}{1-\gamma}.$$

The welfare effects of each experiment are computed as follows. We treat the utmost mortality-regressive case (pooled annuitization) as a benchmark, and we denote the corresponding ex-ante welfare as EW^{BS}. Consider a situation when every agent receives cash transfer Δ every period, and denote the corresponding ex-ante welfare as $EW(\Delta)$. Note that if $\Delta = 0$, we have the benchmark welfare: $EW(0) = EW^{BS}$.

Denote the ex-ante welfare in the experimental economy as EW^{Exp}. We compute the cash transfers needed to make average welfare in the baseline and experimental economies the same (Δ^*) by solving the following equation:

$$EW(\Delta^*) = EW^{Exp}$$

Our welfare measure CEV is expressed as a percentage of average consumption:

$$CEV = \frac{\Delta^*}{\overline{c}}$$

6.2 Parameterization

In our model, we focus on male individuals. Within the same gender group, education and race are important determinants of longevity inequality. To capture this, the ex-ante fixed type in our model is a pair of fixed characteristics, education and race, $j = (ed, ra)$. There are three education types and two race types, thus the total number of types J is equal to 6. Three education types correspond to high-school dropouts ($ed = 1$), people with only high-school degree or high-school degree and some college education but no college degree ($ed = 2$), and people with college or higher degree ($ed = 3$). Two race groups correspond to non-white ($ra = 1$) and white ($ra = 2$).

We set the retirement age $R + 1$ to 65. For age $t > R$, the conditional probability to survive from age t to $t + 1$ (θ_t^j) is estimated using the Health and Retirement Study (HRS) dataset. In our estimation, we use a sample of male individuals and estimate a logit model which depends on a set of age, race, and education dummy variables. Our estimated survival probabilities are plotted in the left panel of Figure 2.

To estimate labor productivity for each type, λ_t^j, we use the Panel Study of Income Dynamics (PSID). We use a sample of male workers, where we define a person as employed if he works at least 520 hours per year, and earns at least the federal minimum wage. We normalize labor income to 2002 base year using the Consumer Price Index (CPI). Our estimated labor income profiles are plotted in the right panel of Figure 2.

We set population growth n and interest rate r to 1%. As in our theoretical section, we maintain the assumption $\beta(1+r) = 1$, hence we set the discount rate β to the inverse of $1+r$. We set risk aversion over consumption, σ, to 2, which is a common value used in structural life-cycle models. We set the initial distribution of each type based on the fraction of people in each race/education group in the HRS at age 65. We consider several alternative values for the aversion to inequality in lifetime utilities (γ), and non-pecuniary utility of being alive (b).

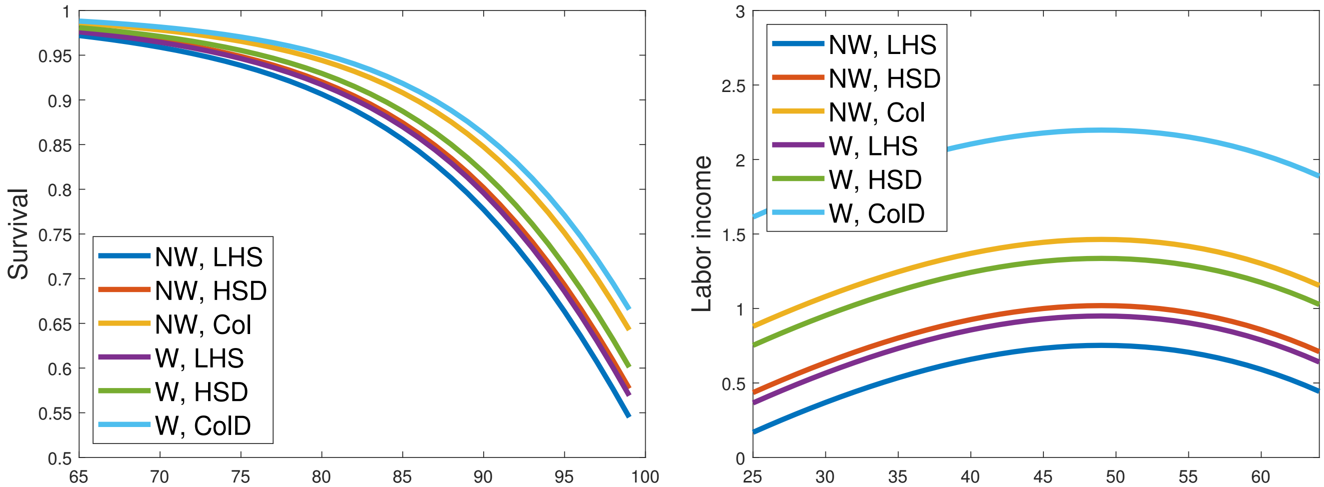

Figure 2: Left panel: survival probabilities by race and education. Right panel: labor income (normalized by average earnings) by race and education. The abbreviations are as follows: LHS - less than high-school degree ($ed = 1$), HSD - high-school degree ($ed = 2$), ColD - college degree and above ($ed = 3$). NW stands for non-whites ($ra = 1$), and W stands for whites ($ra = 2$).

6.3 Results

In this section, we use our quantitative life-cycle model for the comparative welfare analysis of different pension arrangements. We start by describing how we set our quantitative experiments. We then evaluate the welfare effects of restoring mortality-neutrality using the utmost mortality-regressive case as a benchmark. Finally, we compare a wider range of pension policies varying in the degree of mortality regressivity/progressivity.

6.3.1 Setup

To better understand the underlying mechanisms when changing the degree of mortality progressivity in a pension system, we consider two versions of our model. The first model, which we refer to as "only mortality heterogeneity", is when people have the same endowments throughout their lifetime. To construct this model, we assume that all types have the same labor productivity equal to the average in their age group: $\lambda_t^j = \overline{\lambda}_t$ for all j. In this case, all individuals have exactly the same pension balances or contributions to the pension system: $IC_j = IC$ for all j. The second model, which we refer to as "mortality and income heterogeneity", corresponds to the full model, i.e., each type has different labor income and mortality.

6.3.2 Moving to mortality-neutrality

The starting point of our analysis is the utmost mortality-regressive case or pooled annuitization, which is a common feature of many pension systems. We first consider the welfare effects of restoring mortality-neutrality. Below we explain how we compute pension benefits in the utmost mortality-regressive and mortality-neutral cases.

The utmost mortality-regressive case In the utmost mortality-regressive case, pension benefits are the same for all people in the model "only mortality heterogeneity". This is because in this model, labor income and thus pension balances do not differ across types, and converting them into annuity based on the pooled price results in the same pension benefits:

$$ssb_j = \frac{IC}{\overline{q}} \quad \text{for } \forall\ j$$

These pension benefits are plotted as a solid line in the left panel of Figure 3. It is worth noting that in this environment, even though people have the same income and the same pensions, their consumption still differs as they have different optimal savings because of the difference in longevity.

In the "mortality and income heterogeneity" model, pension benefits differ across types because of the difference in pension balances:

$$ssb_j = \frac{IC_j}{\overline{q}}$$

These benefits are plotted as a solid line in the right panel of Figure 3.

Mortality-neutral case In the mortality-neutral case, pension benefits are based on individual mortality. For "only mortality heterogeneity" model, pension benefits take the following form:

$$ssb_j = \frac{IC}{q_j}$$

The benefits are plotted as a dashed line in the left panel of Figure 3. In the "mortality and income heterogeneity" model, we have

$$ssb_j = \frac{IC_j}{q_j}$$

These pension benefits are plotted as a dashed line in the right panel of Figure 3.

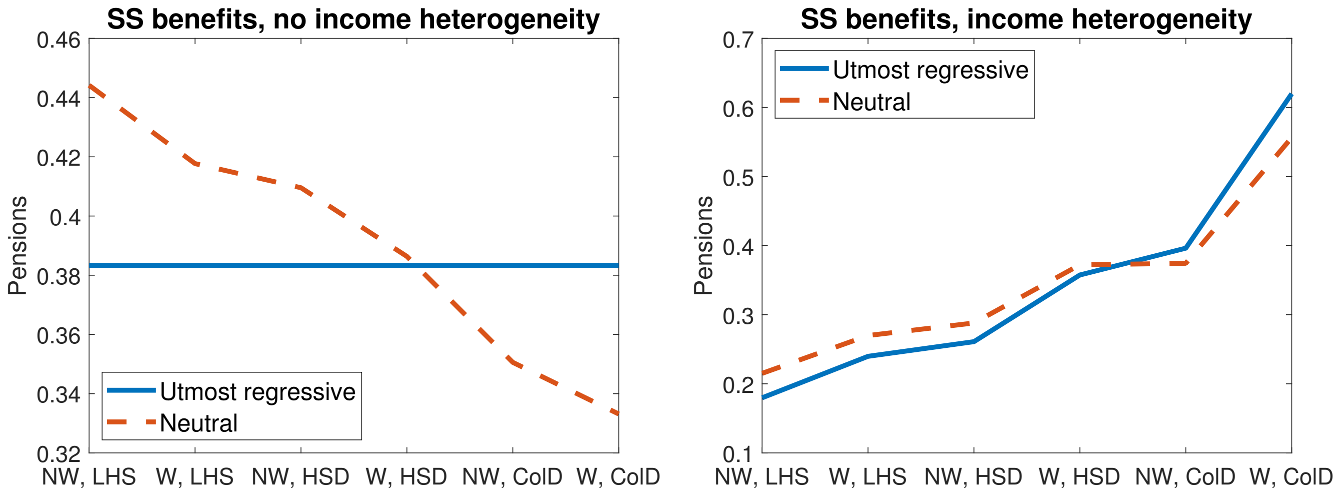

Figure 3: Pensions benefits (normalized by average income), the utmost mortality-regressive (pooled annuitization) versus the mortality-neutral cases. Left panel: only mortality heterogeneity. Right panel: mortality and income heterogeneity. The abbreviations are as follows: LHS - less than high-school degree ($ed = 1$), HSD - high-school degree ($ed = 2$), ColD - college degree and above ($ed = 3$). NW stands for non-whites ($ra = 1$), and W stands for whites ($ra = 2$).

An important observation from Figure 3 is that when there is no heterogeneity in labor income, mortality-neutral pension benefits are higher for short-lived groups. This is because pension balances are the same for all groups, while annuity prices are lower for the short-lived, allowing them to get higher pension benefits. This is, however, no longer the case when labor income differs by type. In this case, there are two forces: on the one hand, the short-lived get their pension balances converted to benefits at a lower price. On the other hand, their balances are lower because of lower labor income. The latter effect dominates, making pension benefits lower for the short-lived.

The welfare effects of moving to mortality-neutrality for the two versions of the model are presented in Figure 4. Each graph in the figure shows CEV for different degrees of aversion to inequality in lifetime utility γ, varying from 0 to 5, and for different levels of non-pecuniary utility of being alive b, varying from 10 to 50.

Consistent with our results in the previous sections, welfare effects are sensitive to non-pecuniary benefits of being alive b, and to the aversion to inequality in lifetime utilities γ. For the model ”only mortality heterogeneity” (left panel) welfare effects of mortality-neutrality become positive even before γ exceeds risk aversion over consumption σ, which is equal to 2. This happens because in our quantitative model, we allow for savings. This generates additional consumption inequality across types, and hence increases the push for mortality progressivity.

This effect becomes more evident when considering the model ”mortality and income heterogeneity” (right panel). Here mortality-neutrality is always welfare-improving, even

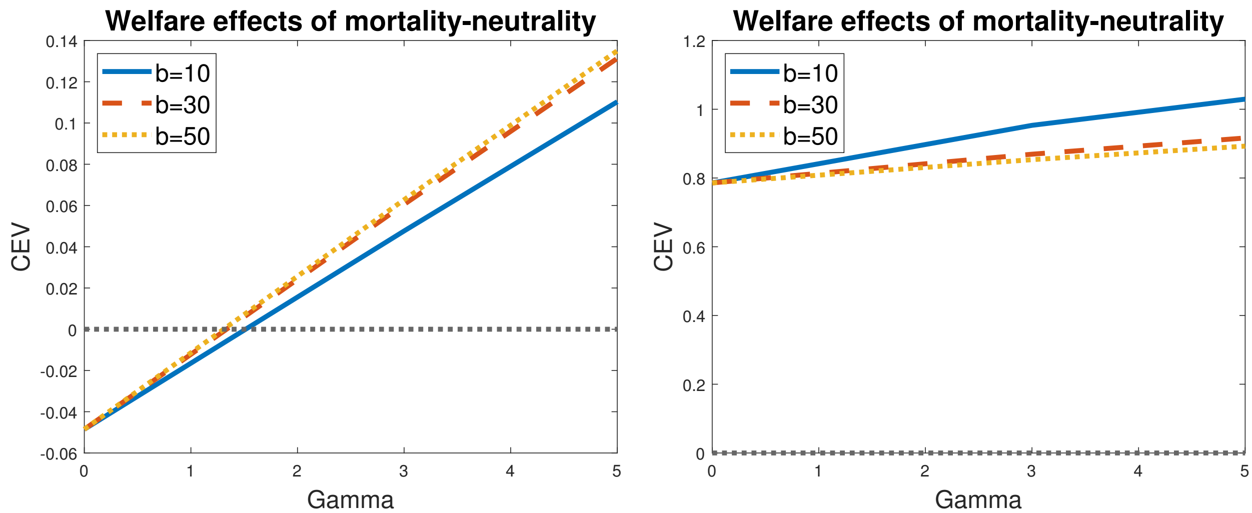

Figure 4: Welfare effects when moving from the utmost mortality-regressive (pooled annuitization) to mortality-neutral pensions. Left panel: only mortality heterogeneity. Right panel: mortality and income heterogeneity.

when there is no aversion to lifetime inequality, i.e., $\gamma = 0$. In this case, the endowment inequality between types is large, and mortality-neutrality reduces this inequality. However, it is important to point out that the reduction in inequality when moving to mortality-neutral pensions is small (see the right panel of Figure 3). This is because inequality in pensions due to inequality in labor income is much more important than that generated by difference in mortality, suggesting it can be optimal to move beyond mortality-neutrality and towards a more progressive system.

6.3.3 Moving to mortality progressivity

We next turn to evaluating welfare effects of different pension designs ranging from the utmost mortality-regressive to mortality-progressive. Specifically, following the analysis in Section 4, we consider pension systems where benefits are determined as follows:

$$ssb_j = (1 - \alpha_1)\,\frac{IC_j}{\overline{q}^{IC}} + \alpha_1\,\frac{IC_j}{q_j} \;+\; \alpha_2\,\frac{\overline{q} - q_j}{q_j}$$

Here $\overline{q}^{IC}$ is the endowment-weighted average annuity price.[4] In this expression, α_2 represents the compensation coefficient that increases consumption of people with low life expectancy ($q_i < \overline{q}$) at the cost of decreasing consumption of people with high life expectancy. By varying α_1 and α_2, we can move from the utmost mortality-regressive to mortality-progressive case

[4] As in Section 4.3, we use the endowment-weighted average annuity price so that the total pension payments are the same, and the pension balance equation is met in each case.

as summarized in Table 1. Importantly, the pension system balance equation is met for all considered combinations of α_1 and α_2.

	α_1	α_2
Utmost regressive (pooled price)	0	0
Weakly regressive	0.5	0
Neutral	1	0
Progressive	1	0.75

Table 1: Parametrization for different degrees of mortality progressivity

The results of this exercise are presented in Figure 5. To construct this figure, we fix the level of non-pecuniary utility of life b at 30. Each graph in the figure reports CEV when moving from regressive to progressive cases described in Table 1 for different levels of aversion to lifetime inequality γ. As in the previous analysis, all welfare results are reported using the utmost mortality-regressive case as a benchmark.

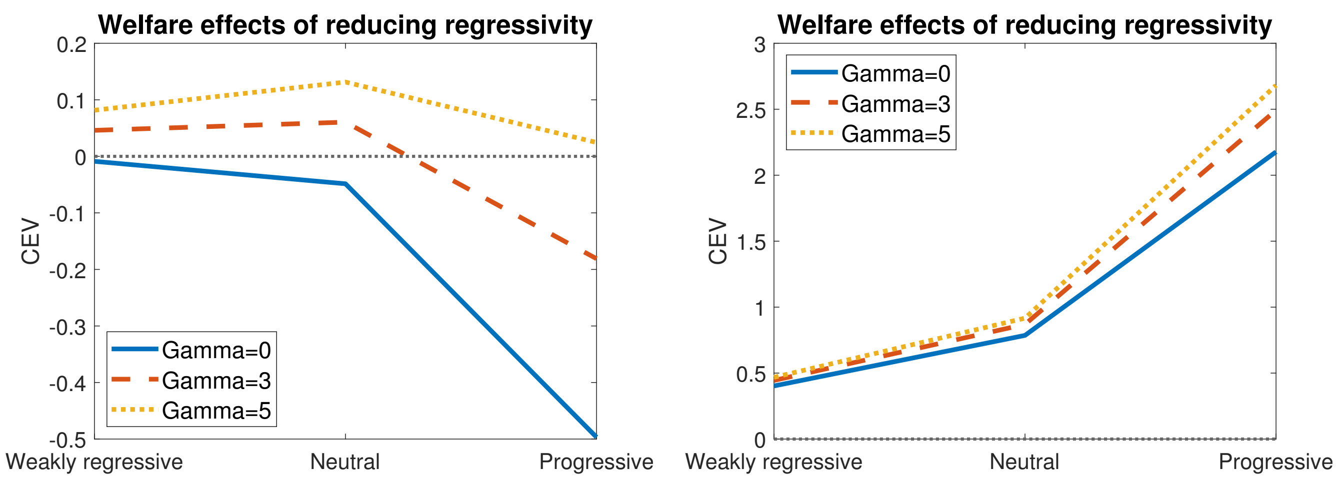

Figure 5: Welfare effects when changing mortality progressivity. Left panel: only mortality heterogeneity. Right panel: mortality and income heterogeneity.

Several important observations from Figure 5 are as follows. For the model "only mortality heterogeneity" (left panel), mortality progressivity is optimal when aversion to lifetime inequality is high. For the model "mortality and income heterogeneity", mortality progressivity is welfare improving for all the combinations of parameters considered, including the case with no aversion to lifetime inequality ($\gamma = 0$). Moreover, the welfare gains of moving to mortality progressivity in the full model can be quite substantial. For example, the CEV is equal to 2.5% when the aversion to lifetime inequality γ is equal to 5. It is important to point out, however, that these gains arise not because high-mortality types get compensated for their short life. Instead, the gains are due to the ability of mortality progressivity to reduce inequality coming from dispersion in pension contributions. This is because mortality and labor income are highly correlated, and there is a constraint that pension contributions cannot be directly redistributed.

7 Conclusion

We develop a framework for understanding the optimality of redistribution along the mortality dimension in isolation from the issue of income redistribution. While income and mortality are correlated, the correlation is not perfect, and the welfare effects of redistribution along these two dimensions may be different. In particular, the standard utilitarian approach favors redistribution from high-income to low-income groups, while at the same time, favoring redistribution from the short-lived to the long-lived.

Our goal is to understand when the redistribution from the short-lived to the long-lived is not optimal. Three important features of our analysis are as follows. First, we deviate from the standard utilitarian welfare criterion only by introducing aversion to inequality in lifetime utilities in addition to aversion to consumption inequality. Second, we define a neutral case which involves no redistribution along the mortality dimension and use it as a benchmark to assess mortality regressivity/progressivity. Third, we treat life as valuable, which implies that people who live long have utility advantage over the short-lived even when there is no difference in their consumption.

Using this framework, we derive several interesting results. First, when people differ in life expectancy but not in income, mortality progressivity is optimal only when life is valuable and aversion to lifetime inequality is stronger than aversion to consumption inequality. Put differently, adding concern about lifetime inequality is not enough to make mortality progressivity optimal unless this concern dominates aversion to consumption inequality.

Second, when people differ in their income, and income and mortality are negatively correlated, mortality progressivity can also optimally arise when income-redistributive tools are limited. This happens because in this case, mortality progressivity can partially substitute for income progressivity.

We then apply our framework to the analysis of pension design. We show that optimal pension benefits are mortality-regressive unless social planner has aversion to lifetime inequality. At the same time, when concern for lifetime inequality enters pension design, it creates an additional effect of increasing the degree of optimal income progressivity. This happens because it becomes optimal not only to compensate high-mortality people for their short life, but also to compensate the poor for their low consumption during working life.

In the final part of our analysis, we construct a quantitative life-cycle model to numerically illustrate the welfare effects of increasing mortality progressivity of pension benefits, while restricting income redistribution and preserving revenue-neutrality. We show that welfare effects of mortality-related redistribution crucially depend on the degree of aversion to lifetime inequality, and (even more so) on the income-mortality correlation.

Our results thus emphasize the distinction between two reasons for why mortality progressivity may be optimal: (i) to compensate the short-lived for their short life; (ii) as a substitute for income progressivity. The first case arises when life is valuable and social welfare incorporates strong aversion to lifetime inequality, and the second - when income and mortality are correlated and there are limited instruments to redistribute income. While similar in its final effect (welfare gains from mortality-progressive pensions), these two cases differ fundamentally. In the first case, the results are driven by the concern for the short-lived, and in the second - by the concern for the poor.

Our results thus create an important framework for better understanding welfare effects of pension policies when people differ in non-pecuniary factors (life expectancy). Pension systems in many countries, including in the US, are under strain from the changing demographics, and various ways to maintain their sustainability are widely discussed. Many considered policies are judged based on the standard utilitarian welfare criteria, and thus are favoring the short-lived over the long-lived. We propose a way to refine the measurement of welfare consequences of pension reforms, and separate the gains accruing to the poor from that to the short-lived.

References

[1] Adema, Y., Bonenkamp, J., Meijdam, L., 2016. Flexible Pension Take-up in Social Security. International Tax Public Finance, 23, pp 316-342

[2] Atkinson, A., Bourguignon, F., 1982. The Comparison of Multi-Dimensioned Distributions of Economic Status. Review of Economic Studies, XLIX, pp 183-201

[3] Bagchi, S., 2019. Differential Mortality and the Progressivity of Social Security. Journal of Public Economics.

[4] Bagchi, S., Jung, J., 2021. Health Risks and Welfare Effects of Social Security. Mimeo, Townson University

[5] Bommier A., Leroux, M-L., Lozachmeur, J-M.., 2011a. Differential mortality and social security. Canadian Journal of Economics, 44(1), pp 273-289

[6] Bommier A., Leroux, M-L., Lozachmeur, J-M.., 2011b. On the public economics of annuities with differential mortality. Journal of Public Economics, 95(7-8), pp 612-623

[7] Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., Cutler, D., 2016. The Association between Income and Life Expectancy in the United States. JAMA, 2016, 316 (6), pp 1750-1766.

[8] Conesa, J., Krueger, D., 1999. Social Security Reform with Heterogeneous Agents. Review of Economic Dynamics, 2, pp 757-795

[9] Coronado, J., Fullerton, D., Glass, T., 2011. The progressivity of Social Security. B.E. Journal of Economic Analysis and Policy, 11(1), pp 1–45.

[10] Cremer, H., Lozachmeur, J-M., Pestieau, P., 2010. Collective Annuities and Redistribution. Journal of Public Economic Theory, 12 (1), pp 23-41.

[11] De Nardi, M., Pashchenko, S., Porapakkarm, P., 2022. The Lifetime Cots of Bad Health. NBER Working Paper No. 23963

[12] Hall, R., Jones, C., 2007. The Value of Life and the Rise in Health Spending. Quarterly Journal of Economics, vol.122(1) , pp. 39-72.

[13] Hosseini, R., Shourideh, A., 2019. Retirement Financing: An Optimal Reform Approach. Econometrica, 87 (4), pp 1205-1265

[14] Huggett, M., Parra, J., 2010. How Well Does the US Social Insurance System Provide Social Insurance? Journal of Political Economy, 118 (1), pp 76-112.

[15] Jones, J., Li, Y., 2022. Social Security Reform with Heterogeneous Mortality. Review of Economic Dynamics

[16] Kitao, S., 2014. Sustainable Social Security : Four Options. Review of Economic Dynamics, Vol17(4), pp. 756-779.

[17] Imrohoroglu, S., Kitao, S., 2012. Social Security Reforms: Benefit Claiming, Labor Force Participation, and Long-Run Sustainability. American Economic Journal: Macroeconomics, 4(3), pp 96-127

[18] Laun, T., Markussen, S., Vigtel, T., Wallenius, J., 2019. Health, Longevity and Retirement Reform. Journal of Economic Dynamics and Control, 103, pp 123-157.

[19] Leroux, M-L., Ponthiere, G., 2013. Utilitarianism and Unequal Longevities. A Remedy? Economic Modeling, 30, pp 888-899

[20] Murphy, K., Topel, R., 2006, "The Value of Health and Longevity", Journal of Political Economy, vol.114, N51, pp. 871-904

[21] Ndiaye, A., 2020. Flexible Retirement and Optimal Taxation. Working Paper No. 2018-18, Federal Reserve Bank of Chicago.

[22] Nishiyama, S., Smetters, K., 2007. Does Social Security Privatization Produce Efficiency Gains? The Quarterly Journal of Economics

[23] Pashchenko, S., 2013. Accounting for Non-Annuitization. Journal of Public Economics, 98, pp. 53-67.

[24] Pashchenko, S., Porapakkarm, P., 2021. Accounting for Social Security Claiming Behavior. Mimeo, University of Georgia

[25] Pashchenko, S., Porapakkarm, P., 2022. Value of Life and Annuity Demand. Journal of Risk and Insurance.

[26] Pestieau, P., Racionero, M., 2016. Harsh Occupations, Life Expectancy and Social Security. Economic Modeling, 58, pp 194-202

[27] Rosen, S., 1988, "The Value of Changes in Life Expectancy", Journal of Risk and Uncertainty, vol.I, pp.285-304.

[28] Sanchez-Romero, M., Lee, R., Prskawetz, A., 2019. Redistributive Effects of Different Pension Systems When Longevity Varies by Socioeconomic Status, NBER Working Paper No. 25944

[29] Sheshinski, E., Caliendo, F., 2020. Social Security and the Increasing Longevity Gap. Journal of Public Economic Theory, 23, pp 29–52.

Appendix

A Proof of Proposition 1.1

In this section, we provide the poof of Proposition 1.1. We first restate the proposition.

Proposition 1.1 Consider a feasible allocation $\{c_i\}$. If this allocation is mortality-regressive (mortality-progressive) then the following is true:

$$\varepsilon^c_{q_i\,|\,a_i} > (<) -1 \quad \forall \ i,$$

where $\varepsilon^c_{q_i\,|\,a_i}$ is the partial elasticity of consumption to mortality, i.e., the elasticity for a given level of endowment a_i:

$$\varepsilon^c_{q_i\,|\,a_i} = \left.\frac{dc_i}{dq_i}\right|_{a_i} \cdot \frac{q_i}{c_i}.$$

Proof We will do the proof for the mortality-regressive case. That for the mortality-progressive case is analogous. Using the definition of the neutral allocation $\{c_i^N\}$, we can rewrite the average tax as follows:

$$AT_i = 1 - \frac{c_i q_i}{a_i}$$

The partial derivative of the average tax with respect to q_i for a given level of endowment a_i takes the form:

$$\left.\frac{\partial AT_i}{\partial q_i}\right|_{a_i} = -\frac{1}{a_i}\left.\frac{\partial(c_i q_i)}{\partial q_i}\right|_{a_i} = -\frac{c_i}{a_i}\,(\varepsilon^c_{q_i\,|\,a_i} + 1),$$

where the last equality follows from the definition of the partial elasticity. Since in the mortality-regressive case $\left.\frac{\partial AT_i}{\partial q_i}\right|_{a_i} < 0$, it follows that $\varepsilon^c_{q_i\,|\,a_i} > -1$. This finished the proof of the proposition.

B Proof of Proposition 4

In this section, we provide the poof of Proposition 4. We first restate the proposition.

Proposition 4 Consider the constrained social planner problem described in Eq (18), and suppose Assumptions 1-4 hold. In addition, assume $R_\Psi = 0$ and $R_u > 1$. The optimal choice of α_1 and α_2 can be summarized as follows:

(1) If $cov(q,a)=0$, then at the optimum $\alpha_1 = 0$,

(2) If $cov(q,a) > 0$, then

(i) At the optimum $\alpha_1 > 0$,

(ii) If $cov(q, \frac{a}{q}) < 0$, then at the optimum $\alpha_2 = 0$,

(iii) If $cov(q, \frac{a}{q}) > 0$, then at the optimum $\alpha_2 > 0$.

Proof The first-order condition for the choice of α_1 can be written as follows:

$$(\alpha_1): \qquad \int\limits_q \int\limits_a u_c\, q \left(\frac{a}{q} - \frac{a}{\overline{q}^a}\right) H(a,q)\, da\, dq \;=\; 0 \tag{30}$$

Consider this equation evaluated at $\alpha_1 = 0$. Based on the consumption allocation rule in Eq(17), we have $c_i = \frac{a_i}{\overline{q}^a}$ (since when $\alpha_1 = 0$, we have $\alpha_2 = 0$ as well).

Consider first the case when $cov(q,a) = 0$. In this case, c_i varies with a_i but not with q_i. Thus, $cov(q,c) = 0$ and hence $cov(q,u_c) = 0$. Based on this, we can transform Eq(30) as follows:

$$\int\limits_q \int\limits_A u_c a\, H(a,q)\, da\, dq - \frac{1}{\overline{q}^a} \int\limits_q \int\limits_a u_c a\, q\, H(a,q)\, da\, dq =$$

$$= \int\limits_q \int\limits_a u_c a\, H(a,q)\, da\, dq - \frac{\overline{q}}{\overline{q}^a} \int\limits_q \int\limits_a u_c a\, H(a,q)\, da\, dq = 0$$

where in the last equality we used the fact that $\overline{q} = \overline{q}^a$ when $cov(q,a) = 0$. The first-order condition holds with equality when $\alpha_1 = 0$ implying that this represents the optimum. This finishes the proof of part (1) of the proposition.

Consider next the case when $cov(q,a) > 0$. Using the fact that $u(\cdot)$ is the CRRA function with risk aversion R_u (Assumption 4), we have $u_c = c^{-R_u}$, and hence we can rewrite the left-hand side of the FOC as follows:

$$(\overline{q}^a)^{R_u} \int\limits_q \int\limits_a a^{1-R_u} \left(1 - \frac{q}{\overline{q}^a}\right) H(a,q)\, da\, dq \;= (\overline{q}^a)^{R_u} \int\limits_q \int\limits_a \phi(a)\mu(q) H(a,q)\, da\, dq, \tag{31}$$

where $\phi(a) \equiv a^{1-R_u}$ and $\mu(q) \equiv 1 - \frac{q}{\overline{q}^a}$. Both functions are monotone decreasing in its respective arguments (for the first function it follows from $R_u > 1$), and have positive means (for the second function it follows from the positive correlation between q and

a, which implies $\overline{q} < \overline{q}^a$). Since $cov(q,a) > 0$, we thus have $cov(\mu(q),\phi(a)) > 0$, and $\int\limits_q \int\limits_a \phi(a)\mu(q)H(a,q)\,da\,dq > 0$. Thus, the FOC in Eq (31) is positive when evaluated at $\alpha_1 = 0$, meaning it is optimal to increase the value of this variable. This finishes the proof of part 2(i) of Proposition 4.

To prove parts (ii) and (iii) of the proposition, we turn to the FOC for the choice of α_2, which can be written as follows:

$$\int\limits_q \int\limits_a u_c\,(\overline{q} - q)\,H(a,q)\,da\,dq \,=\, 0 \tag{32}$$

Consider this equation evaluated at $\alpha_1 = 1$ and $\alpha_2 = 0$. In this case, the consumption allocation rule in Eq (17) becomes $c_i = \dfrac{a_i}{q_i}$. The FOC in Eq (32) can be transformed as follows:

$$\int\limits_q \int\limits_a u_c\,(\overline{q} - q)\,H(a,q)\,da\,dq \,=\, -cov(u_c,q)$$

Since $u_c = \left(\dfrac{a}{q}\right)^{-R_u}$, the correlation between u_{ci} and q_i depends on the sign of $cov(q, \dfrac{a}{q})$. Let us consider two cases.

First, when $cov(q, \dfrac{a}{q}) < 0$, we have $cov(u_c, q) > 0$. This means that the left-hand side of the FOC in Eq (32) is negative, implying that it is not optimal to increase α_2 above zero.[5] This finishes the proof of part 2(ii) of Proposition 4.

Second, when $cov(q, \dfrac{a}{q}) > 0$, we have $cov(u_c, q) < 0$, implying that the left-hand side of the FOC in Eq(32) is greater than zero, thus it is optimal to set $\alpha_2 > 0$. This finishes the proof of part 2(iii) of Proposition 4.

Comparing conditions $cov(q,a) > 0$ and $cov(q, \dfrac{a}{q}) > 0$ We argue that the condition $cov(q, \dfrac{a}{q}) > 0$ is stronger than the condition $cov(q,a) > 0$, and we formally prove it in the auxiliary proposition below.

Auxiliary proposition If $cov(q, \dfrac{a}{q}) > 0$ then $cov(q,a) > 0$, while the reverse is not true.

[5] In fact, based on the FOC in Eq (32) it is optimal to make α_2 negative. This happens because in this case, setting $\alpha_1 = 1$ is not optimal.

Proof We can use the definition of covariance to express $cov(q, \dfrac{a}{q})$ as follows:

$$cov(q, \frac{a}{q}) = \int\limits_q \int\limits_a q \cdot \frac{a}{q} \; H(a, q) \, da \, dq \;\; - \;\; \overline{q} \int\limits_q \int\limits_a \frac{a}{q} \; H(a, q) \, da \, dq \;\; =$$

$$= \overline{q} \left(\frac{\overline{A}}{\overline{q}} - \overline{\left[\frac{a}{q} \right]} \right),$$

where

$$\overline{\left[\frac{a}{q} \right]} \equiv \int\limits_q \int\limits_a \frac{a}{q} \; H(a, q) \, da \, dq$$

To approximate $\overline{\left[\dfrac{a}{q} \right]}$, we can use Taylor expansion around the means of a and q, $\overline{A}$ and $\overline{q}$, respectively. This results in the following expression:

$$\overline{\left[\frac{a}{q} \right]} \approx \frac{\overline{A}}{\overline{q}} + \frac{\overline{A}}{\overline{q}} \cdot \frac{Var(q)}{\overline{q}^2} - \frac{cov(q, a)}{\overline{q}^2}$$

If $cov(q, \dfrac{a}{q}) > 0$, we have $\dfrac{\overline{A}}{\overline{q}} > \overline{\left[\dfrac{a}{q} \right]}$. This implies

$$\frac{cov(q, a)}{\overline{q}^2} > \frac{\overline{A}}{\overline{q}} \cdot \frac{Var(q)}{\overline{q}^2} > 0$$

Thus, $cov(q, a) > 0$.

On the other hand, when $cov(q, a) > 0$, it is still possible to have $\dfrac{\overline{A}}{\overline{q}} < \overline{\left[\dfrac{a}{q} \right]}$, and thus $cov(q, \dfrac{a}{q}) < 0$. This finishes the proof of the auxiliary proposition.

C Proof of Proposition 2.1

In this section, we provide the poof of Proposition 2.1. We are going to start by giving two additional definitions and by formulating and proving an additional proposition.

We define two partial elasticities of pension benefits, with respect to mortality and endowment (or lifetime pension contributions). These elasticities show how pension benefits change with mortality (endowment), while keeping endowment (mortality) fixed:

$$\varepsilon^{ssb}_{q_i \,|\, IC_i} = \left. \frac{dssb_i}{dq_i} \right|_{IC_i} \cdot \frac{q_i}{ssb_i}$$

$$\varepsilon^{ssb}_{IC_i\,|\,q_i} = \left.\frac{dssb_i}{dIC_i}\right|_{q_i} \cdot \frac{IC_i}{ssb_i}$$

To understand whether pensions benefits are progressive/regressive along mortality and endowment dimensions, we can use the modified version of Proposition 1 from Section 3, which is stated below.

Proposition 1.2 Consider feasible pension benefits $\{ssb_i\}$. If these benefits are mortality-regressive/progressive then the following is true:

$$\varepsilon^{ssb}_{q_i\,|\,IC_i} > (<) -1 \quad \forall \;\; i$$

If these benefits are endowment-regressive/progressive then the following is true:

$$\varepsilon^{ssb}_{IC_i\,|\,q_i} > (<)\; 1 \quad \forall \;\; i$$

Proof We start by doing the proof for the mortality-regressive case. That for the mortality-progressivity case is analogous. In the mortality-regressive case, we have

$$\left.\frac{\partial AT_i}{\partial q_i}\right|_{IC_i} = -\frac{1}{IC_i}\left.\frac{\partial(ssb_i q_i)}{\partial q_i}\right|_{IC_i} = -\frac{ssb_i}{IC_i}(\varepsilon^{ssb}_{q_i\,|\,IC_i} + 1) < 0.$$

From here it follows that $\varepsilon^{ssb}_{q_i\,|\,IC_i} > -1$, which finishes the proof of the proposition.

We next consider the case of endowment regressivity. That for the endowment progressivity case is analogous. In the endowment-regressive case, we have

$$\left.\frac{\partial AT_i}{\partial IC_i}\right|_{q_i} = -q_i \left.\frac{\partial\left(\dfrac{ssb_i}{IC_i}\right)}{\partial IC_i}\right|_{q_i} = -\frac{q_i\;ssb_i}{IC_i^2}(\varepsilon^{ssb}_{IC_i\,|\,q_i} - 1) < 0.$$

From here it follows that $\varepsilon^{ssb}_{IC_i\,|\,q_i} > 1$, which finishes the proof of the proposition.

We next will restate and then prove Proposition 2.1.

Proposition 2.1 Consider pension benefits $\{ssb_i\}$ that represent the solution to the social planner problem described in Eqs (22)-(23). Under Assumptions 1-4, whether these benefits are mortality- and endowment-regressive/progressive can be determined as follows:

1. If $\Psi(\cdot)$ is linear, $\{ssb_i\}$ are mortality-regressive.

2. If $\Psi(\cdot)$ is strictly concave, then $\{ssb_i\}$ are mortality-regressive (-progressive) if

$$\frac{u_{ssb_i} ssb_i}{v_i^R} + \frac{R_u}{R_\Psi} \frac{V_i}{\beta^{R-1} V_i^R} > (<)\ 1 \qquad \forall \ \ i \tag{33}$$

3. $\{ssb_i\}$ are always endowment-progressive.

Proof:

Following the same steps as when proving Proposition 2, we can take a full differential of Eq (24) around the optimal allocation while keeping either IC_i or q_i fixed. This allows us to obtain the following expressions for partial elasticities:

$$\varepsilon^{ssb}_{q_i \,|\, IC_i} = - \frac{R_\Psi}{R_\Psi \dfrac{u_{ssb_i} ssb_i}{v_i^R} + R_u \dfrac{V_i}{\beta^{R-1} V_i^R}}$$

$$\varepsilon^{ssb}_{IC_i \,|\, q_i} = - \frac{R_\Psi \dfrac{u^W_{c_i} c^W_i}{v^W_i} \dfrac{V^W_i}{V_i}}{R_\Psi \dfrac{u_{ssb_i} ssb_i}{v_i^R} \dfrac{\beta^{R-1} V_i^R}{V_i} + R_u}$$

Consider first the partial elasticity of pension benefits with respect to q, $\varepsilon^{ssb}_{q_i \,|\, IC_i}$. Based on Proposition 1.2, $\{ssb_i\}$ are mortality-regressive(-progressive) if this expression is greater (less) than negative one. From here, parts 1 and 2 follow directly.

Consider next the partial elasticity of pension benefits with respect to IC, $\varepsilon^{ssb}_{IC_i \,|\, q_i}$. Based on Proposition 1.2, $\{ssb_i\}$ are endowment-regressive(-progressive) if this expression is greater (less) than one. Given that all agents have positive flow utility of being alive every period ($v^W_i > 0$ and $v^R_i > 0$ for all i), $\varepsilon^{ssb}_{IC_i \,|\, q_i}$ is always negative and hence less than one, implying endowment progressivity. This proves part 3 of the proposition.