The Optimal Earnings Test and Retirement Behavior

Masayuki Okada*

Abstract

This paper quantitatively derives the welfare-improving earnings test within an optimal income tax framework. I construct a life cycle model of labor supply and savings to compute social welfare. The preference parameters are estimated by the method of simulated moments using Japanese data. I find that social welfare under the current earnings test with large changes of marginal tax rates at thresholds is substantially lower than social welfare under the earnings test with a linear tax rate. In addition, an earnings test with negative marginal tax rates will increase social welfare more than a system without negative marginal tax rates.

JEL codes: J22; J26; J32; H21; H55

Keywords: earnings test, retirement, optimal income tax, negative marginal tax rate.

1 Introduction

An aging population globally causes the decline of working-age populations and increases fiscal pressure. To maintain fiscal sustainability, the Organisation for Economic Co-operation and Development (OECD) countries have changed social security systems since the 1990s. Despite these

*Ph.D. student at New York University. Email: mo2192@nyu.edu. This paper is based on my master's thesis at the University of Tokyo. I am extremely grateful to my advisor Shintaro Yamaguchi for his guidance and support throughout this project. I appreciate two anonymous referees and editors of International Tax and Public Finance, especially David R. Agrawal, for their critical comments that improve this paper substantially. I thank Kosuke Aoki, Kei Ikegami, Sagiri Kitao, Daiji Kawaguchi, Andrew Griffin, Daniel Marszalec, Shiko Maruyama, Tomoyuki Nakajima, Koki Okumura for their helpful comments. All remaining errors are mine. Data used in this article can be obtained from Research Institute of Economy, Trade and Industry. The author has no relevant or material financial interests that relate to the research described in this paper.

reforms, both increasing life expectancies and low fertility rates still make fiscal sustainability a serious problem. In fact, social security reforms are currently on the agenda for many OECD countries (OECD 2016).

Social security reform includes increasing the retirement age, cutting pension benefits, and reforming the tax rates of the earnings test. Increasing the retirement age and cutting pension benefits effectively alleviate fiscal pressure. However, people receive fewer pension benefits as a result, so their overall welfare declines. These policies are often unpopular among voters (Börsch-Supan et al., 2018a).

In contrast, reforming the tax rates of the earnings test could improve welfare. Under the earnings test, Social Security benefits for workers who have claimed pension benefits are reduced at a high rate once earnings pass a "test" threshold amount (Song and Manchester 2007). One of the purposes of the earnings test is to reduce pension benefits based on labor income. As a result, the earnings test imposes an implicit income tax on the elderly by reducing their pension benefits. More efficient tax rates for the earnings test would improve social welfare by changing the labor supply. Therefore, although there are many policies for reducing fiscal pressure, this paper focuses specifically on the earnings test . My research has found that reforming the tax rates of the current earnings test would increase social welfare without additional government pension expenditure.

Reforming the earnings test would increase social welfare because the current earnings test system is inefficient. The earnings test exists in many countries; however, the current design is inefficient in Japan and the U.S[1]. There are two main issues. First, the tax rates of the earnings test are 0%, 50%, or 100%; this implies that when earnings are increased by 1 yen, pension benefits are deducted by 0 yen, 0.5 yen, or 1 yen depending on an individual's earnings. As a result, the marginal tax rates jump disproportionately at the threshold. Individuals just below the threshold thus avoid working more. If these disproportionate jumps disappear, productive people earn and consume more, which improves welfare. The other problem of the earnings test is that the threshold where marginal tax rates jump from 0% to 50% is much higher for people over 65 than for people aged 60-64. The current age-dependent system, under which people aged 60-64 are more likely to face higher tax rates, encourages people work more after 65. On average, older people are less productive, so the current system encourages working when less productive. This is the other

[1]In the U.S, the marginal tax rates are 0%, 33%, or 50%. There are disproportionate jumps at the threshold.

source of inefficiency.

Focusing on the Japanese case, this paper aims to derive an optimal design for the earnings test in an optimal income tax framework (Mirrlees 1971). Since the earnings test implicitly imposes an income tax on elderly people through the pension system, this paper applies the optimal tax framework to the earnings test. In the optimal income tax framework, the government chooses the design of the earnings test to maximize social welfare, which is the weighted sum of the utility of individuals. This counterfactual earnings test would be subject to the constraint that its pension expenditure not exceed the expenditure under the current earnings test.

To model labor supply and retirement behavior, I construct a life cycle model with the current earnings test based on French (2005). Individuals in the model choose their consumption, hours worked, labor force participation decisions, and savings between the ages of 50 to 79. They face health, productivity, and survival uncertainties when making decisions. Since individuals can save and decumulate assets, they supply more labor when their wages are high and retire when wages are low.

This paper estimates preference parameters using the method of simulated moments. I match life cycle profiles from Japanese data (JSTAR) to the profiles generated by the life cycle model. Matched moments include labor force participation rates, hours worked, earnings, and asset profiles. The profiles identify key structural parameters, such as Frisch elasticity and the coefficient of relative risk aversion. These simulated profiles can capture important data features such as a sharp decline in labor force participation after 55, given the current earnings test scheme.

To analyze the inefficiency of the current earnings test, I compute social welfare under the counterfactual earnings test with a linear tax rate. I find that social welfare under the counterfactual system with a linear tax rate is substantially higher than under the current earnings test. The consumption equivalence turns out to be 0.72%. The source of welfare gain is the smoothed profile of the labor supply. Under the current earnings test, tax rates disproportionately drop at age 65. While the high marginal tax rates make individuals aged 60-64 work less, they would work more after age 65. This labor supply profile reduces their welfare because disutility from labor is a convex function. In contrast, under the counterfactual earnings test with age-independent tax rates, the labor supply profile would be smooth.

Second, I apply the optimal income tax theory (Saez 2002) to the earnings test. Saez (2002) shows that when the elasticity of an extensive margin relative to an intensive margin is large, the negative marginal tax rates on low-earnings are optimal. Since data shows that elderly people make retirement decisions, their labor supply elasticity of an extensive margin relative to an intensive margin is large. In my estimation, the matched profile of the labor supply identifies the elasticity of an extensive margin. I found that the welfare-maximizing earnings test has the negative marginal tax rates on low earnings. The intuition for why the negative marginal tax rates are optimal is as follows. When the elasticity of an extensive margin relative to an intensive margin is large, the negative marginal tax rates on low-earnings do not make high-earning people reduce their hours worked because it is an intensive change. A small decrease in labor supply leads to a small decrease in tax revenue. The decrease in tax revenue is lower than the marginal social value of reducing tax rates on low-earning people, so the negative marginal tax rates improve social welfare.

This paper also analyzes other social security reforms that would affect the labor supply of the elderly. The analysis suggests two major policy outcomes. First, I simulate the effect of a repeal of the earnings test on labor supply. The Japanese government currently plans to abolish the earnings test to encourage the labor supply among the elderly[2]. I find that the labor force participation rate under the counterfactual earnings test with the negative marginal tax rates is higher than when the earnings test is just abolished. Second, I compare the effect of reducing pension benefits with the effects of raising the pension eligibility age on welfare. Both of these policies aim to mitigate government fiscal pressure, but this research finds that raising the pension eligibility age is better in terms of improving social welfare.

This paper is related to several strands of literature. The life cycle model in this analysis is based on French (2005) and French and Jones (2011). Both papers estimate a life cycle model of labor supply and retirement using a method of simulated moments. French (2005) simulate the effect of the earnings test's repeal on labor supply, but they do not conduct welfare analysis. In contrast, this paper analyzes the effect of social security reform on social welfare.

Few papers focusing on the earnings test analyze social welfare. Instead, they focus on estimating the effect of the earnings test on labor supply by using natural experiments. For instance, Friedberg (2000) finds a substantial bunching of elderly people around the threshold where the marginal tax rate changes. Studying the earnings test's repeal in the United States in 2000, Song and Manchester

[2]"Basic Policy on Economic and Fiscal Management and Reform 2019".

(2007) find significant labor supply responses. A similar result appears in the case of abolition in the United Kingdom (Disney and Smith 2002). The approach of these studies is the reduced form, and their analysis does not examine the inefficiency of the earnings test in terms of social welfare.

Responses of labor force participation to income tax are usually studied in the context of low-income families. Blundell et al. (2009) examines the optimality of in-work credit for single mothers and finds that it is not optimal. Income tax on low-income families has been intensively studied, but fewer studies have explicitly analyzed the earnings test's optimal design for the elderly.

This paper is also related to literature about means-tested social insurance programs for retirees such as Medicaid and Supplemental Security Income. Braun et. al. (2017) find that welfare gains from these programs are large because they works as an insurance against risks for those who cannot easily self-insure by re-entering the labor market. This paper finds that the optimal earnings test encourages agents to re-enter the labor market and accumulate savings, which makes it easier for agents to self-insure against medical or spousal death event. The optimal earnings test is related to the means-test in that both work as insurance against risks.

The remainder of the paper proceeds as follows. Section 2 describes the optimal income tax theory. In section 3, I describe the life cycle model which will be estimated. Section 4 explains the estimation strategy. Section 5 describes the data, and section 6 gives the estimation results. Section 7 explains the framework of the optimal earnings test problem, while section 8 gives counterfactual experiments.

2 Optimal Income Tax Theory

Saez (2002) shows that the negative marginal tax rates on low earnings are optimal when the elasticity of an extensive margin relative to an intensive margin is large. In this section, I briefly describe the result and its intuition in Saez (2002).

The general framework of Saez (2002) is described as follows. The government chooses tax on earnings to maximize social welfare such that it satisfies the government budget constraint. While people are heterogeneous in productivity, the government cannot observe this heterogeneity. Social welfare is the weighted sum of individuals' utility. Saez (2002) shows that when the elasticity of an extensive margin is zero, the marginal tax rates are nonnegative everywhere. However, the larger the elasticity of an extensive margin relative to an intensive margin, the more likely the marginal tax rates on low-earnings are negative.

The intuition is as follows. Suppose the government marginally increases transfer for a lower-skill occupation. The government cares for the lower-skill occupation, so the government accepts the policy even if it reduces tax revenue. Change in tax revenue is caused by an intensive and extensive labor supply response. An intensive labor supply response has an ambiguous effect on tax revenue, but it is negligible if the intensive margin elasticity is small. On the other hand, when tax on the unemployed is higher than tax on lower skill occupation, namely marginal tax rate is negative , an extensive labor supply response reduces tax revenue. The government accepts this tax decrease. Therefore, if the elasticity of an extensive margin relative to an intensive margin is large, the optimal marginal tax rate is negative.

Since the elderly typically make retirement decisions, their elasticity of an extensive margin is large. However, it is a quantitative question of whether the extensive margin elasticity for elderly people is sufficiently large to make the negative marginal tax rates negative. In the following section, I construct the life cycle model, which computes social welfare. I estimate the elasticity of an extensive margin for older people using the method of moments in section 6. Quantitative exercise in section 8.1 confirms that the welfare-maximizing earnings test has the negative marginal tax rates .

3 Life Cycle Model

3.1 Setup

In this section, I construct a life cycle model. The model has two objectives. First, the model calculates social welfare given a certain design of the earnings test. Since social welfare is the weighted sum of the lifetime utility of individuals, the life cycle model is used to compute their lifetime utility. The second goal is to estimate the elasticity of an extensive margin with the life cycle model. Section 2 shows that the elasticity of an extensive margin determines the design of the earnings test. I structurally estimate the elasticity of an extensive margin with the life cycle model.

The design of the earnings test affects retirement behavior, which is a lifetime decision. I derive the welfare-improving earnings test in a dynamic environment. French (2005) is the basis on which the life cycle model is developed. Individuals decide the lifetime profiles of consumption, saving, and labor supply, facing survival, health, and productivity uncertainty. They enter the market at 50, and the remaining surviving individuals exit the economy at 79. By assumption, they enter the market at 50 because of data constraints. I discuss the justification of this assumption in online Appendix G.2.

Individuals maximize their expected lifetime utility. The utility in the current period comes from C_t consumption, H_t hours worked, M_t health status, and L time endowment. The within-period utility function is

$$U(C_t, H_t, M_t) = \frac{1}{1-\nu}\big(C_t^{\gamma}(L - H_t - \theta P_t\{H_t > 0\} - \phi I\{M_t = \text{bad}\})^{1-\gamma}\big)^{1-\nu}, \tag{1}$$

where θ is the fixed cost of working, and P_t is the indicator of labor force participation. The fixed cost of working is the cost that all workers need to pay, such as commuting time. The extensive margin of labor supply is created by the fixed cost of working. The fixed cost of working makes it difficult for individuals to work for only a few hours. Individuals who work for a few hours obtain a slight amount of earnings, but they lose the fixed size of leisure time. By assumption, the unemployed can reenter the market without any cost. The reason why elderly people are more likely to retire is wage. As they get old, they face lower wages and therefore are more likely to retire.

In equation (1), ϕ is the cost of poor health, and M_t is the health indicator, which takes good or bad. $I\{M_t = \text{bad}\}$ is a health indicator that takes 1 when an individual's health status is bad. M_t follows a Markov process. In the utility function, γ represents the weight of consumption, while ν is the coefficient of risk aversion. This functional form is used in most general equilibrium studies of social security reform (İmrohoroğlu and Kitao, 2009).

Lifetime utility is

$$U(C_t, H_t, M_t) + E_t\left[\sum_{j=t+1}^{T} \beta^j S(j-1,t)\big(s_j U(C_j, H_j, M_j) + (1-s_j)b(A_j)\big)\right], \tag{2}$$

where A_t is asset, and β is the time discount factor. Let s_j be the probability of being alive at j conditional on being alive at $j-1$, and $S(j,t) = (1/s_t)\Pi_{k=t}^{j} s_k$ denote the survival rate of being alive at age j conditional on living at t. Since age T is the terminal period, s_T is zero. Upon their death, they receive utility from bequest

$$b(A_t) = \theta_B \frac{(A_t + K)^{(1-\nu)\gamma}}{1-\nu}, \tag{3}$$

where K determines the curvature of bequest utility.

The individual's budget constraint is

$$A_{t+1} = (1+r)A_t + w_t H_t + Pb_t - C_t, \tag{4}$$

where wages are w_t, and pension after deduction by the earnings test is Pb_t. The earnings test is included in Pb_t. People start to receive pension benefits after 60[3]. I discuss the detailed part of Pb_t in section 3.2. After-tax wages in data are used to generate w_t, so the income tax is implicitly included in the model.

Wage at time t is exogenous. Wage is a function of an age-dependent part, $W(t)$, and an autoregressive component, AR_t.

$$\ln w_t = W(t) + AR_t. \tag{5}$$

The autoregressive component of wage has a correlation ρ and a normally distributed innovation η_t.

$$AR_t = \rho AR_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2). \tag{6}$$

In this model, people are heterogenous in productivity, which is represented as η_t.

Individuals maximize lifetime utility, subject to asset accumulation equation (4) and wage determination equation (5) and (6).

3.2 Earnings Test

The earnings test is included as Pb_t in the individual's budget constraint. Their past earnings decide the pension benefit before deduction by the earnings test. A detailed way to compute pension benefits before a deduction is described in Appendix A.2. Pb_t is the pension benefit after deduction by the earnings test. Earnings and pension benefits decide the amount of deduction. The life cycle model fully incorporates the rules of the current earnings test in Japan.

[3]This model is based on the pension system in 2013, so I assume that all people start to receive pension benefit after 60.

According to the concept of the earnings test, elderly people with more earnings receive fewer pension benefits. People above 60 years old are affected by the earnings test. The rule of the earnings test changes at 65. The detailed rule of the current earnings test is fully described in Appendix A.1. Importantly, the marginal tax rates jump up disproportionately at the threshold; the tax rates are 0%, 50%, or 100%. This implies that when labor income increases by 1 yen, the sum of labor income and pension benefit deducted by the earnings test increases by 0 yen, 0.5 yen, or 1 yen.

Figure 1 shows the example. The horizontal axis is labor income. The vertical axis is pension benefit deducted by the earnings test. It shows two pension benefits under two systems: the system that people aged between 60 and 64 face and the system that people aged above 65 face. Low-income people receive the full amount of pension benefits. The marginal tax rate is 0%. When the sum of pension and labor income is higher than a certain threshold, an additional 1 yen of earnings leads to a deduction of 0.5 yen. The marginal tax rate is 50%. When the total income exceeds a certain threshold, no pension benefit to be deducted is left. Notably, the threshold where the marginal tax rate changes for people aged 60-64 is lower than that for people above 65. This pension system discourages individuals aged 60-64 work more and encourages work more after 65. Fluctuated labor supply is inefficient because the disutility of labor supply is convex.

3.3 Heterogeneity and Model Solution

Individuals decide consumption and leisure based on state variables, preference parameters, and parameters determining exogenous variables. The heterogeneous variables in this model are health condition and productivity. The value function solves

$$\begin{aligned} V_t(X_t) &= \max_{C_t, H_t}\Big\{\frac{1}{1-\nu}\left(C_t^{\gamma}(L - H_t - \theta P_t - \phi I\{M_t = \text{bad}\})^{1-\gamma}\right)^{1-\nu} \\ &+ \beta s_{t+1} \sum_{M_{t+1}\in\{\text{good, bad.}\}} \sum_{W_{t+1}} V_{t+1}(X_{t+1})\text{prob}(M_{t+1}|M_t, \text{age})\text{prob}(w_{t+1}|w_t, \text{age}) \\ &+ \beta(1 - s_{t+1})b(A_{t+1})\Big\}. \end{aligned} \tag{7}$$

State variables are denoted as $X_t = (A_t, w_t, M_t)$. Since the model has no closed form solution, the decision rules must be found numerically. The decision rules are solved by backward induction starting at time T. I discretize assets and labor supply decision space, and the value function is

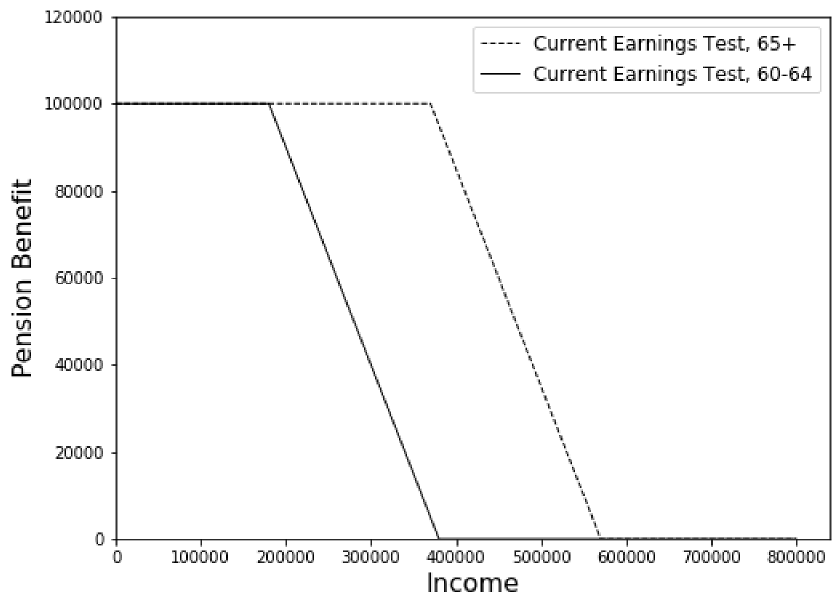


Figure 1: The Design of Earnings Test

This figure shows how much pension benefit people receive given the amount of monthly income. The marginal tax rate is 0% if income is lower than the threshold. After income exceeds it, the marginal tax rate is 50%. It is assumed that an individual receives 100,000 yen ($\approx$ 1000 USD) of pension per month.

calculated at each point. I search over the grids to find the optimal assets and labor supply level. The Tauchen method is used to discretize heterogeneity in wage into five grids.

By assumption, counterfactual policies do not change the distribution of exogenous variables, which include health, productivity, age, and assets at the initial period. It is reasonable that policies will not affect health and productivity. In addition, counterfactual policies do not change the distribution of assets in the initial period. This assumption is justified if the government suddenly enacts counterfactual policies without announcing them beforehand. The individuals cannot adjust their assets because the counterfactual policies are suddenly enacted.

4 Estimation Strategy

This section describes the method of simulated moments (MSM) estimation strategy. The estimation method is based on French (2005). It aims to estimate the preference parameters given the data generating process of the exogenous variables. Since it is difficult to estimate all parameters simultaneously, this paper uses a two-step strategy. First, I estimate some parameters of exogenous state variables and calibrate others. By assumption, individuals know their state variables and the Markov process and maximize their utility. In the second step, I estimate the preference parameters that creates simulated profiles that match the profiles from data. To simulate the life cycle profiles for hypothetical individuals, I use the numerical methods in section 3.3 and parameters from the first step. The next subsection describes the second step in detail.

4.1 Estimation of Preferences: The Method of Simulated Moments

The objective is to find preference parameters simulating profiles that are close to the profiles from data. The method of simulated moments (MSM) estimation strategy is described as follows. First, I estimate the life cycle profile of the labor force participation rate, hours worked, and savings from data (JSTAR). Second, I estimate the data generating process to simulate matrices for health and productivity shocks, including an initial distribution for health, wage, and assets. These are sequences of 20,000 simulated individuals. They live for T periods, so I obtain a 20,000$\times$T matrix of health and productivity. Third, I pick an arbitrary vector of preference parameters and compute decision rules with the methods described in section 3.3. Fourth, I simulate a hypothetical life cycle profile of decision variables. The fifth step is to calculate the difference between data profiles and simulation profiles. Finally, a new vector of parameters is chosen to minimize the distance, and the whole procedure is repeated. Estimated parameters are the parameters that minimize the distance between the data and the simulation. I discuss the distribution of the parameter estimates and the weighting matrix in online Appendix C.

Matching moments include the distribution of hours worked, the distribution of earnings, and the life cycle profile of savings. The choice of matching moments follows French (2005) and Blundell and Shephard (2012)[4]. Online Appendix D explains how parameters are identified by the

[4] In French (2005), matching moments are the life cycle profile of savings, hours worked, and labor force partic-

matching moments.

The hours worked distribution is categorized into three groups, 0, 1–36, and 37+ hours. The shares of three categories for people above 60 years old are matched. The second moment is the distribution of earnings. The distribution of earnings determines the level of pension benefits through the earnings test. The pension expenditure under the current earnings test determines the government budget constraint, so the distribution of earnings is critical to measure the current earnings test's inefficiency. The 20th, 40th, 60th, and 80th percentiles of labor income for people above 60 are matched. Let $j \in \{20, 40, 60, 80\}$ index asset quantiles. The π_j the earnings quantile, $Q_{\pi_j}(w_{it}H_{it})$, is defined as

$$Pr(w_i H_i \leq Q_{\pi_j}(w_i H_i)) = \pi_j. \tag{8}$$

This implies that the fraction of individuals with less than $Q_{\pi_j}(w_{it}H_{it})$ in earnings is π_j. As is well known (Powell(1994)), above equation can be rewritten as a moment condition:

$$E[1\{w_i H_i \leq Q_{\pi_j}(w_i H_i)\}] = \pi_j. \tag{9}$$

Finally, the mean of life cycle profile of saving is matched for all ages.

Moment conditions are

$$E[1\{H_i = 0\}] - \int 1\{H = 0\}(X, \theta, \chi) dF(X) = 0 \tag{10}$$

$$E[1\{0 < H_i \leq 36\}] - \int 1\{0 < H \leq 36\}(X, \theta, \chi) dF(X) = 0 \tag{11}$$

$$E[1\{36 < H_i\}] - \int 1\{36 < H\}(X, \theta, \chi) dF(X) = 0 \tag{12}$$

$$E[1\{w_i H_i \leq g_{\pi_j}(\theta_0, \chi_0)\} - \pi_j] = 0 \quad \forall \pi_j \in \{0.2, 0.4, 0.6, 0.8\} \tag{13}$$

$$E[A_{it}|t] - \int A_t(X, \theta, \chi) dF_{t-1}(X|t) = 0 \quad \forall t \in \{1, \cdots, T\}, \tag{14}$$

where X is state variables, θ is preference parameters, χ is the Markov process that determines state variables, and $F_t(X)$ is the cdf of the state variables at time t. There are T+7 moment conditions. Equations (10), (11), and (12) are the moments of distribution of hours worked. $1\{\cdot\}$ is

ipation rate. The life cycle profile of savings is included in this paper. The life cycle profile of hours worked and labor force participation rate are not directly included in the matching moments, but they are used as an out of sample validation.

the indicator function. Equation (13) is the moment of earnings. $g_{\pi_j}(\theta_0, \chi_0)$ is the π_j the quantile of the simulated earnings distribution. $g_{\pi_j}(\theta_0, \chi_0)$ is the model analog to $Q_{\pi_j}(w_{it}H_{it})$ and equation (14) is the moment of mean assets profile. The integrals are computed using a Monte Carlo integration. I discuss how the matching moments identify the preference parameters in section 6.3.

5 Data

5.1 Data

I use Japanese panel data named 'Japanese Study of Aging and Retirement' (JSTAR) for 2007, 2009, 2011, and 2013. JSTAR surveys the Japanese elderly to compare it to other data, so its design have many things in common with that of HRS in the U.S. JSTAR covers a wide range of information, including social, economic, and health conditions for people aged above 50. I drop people aged above 80. The sample includes salaried job workers, the self-employed, and both sexes.

The project of JSTAR started in 2007. The samples are collected from five municipalities in Japan. JSTAR conducts stratified random sampling within each municipality. City-level representative data have been conducted every two years, and new samples from other cities were added in 2009 and 2011. Table 1 represents the summary statistics of JSTAR. Ichimura et al. (2009) provide a detailed description of the JSTAR's design and sample methodology.

5.2 Profile of Hours Worked in Data

Based on the optimal income tax theory in section 2, the larger the elasticity of an extensive margin relative to an intensive margin is, the more likely it is that the negative marginal tax rates on low-earnings are optimal. This section shows that the elderly have a large elasticity of an extensive margin relative to an intensive margin.

An extensive margin elasticity is defined as the change in the aggregate labor force participation rate when the wage changes. When the change in labor force participation rate is large relative to the change in hours worked, the elasticity of an extensive margin relative to an intensive margin is large.

Table 1: Summary Statistics of Each Wave

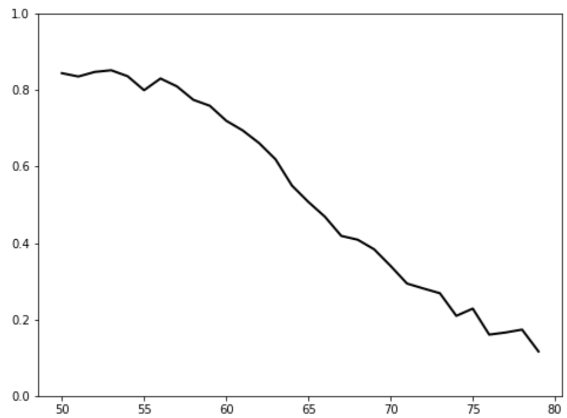
Year	2007		2009		2011		2013	
Variable	Value	N	Value	N	Value	N	Value	N
Average age	62.8	3739	64.5	4188	65.6	4643	67.2	3930
LFP Rate in %	56.8	3704	51.0	4072	50.9	4535	48.6	3797
Mean Hours Worked	41.1	1606	39.6	1595	37.2	1791	36.7	1478
20 Percentile of earnings	137	808	120	422	130	810	120	638
40 Percentile of earnings	220	808	200	422	200	810	200	638
60 Percentile of earnings	350	808	300	422	300	810	281	638
80 Percentile of earnings	450	808	400	422	421	810	400	638
Mean Savings	1044	684	1009	789	857	364	739	396
Good Health in %	80.4	3739	80.8	4188	83.9	4643	84.4	3930

1. Source: "JSTAR". Samples include both sex and self-employed. Age is 50-79.

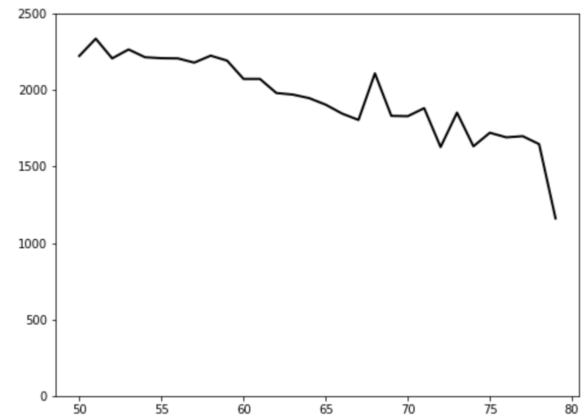
2. LFP stands for labor force participation. Mean hours worked are weekly. 'Earnings' in the table is monthly earnings. 'N' of the each percentile of earnings is the sample size of earnings. Earnings and savings are in 1000 yen ($\approx$ 10 USD). Good health is based on self-reported health.

Figure 2 shows the life cycle profile of the labor supply in the data. Figure 2-(a) shows the labor force participation rate. As they age, fewer people participate in the labor market. Figure 2-(b) shows the mean annual hours worked, excluding zero hours worked. The labor force participation rate approaches zero in (a), which implies that individuals respond along an extensive margin. In contrast, the hours worked in 2-(b) are relatively constant over the life cycle. Those who participate in the labor market annually work for more than 1500 hours on average, leading to a smaller intensive change.

I check whether individuals directly respond along an extensive margin with panel data. Table 2 shows the transition of hours worked. 14% of people who work more than 30 hours per week are unemployed in the next wave of data. 3% of them work between 1 and 15 hours per week in the next wave. Direct transition from full-time to zero hours worked is more common. This implies



(a) Labor Force Paricipation Rate



(b) Mean Hours Worked Excluding Zero

Figure 2: Life Cycle Profile of Labor Force Participation Rate and Hours Worked

Source: JSTAR. The horizontal axis of (a)-(b) is age. (a) is labor force participation rate. (b) is the annual mean hours worked conditional on age, but people who do not work are excluded when computing the average. The sample includes people aged between 50 and 79, both sexes, and the self-employed.

Table 2: The Transition of Hours Worked

	Hours worked in next wave			
Hours worked	0	(0,15]	(15,30]	>30
0	0.97	0.01	0.01	0.01
(0,15]	0.32	0.46	0.11	0.11
(15,30]	0.25	0.20	0.41	0.13
>30	0.14	0.03	0.08	0.73

Source: JSTAR. Samples include both sex and self-employed. People who are aged between 60 and 70 in JSTAR are used.

The table represents how people change their hours worked in the next wave. The survey is conducted once per two years.

that the elasticity of an extensive margin relative to an intensive margin is large.

The fixed cost of working creates the large elasticity of an extensive margin relative to an intensive margin. When the fixed cost of working is large, it is suboptimal to work for a few hours. When the fixed cost of working is small, more individuals work for a few hours and there are fewer unemployed people. The fixed cost of working determines an extensive margin's elasticity relative to an intensive margin, which is crucial in this paper. I estimate the fixed cost of working by the method of simulated moments in section 6.3.

6 Estimation Results

The MSM strategy has two steps. In the first step, I estimate some data-generating processes for exogenous state variables. I calibrate other elements. The data generating process for exogenous state variables includes the profile of wage, health transition matrices, the distribution of savings in the initial period, and survival probabilities. In section 6.1 and 6.2, some of them are estimated, and others are calibrated. In the second step, I estimate preference parameters. The preference parameters that simulate the decision variables are estimated in section 6.3. The decision variables are the profile of hours worked, the level of earnings, and the labor force participation rate.

6.1 Wage

This section estimates the life cycle profiles of wages to be fed into equation (5). The procedure is based on French (2005). The monthly earnings divided by hours in JSTAR is used. The wage at age t, w_t, is given by

$$\ln w_t = W(t) + AR_t. \tag{15}$$

Idiosyncratic shock is included in AR_t. $W(t)$ is estimated as follows.

$$W_i(t) = f_i + \beta_1 age_{it} + \beta_2 age_{it}^2 + \Pi_d I\{60 \leq age_{it}\} + \Pi_U U_t + u_{it}, \tag{16}$$

where f_i is an individual-specific effect, age is an age, $I\{60 \leq age_{it}\}$ is a dummy variable, and U_t is the unemployment rate. I am concerned about the individual-specific effects and year effects. The dummy variable, $I\{60 \leq age_{it}\}$, is included to capture the mandatory retirement system in Japan. Full-time workers are forced to switch their jobs to part-time work at 60 in Japan. Wage profiles discontinuously drop at 60 years old. To capture this effect, I include the dummy variable. I estimate β_1, β_2, Π_d, and Π_U[5].

When generating $W(t)$, I set f_i as a mean individual-specific fixed effect and U_t as an average unemployment rate over the sample period. Figure 3 shows the profile of $W(t)$. The striking feature is the downward slope of the wage profiles. It shows a large decline between 50 and 65. Wage discontinuously drops at 60 because of a mandatory retirement system. After 65, the profile of wages is relatively flat. Fixed-effects estimates imply a more rapid drop in wages than OLS estimates do (Heckman 1976). The reason for this bias is that highly productive people tend to remain in the labor market. While the OLS estimation produces composition bias, fixed-effects estimates capture this bias.

The autoregressive component of wages, ρ and σ_η in equation (15), are estimated with minimum distance techniques[6]. Table 3 shows the estimates.

[5]The estimated coefficient of $I\{60 \leq age\}$ is -0.13 (0.068). The coefficient is statistically significant. Estimated values of β_1 and β_2 are -0.12 (0.08) and 0.0008 (0.0007). Values in parentheses are the standard error.

[6]From equation (15), $AR_t = \ln w_t - W(t)$. I have $\ln w_t$ in data. $W(t)$ is computed from equation (16). I derive AR_t, and AR_{t-1}. Since AR_t follows $AR_t = \rho AR_{t-1} + \eta_t$, ρ and σ_η are estimated by OLS.

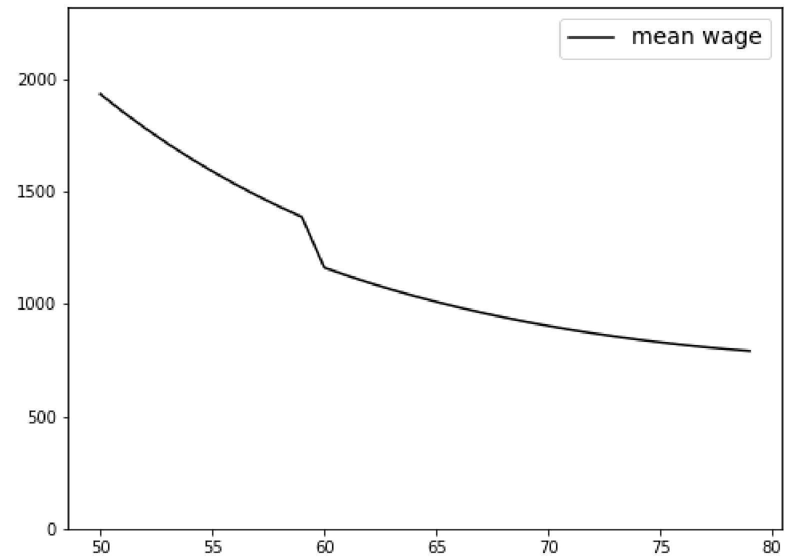


Figure 3: Life Cycle Profile of Mean Wage.

The vertical axis is mean hourly wage in yen. The horizontal axis is age. Fixed-effect estimation. 2000 yen ≈ 20 USD.

6.2 Remaining Calibrations

I calibrate the transition matrix of health condition based on self-rated health status. Health status takes on the status of 'good' or 'bad' in my model. Health status follows a Markov process. The health transition matrices are separately calibrated in three age groups; 50–59, 60–69, and 70-79. Estimation shows that as people age, they are more likely to be unhealthy. The online appendix B describes the details.

The wealth distribution at 50 years old is estimated. In my model, people enter the market at

Table 3: Variance and Persistance of Wages

Parameter	Variable	Estimate
Standard error of shock in wages	σ_η	0.33
Autoregressive coefficient of wages	ρ	0.57

50 years old with some wealth. Non-parametric estimation is adopted to estimate the distribution of wealth[7]. By assumption, initial wealth is dependent on health and productivity in my model. People enter the market with their health condition and their productivity. There are two groups of health (good and bad) and three groups of productivity (low, middle, and high). If the initial health is bad, their initial wealth is randomly drawn from the bad health distribution. If their initial health is good, their productivity decides their initial wealth. There are four distributions of initial wealth; bad health, good health and high productivity, good health and middle productivity, and good health and low productivity.

The survival probability (s_t) is calibrated from the "Simple Life Expectancy Table" generated by the Ministry of Health, Labor, and Welfare in Japan in 2017. Online Appendix B describes the details. The survival probability depends on age, but not on health status. The interest rate (r) is 0.01. The discount factor (β) is 0.98. The curvature of the bequest motive (K) is 1,500,000 yen ($\approx$ 15,000 USD).

6.3 Preference Parameters Estimates

The preference parameters to be estimated include the coefficient of risk aversion (η), consumption weight (γ), time endowment (L), the fixed cost of working (θ), the fixed cost of bad health condition (ϕ), and the parameter of bequest (θ_b). Table 4 gives the results. Values in parentheses are standard errors. The second and third columns show the estimation results in French (2005), using the U.S. data. I use the same functional form and estimation method, so they are comparable. Frisch elasticity and the coefficient of relative risk aversion in my estimation results are similar to those in French (2005).

A willingness to intertemporally substitute their hours worked is the key parameter in this research. The Frisch elasticity of the labor supply is 0.86 at 55 years old and 1.15 at 60 years old[8]. The elasticity increases with age. Kuroda and Yamamoto (2008), who estimate the Frisch elasticity with Japanese data, show that Frisch elasticity is in the range of 0.2-0.7 for males, 1.3-

[7] To compensate for the small sample size, people aged between 50 and 52 are used to estimate the wealth distribution at the initial period.

[8] Assuming certainty and the interior condition, the Frisch elasticity of labor is $-\frac{L-H_t-\theta}{H_t} \times \frac{\gamma(1-\nu)-1}{\nu}$. The elasticity of labor depends on the ratio of leisure to labor, so it is age-dependent.

Table 4: Preference Parameter Estimates

Parameters	My Results	French (2005)	
		Specification 1	Specification 2
ν: Coefficient of risk aversion, utility	2.9 (0.035)	3.34 (0.07)	3.19 (0.05)
γ : Consumption weight	0.49 (0.0043)	0.578 (0.003)	0.533 (0.003)
L : Time endowment	5700 (22)	4466 (30)	3900 (24)
θ : Fixed cost of working	1000 (7)	1313 (14)	335 (7)
ϕ : Fixed cost of bad health	500 (11)	318 (14)	196 (8)
θ_b : Parameters of bequest	0.01 (0.006)	1.69 (0.05)	1.70 (0.04)
Frisch elasticity, age 55	0.86	–	–
Frisch elasticity, age 60	1.15	1.24	1.10
Coefficient of relative risk aversion	1.93	2.35	2.17

1. Methods of simulated moments estimates. Values in parentheses are the standard error.

2. The first column shows my estimatin results. The second and third columns show the results in French (2005) under different specifications. Results in the U.S.

3. Frisch elasticity at time t is given by $-\frac{L-H_t-\theta}{H_t} \times \frac{\gamma(1-\nu)-1}{\nu}$. When computing Frisch elasticity at 60, H_t is average hours worked at 60. Frisch elasticity at 55 is not listed in French (2005).

4. The coefficient of relative risk aversion is $-C_t \times \frac{\partial^2 U_t/\partial^2 C_t}{\partial U_t/\partial C_t} = -\left(\gamma(1-\nu)-1\right)$.

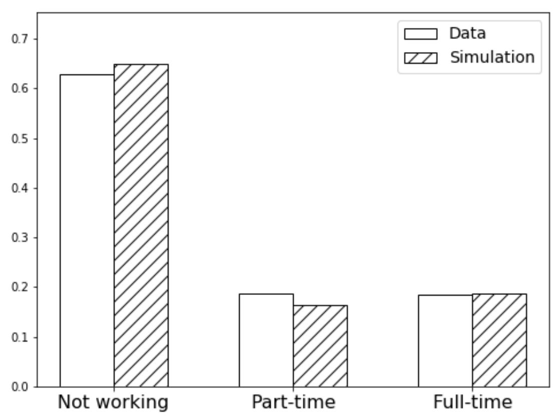
1.5 for females, and 0.7-1.0 for both sexes. Although Kuroda and Yamamoto (2008) estimate Frisch elasticity with Japanese aggregate data and different methods, my estimates from Japanese microdata are comparable with their paper estimates.[9]. The coefficient of relative risk aversion is 1.93[10]. The estimate is close to the previous estimates (Attanasio and Weber 1995).

Figure 4 shows the simulation and data of profiles for decision variables. The procedure to obtain them is described in online Appendix C. Figure 4-(a) represents the distribution of hours

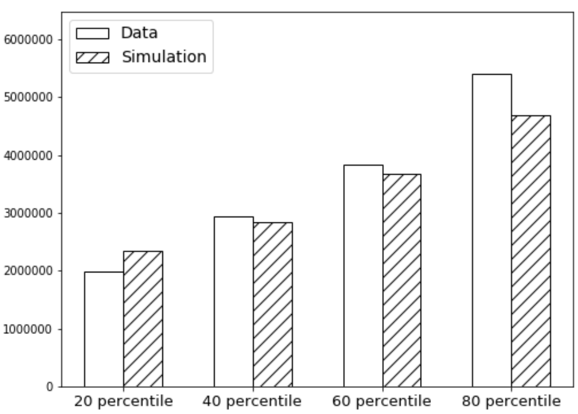
[9]The different estimation results come from the different method. My estimation results are based on the methods of simulated moments. Kuroda and Yamamoto (2008) is based on the method by Blundell and MaCurdy (1999) and MaCurdy (1981).

[10]The relative risk aversion is from $-C_t \times \frac{\partial^2 U_t/\partial^2 C_t}{\partial U_t/\partial C_t} = -\left(\gamma(1-\nu)-1\right)$.

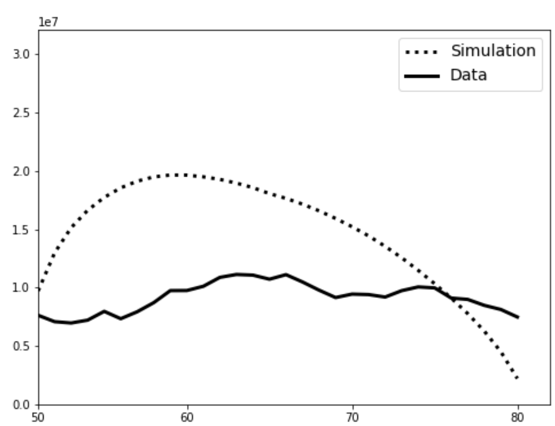
worked for people above 60. It shows the ratio of the unemployed, part-time workers, and full-time workers. The simulated distribution of hours worked has a high ratio of the unemployed. Figure 4-(b) portrays the distribution of earnings. It shows the 20th, 40th, 60th, and 80th percentile of earnings. It excludes zero earnings. Figure 4-(c) shows the life cycle profile of average savings. The simulated profiles peak in the 60s, and people aged 80 have enough savings for a bequest. Individuals in the model accumulate savings more than they do in data, which is often observed in life-cycle model (Imrohoroğlu and Kitao 2012, Gourinchas and Parker 2002, De Nardi, French, and Jones 2010).



(a) Distribution of Hours Worked



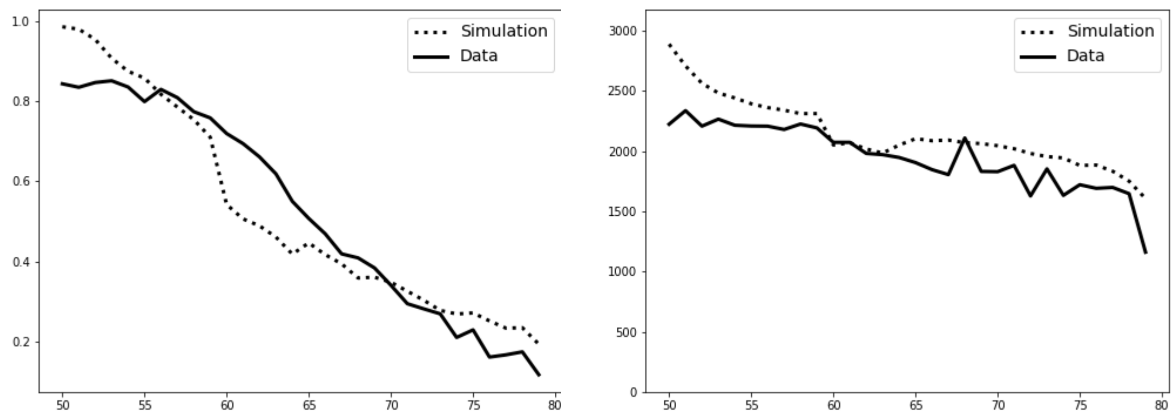
(b) Distribution of Earnings



(c) Profile of Mean Savings

Figure 4: Simulation and Data of Matching Moments[11].

Figure 5 shows the simulated and data profile of the labor supply. Targeted moments do not include the labor supply profiles, so they give a validation. Figure 5-(a) shows the labor force participation rate. Figure 5-(b) shows the mean of hours worked, excluding zero hours worked. In the figure 5-(a), labor force participation decreases as individuals get older, which is an extensive response. The simulated profile in figure 5-(b) shows a smaller intensive response. While the labor force participation rate approaches to zero, the profile of the mean hours worked is relatively flat. The simulated profiles capture the features of labor supply for the elderly.



(a) Labor Force Paricipation Rate (b) Mean Hours Worked Excluding Zero

Figure 5: Simulation and Data of Profile of Labor Supply.

This figure represents the profile of labor supply in simulation and data. The horizontal axis of (a)-(b) is age. (a) is labor force participation rate. (b) is the mean hours worked, excluding zero hours worked.

[11]The figures show the simulation and data of matching moments. (a) is the share of not working, part-time, and full-time for people aged above 60. Part-time workers are defined as people who work between 1 and 36 hours per week. Full-time workers are defined as people who work more than 36 hours per week. (b) is the 20th, 40th, 60th, and 80th percentile of earnings for people above 60. (c) is the profile of mean savings. The horizontal axis is age.

7 The Optimal Earnings Test Problem

The framework to derive the welfare-improving earnings test is as follows. The government chooses the earnings test to maximize the social welfare function. The counterfactual earnings test satisfies the government budget constraint. The government cannot observe the productivity of individuals but can observe their earnings and pension, so the government imposes taxes on them.

The social welfare function is defined as the weighted sum of utilities of individuals:

$$W(Ta) = \int_{\epsilon} \Gamma\Big(U\big(C^*(\epsilon, Ta), H^*(\epsilon, Ta)\big), \mu\Big) dF(\epsilon), \tag{17}$$

where Ta represents the pension system, which is the government choice variable, and ϵ represents the heterogeneity of individuals. $U(C^*(\epsilon, Ta), H^*(\epsilon, Ta))$ is defined as the lifetime utility given in equation (2). Function Γ represents the government's preference for equality, and it puts more weight on low-income people. Following the Blundell and Shephard (2011), Γ is defined as

$$\Gamma\Big(U\big(C^*(\epsilon, Ta), H^*(\epsilon, Ta)\big), \mu\Big) = \frac{\Big(\exp\Big(U\big(c^*(\epsilon, Ta), H^*(\epsilon, Ta)\big)\Big)\Big)^{\mu} - 1}{\mu}. \tag{18}$$

The preference for equality is represented as μ. The welfare-improving earnings test is derived under $\mu = -1$[12]. The way to compute the social welfare is explained in online Appendix C in detail.

The government budget constraint is

$$\int_{\epsilon} Ta(H^*, \epsilon) dF(\epsilon) \leq \bar{T}. \tag{19}$$

The left-hand side is pension expenditure from the counterfactual earnings test[13]. The right hand side is exogenously given. It is set to be pension expenditure under the current earnings test because

[12] The value of μ is close to the value in Blundell and Shephard (2011). $-\mu = -\Gamma''(U, \mu)/\Gamma'(U, \mu)$ so that $-\mu$ can be interpreted as the absolute inequality aversion.

[13] In this model, the government's expenditure is only pension benefit. The pension benefit which individuals receive is the full amount of pension benefit minus the deduction by the earnings test. The full amount of pension benefit is described in Appendix A.2. The earnings test's deduction amount depends on the full amount of pension benefit and earnings. Detail is described in Appendix A.1. The government budget constraint states that this expenditure should not exceed exogenously given $\bar{T}$. There is no source of income for the government.

this paper aims to measure the inefficiency of the current earnings test[14]

Next, the counterfactual earnings test is described. People above 60 years old face the same counterfactual design of the earnings test. The amount of pension benefit deducted by the counterfactual earnings test is defined as

$$Pb_t = \begin{cases} Pb_{b,t} & (H_t = 0) \\ \max\{0, Pb_{b,t} - \tau w_t H_t + tf\} & (H_t > 0), \end{cases} \tag{20}$$

where $Pb_{b,t}$ is the pension benefit before deduction, τ is a marginal tax rate, and tf is a lump-sum transfer for workers. $Pb_{b,t}$ is exogenously given for workers. Policy variables are a transfer (tf) and a linear tax rate (τ). The first variable is the lump-sum transfer, which workers can receive but the unemployed cannot receive. The transfer makes working attractive. The second policy variable is a linear tax rate. To satisfy the government budget constraint, a linear tax rate is set to be positive.

Importantly, the counterfactual pension system satisfies the property implied by the optimal income tax theory if the tf is positive. The optimal income tax theory described in section 2 implies that transfer for low-earnings people should be higher than transfer for the unemployed if the elasticity of an extensive margin is large. Based on the counterfactual pension system, low-earners receive more pension than the unemployed if the transfer is positive. In section 8.1, I confirm that the welfare-maximizing level of tf is positive.

[14] When the counterfactual earnings test is defined as (20), equation (19) can be rewritten as

$$\begin{aligned} & \text{Average pension before deduction} - \tau \times \text{Average labor income} + \text{ Transfer} \times \text{Labor force participation rate} \\ \leq \quad & \text{Average pension before deduction} - \text{Average deduction under the current earnings test} \end{aligned}$$

The left-hand side is the pension expenditure under the counterfactual earnings test, and the right-hand side is the pension expenditure under the current earnings test. In the later quantitative exercise, average annual pension expenditure before deduction is approximately 800,000 yen. Under the optimal earnings test derived in section 8.1, $\tau = 0.1$, and annual transfer is 200,000 yen. In the later simulation, $\tau \times$ Average labor income is 100,000 yen. Transfer $\times$ Labor force participation rate is 75,000 yen. Average pension deduction under the current earnings test is 25,000 yen. Therefore, equation (19) holds with equality.

8 Policy Experiments

In section 8.1, I derive the welfare-improving earnings test. First, I analyze how much the counterfactual earnings test raises social welfare compared to the current earnings test. I also confirm that the welfare-maximizing earnings test has negative marginal tax rates. Section 8.2 simulates the effect of the repeal of the earnings test on the labor force participation rate. Section 8.3 compares cutting pension benefits to raising the pension eligibility age in terms of social welfare.

The situation in which the government implements a counterfactual policy is as follows. The government implements the policy without an announcement beforehand. When people are 50 years old, a counterfactual policy is implemented. Before individuals are 50 years old, they make decisions under the current policy. They choose their decision variables under the counterfactual policy after 50 years old.

Table 5 gives accounting statistics for each experiment. The first row shows results under the current earnings test. The second row displays results under the optimal earnings test described in section 8.1. Importantly, the labor force participation rate is higher than any other result. The lump-sum transfer for workers encourages labor force participation. The third row shows results when the earnings test is abolished. The fourth row displays the result when the pension eligibility age is raised by one year. The bottom row shows that cutting pension benefits by 40,000 yen ($\approx$ 400USD)[15].

8.1 The Welfare Improving Earnings Test

I search for the welfare-maximizing transfer (tf) and a linear tax rate (τ) in equation (20). First, I confirm that the welfare-maximizing earnings test satisfies the implication of the optimal income tax theory. Second, I compute the counterfactual policy's consumption equivalence that makes social welfare under the current policy equal. I find that the consumption equivalence is substantial.

The procedure to find the optimal policy variables is as follows. First, the amount of transfer (tf) is fixed. The policy variable to be chosen is only a linear tax rate (τ).

The optimal linear tax (τ) rate is determined from the government budget constraint. The least

[15]Pension expenditure when raising pension eligibility age by 1 year is approximately equal to that when cutting pension benefits by 40,000 yen in my model.

Table 5: Policy Experiments

	LFP rate in %	Hours worked per year	PDV of mean income	PDV of mean consumption	Assets at age 62
Current earnings test	53.0	1213	6932	6933	1822
The optimal earnings test	55.1	1208	6934	6978	1920
The repeal of the earnings test	53.2	1226	7088	7033	1852
Raise pension eligibility age by 1 year	53.4	1224	6975	6923	1802
Cut pension benefit by 40,000yen	53.4	1225	6919	6919	1846

1. LFP stands for labor force participation. Labor force participation rate and hours worked are average for 50-79.

2. PDV stands for present discounted value. Consumption, income, and assets are measured in ten thousands yen ($\approx$ 100USD). Income includes labor income and pension benefit. The difference between PDV of mean income and consumption is bequest.

3. Pension expenditure when raising pension eligibility age by 1 year is the same as that when cutting pension benefits by 40,000 yen.

tax rate that satisfies the government budget constraint maximizes social welfare. A higher linear tax rate necessarily reduces social welfare because a higher tax rate reduces labor income, and excess tax revenue collected by the government is not consumed.. Second, I choose the amount of transfer. Given the transfer, the first step gives the optimal linear tax rate. I compute the social welfare given the transfer and a linear tax rate. Figure 6 gives a social welfare function. It shows that positive transfer maximizes social welfare.

I find that the earnings test with 200,000 yen ($\approx$ 2,000 USD) of transfer and 10.0% of a marginal tax rate maximize social welfare. I calculate the consumption equivalence, which makes the welfare under the current policy equal to the welfare under the counterfactual policy. It is 0.96%[16]. The welfare gain is substantial. The detailed decomposition of welfare gain into a linear tax rate and transfer is described in online Appendix F.

Figure 7 shows the current and counterfactual earnings test. It shows the pension benefit under three systems: the current system targeted to people above 65, the current system targeted to people between 60 and 64, and the counterfactual earnings test. It shows that the counterfactual

[16]For comparison, I list the example of welfare gain from policies. Krusell et al. (2009) show that the welfare gain from eliminating business cycles is 1% in consumption equivalence. Lucas (2003) shows that welfare gain from reducing the annual inflation rate from 10 to 0 percent is a perpetual consumption flow of 1 percent of income.

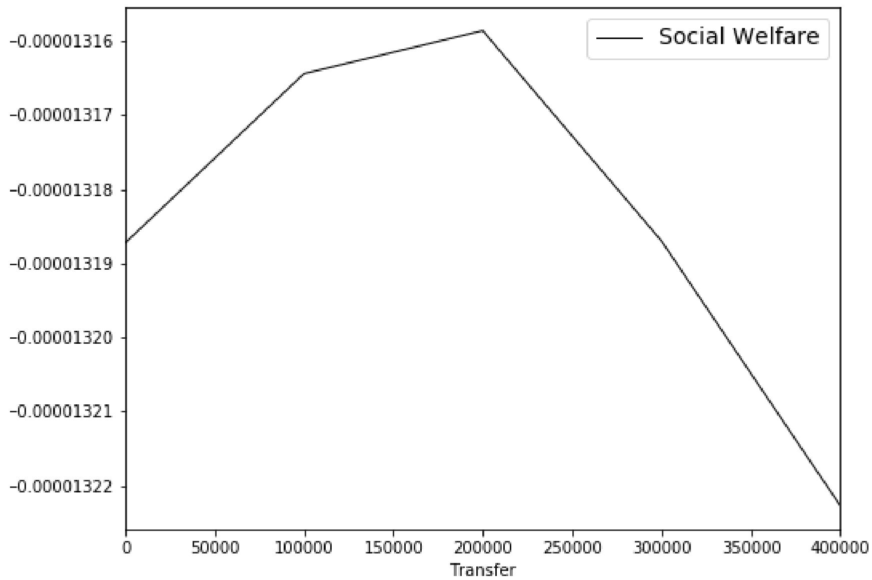


Figure 6: Social Welfare Function.

The horizontal axis is the amount of transfer. The vertical axis is the social welfare given by equation (18). The value of social welfare does not have any economic interpretation by itself.

earnings test has smoother tax rates than the current earnings test. Under the counterfactual system, the unemployed do not receive the transfer, but low-earning people receive the transfer. The counterfactual system gives low-earners more pension than the current system does, thereby reducing inequality.

The welfare-maximizing earnings test satisfies the implications of the optimal income tax theory. Based on the theory, the optimal marginal tax rates on low earnings are negative if the elasticity of an extensive margin relative to an intensive margin is sufficiently large. My quantitative exercise shows that the optimal policy in my model has the negative marginal tax rate at the bottom of earnings because the transfer for workers (tf) is positive.

Figure 8 shows the distribution of hours worked under the current and the optimal earnings test for people above 60. The share of the unemployed under the counterfactual system is lower than that of the current system. The ratio of part-time workers under the optimal policy is higher than that under the current policy. The transfer for workers encourages the unemployed to work as a

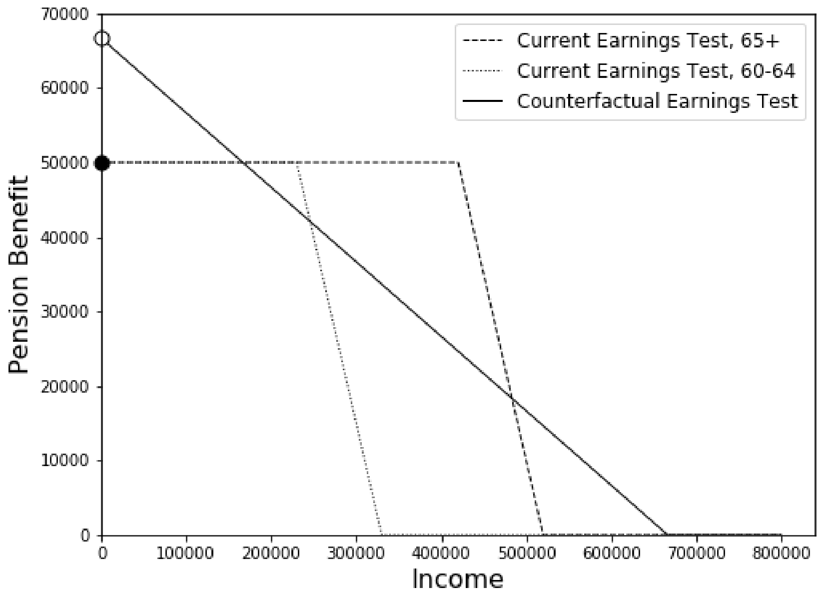


Figure 7: Current and Counterfactual Earnings Test.

This graph shows monthly labor income and monthly pension benefit deducted by the earnings test. It is assumed that people receive 50,000 yen ($\approx$ 500 USD) of pension per month. 'Current Earnings Test, 65+' implies the current system targeted to people above 65 years old. 'Current Earnings Test, 60-64' implies the current system targeted to people between 60 and 64 years old. White and black circles in the figure show that the unemployed do not receive the transfer, but workers receive it under the counterfactual system.

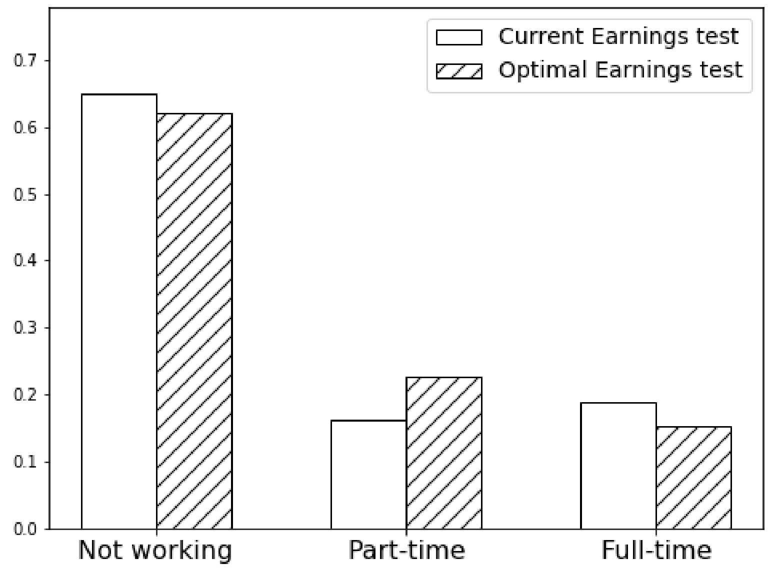
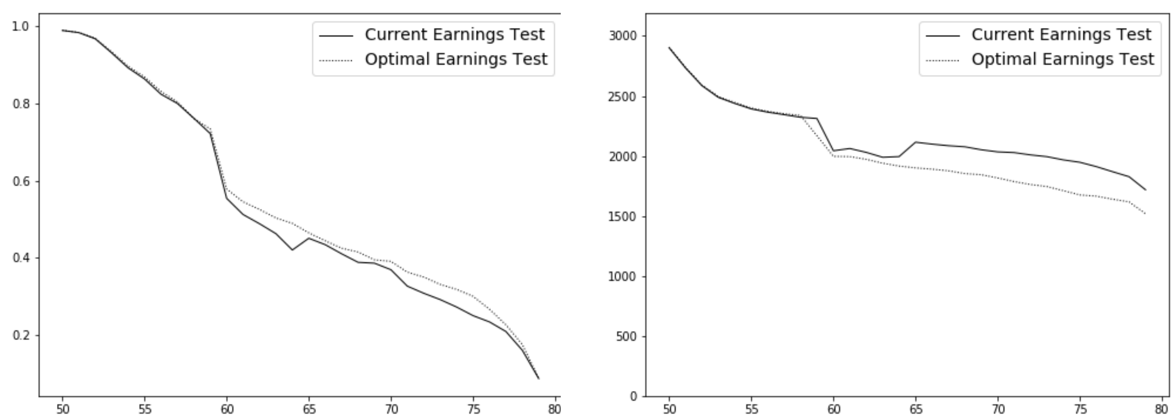


Figure 8: Distribution of Hours Worked under the Current and Optimal Earnings Test.

People above 60 years old. 'Not working' is people whose hours worked is zero. 'Part-time' is people whose weekly hours worked is between 1 and 36 hours. 'Full-time' is people whose weekly hours worked is above 36 hours. The same definition is used in the all following figures.

part-time worker. On the other hand, the linear tax discourages people from working as a full-time worker. The distribution of hours worked under the counterfactual system is more desirable because it reduces inequality

Figure 9 shows the simulated profile of the labor supply under the current earnings test and optimal earnings test. Figure 9-(a) and (b) show the downward profile of labor supply. There is a discontinuous decline at 60 because of the earnings test and the mandatory retirement system, which starts at 60. These systems discontinuously reduce after-tax wages. In figure 9-(a), the labor force participation rate under the optimal earnings test is higher than that under the current earnings test after 60 because of transfer for workers. In contrast, in figure 9-(b), the mean hours worked (excluding zero hours) under the counterfactual system is lower than that under the current system. Since the transfer for workers encourages the unemployed to work as part-time workers, the mean hours worked are low.



(a) Labor force participation rate (b) Mean hours worked excluding zero

Figure 9: Profile of Labor Supply under Current and Optimal Earnings Test

Figure 9 represents the profile of labor supply under the optimal earnings test and the current earnings test. The horizontal axis of (a) and (b) is age. (a) represents the labor force participation rate. (b) represents the mean of hours worked. The unemployed is considered as zero hours. The solid line represents labor supply under current earnings test. The dotted line represents labor supply under the optimal earnings test.

The consumption equivalence of the counterfactual system is 0.96%. There are two sources of welfare gain. The first one is the elimination of the inefficient current earnings test. The second one is the negative marginal tax rates of the counterfactual earnings test. The first welfare gain is obtained by comparing the current earnings test to the counterfactual earnings test with a linear tax rate. The second one is obtained by comparing the counterfactual earnings test with no transfer to the counterfactual earnings test with a positive transfer.

First, I consider the inefficiency of the current earnings test. Under the current system, people aged above 65 face much lower tax rates than people between 60 and 64 face. To avoid high tax rates targeted to their age (60-64), individuals aged 60-64 work less, and they work more when they are not 60-64 years old. Figure 9-(b) shows such a labor supply profile. The fluctuation of hours worked lowers utility because the disutility of labor supply is a convex function. When an age-independent linear tax rate replaces the current system, the labor supply profile is smooth.

The other source of inefficiency is that it imposes higher tax rates when their wages are high. Under the current system, before-tax wages for 60-64 years old are higher than those for those above 65. Because of high tax rates on those between the ages of 60-64, their after-tax wages are lower than those for individuals above 65. As a result, their hours worked are high though their earnings are low.

Under the counterfactual system, after-tax wages for 60-64 years old are higher than those under the current system. Individuals earn more with fewer hours worked under the counterfactual earnings test. Table 5 shows that hours worked under the optimal earnings test is lower than hours under the current system, but the present discounted value of mean income is higher. Under the counterfactual earnings test, individuals consume more with fewer hours worked, leading to higher utility.

Next, I consider how much positive transfer for low earnings improves social welfare. Figure 6 shows that the social welfare function is welfare-maximizing when the transfer is positive. The earnings test with zero transfer is linear. Its consumption equivalence is 0.72%. The consumption equivalence of 200,000 yen ($\approx$ 2,000USD) transfer is 0.96%.

Why do the negative marginal tax rates improve social welfare? Note that social welfare is improved only when the elasticity of the extensive margin is large. The negative marginal tax rates are evaluated in two terms: the marginal social value of reducing tax rates and the decline in tax revenue due to reducing tax rates. The negative marginal tax rates are optimal because the former is high and the latter is low. First, the marginal social value of reducing tax rates on low-earnings is high. The government cares for low-earning people, so the transfer for them greatly improves social welfare. Second, and more importantly, transfer for low earners does not reduce tax revenue much when the elasticity of an extensive margin relative to an intensive margin is large. The negative marginal tax rates on low earners do not necessarily make high earners reduce their hours worked because it is an intensive change. Since a decrease in labor supply is small, the negative marginal tax rates on low-earnings do not reduce tax revenue much. The marginal social value of reducing tax rates on low earners is higher than the decrease in tax revenue, so the negative marginal tax rates improve social welfare.

8.2 The Repeal of the Earnings Test

This section shows the effect of the repeal of the earnings test on the labor supply. An announcement representing the Japanese government's basic attitude in 2019 includes the repeal of the earnings test[17]. The government aims to encourage the elderly to work by repealing the earnings test's implicit income tax. Nevertheless, if the earnings test were abolished, the expenditure of the government would increase substantially: the repeal of the earnings test would lead to an increase in the pension expenditure by 1 trillion yen ($\approx$ 10 billion USD) annually[18]. The government faces a trade-off between increasing the labor force participation rate and reducing pension expenditure. This section shows that the optimal earnings test increases labor force participation rates more than the earnings test's repeal, so the government can avoid the trade-off.

I consider the effects of the optimal earnings test described in section 8.1 and the repeal of the earnings test on the labor force participation rate. When the earnings test is repealed, all individuals receive the full amount of pension benefit. I find that the labor force participation rate under the counterfactual earnings test is 1.9% higher than the value with no earnings test[19]. No earnings test implies that pension benefit is not deducted. The marginal tax rates are zero everywhere. In

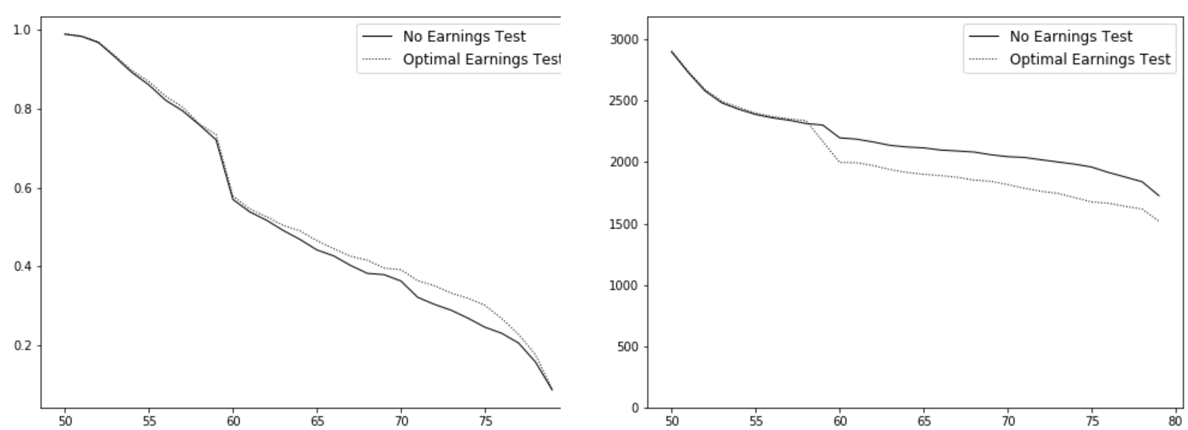
[17] "Basic Policy on Economic and Fiscal Management and Reform 2019"

[18] Source: "The fourth Social Security Council in 2011" by The Ministry of Health, Labor and Welfare. One trillion yen is 1% of the total annual government expenditure in Japan.

[19] The labor force participation rate for people aged above 50 years old is analyzed.

contrast, the optimal earnings test gives low-earning people more transfer than the unemployed. The marginal tax rate at zero earnings is negative. Therefore, the optimal earnings test raises the labor force participation rate more effectively than the earnings test's repeal. While the earnings test's repeal leads to a tremendous increase in fiscal expenditure, it does not effectively increase the labor force participation rate more than the earnings test to the negative marginal tax rates.

Figure 10 shows the profile of the labor force participation rate and mean hours worked under the optimal earnings test and no earnings test. Under the optimal earnings test, the labor force participation rate is higher than that under no earnings test for all ages. The mean hours worked under no earnings test are higher. While the marginal tax rates are zero everywhere under no earnings test, the counterfactual system's linear tax discourages high earners from working more.



(a) Labor Force Participation Rate (b) Mean Hours Worked Excluding Zero

Figure 10: Labor Supply under the Optimal Earnings Test and No Earnings Test.

Figure 10 represents the profile of labor supply under the optimal earnings test and no earnings test. The horizontal axis of (a) and (b) is age. (a) represents the labor force participation rate. (b) represents the mean of hours worked. The unemployed is considered as zero hours. The solid line represents labor supply under no earnings test. The dotted line represents labor supply under the optimal earnings test.

8.3 Raise Pension Eligibility Age and Cut Pension Benefit

Because of an aging society, it is quite likely that the Japanese government will implement policies to reduce the pension expenditure. Two policies can achieve it; raising the pension eligibility age and cutting the pension benefit. The first policy keeps the annual pension benefit constant, but people start to receive pension benefits several years later. The second policy keeps the pension eligibility age the same, but people receive a lower pension annually. It is not clear which policy is better in terms of social welfare. French (2005) simulates the effect of social security reform on labor supply, but does not analyze the social welfare and pension expenditure. This section quantitatively analyzes which policy achieves higher social welfare such that pension expenditure under the two policies is the same.

The first policy to be simulated is to raise the pension eligibility age from 60 to 61, keeping the annual pension benefit constant. The second policy is to cut the pension benefit by τ yen, holding the pension eligibility at age 60. I set τ yen so that the pension expenditures under the two policies are the same. In my model, τ is equal to 40,000 yen ($\approx$ 400USD). I compute social welfare under two policies. I find that raising the pension eligibility age is better than cutting pension benefits in terms of social welfare. Its consumption equivalence is 0.05%.

Why is raising the pension eligibility age a better policy? The key mechanism is the repeal of the earnings test. When pension eligibility is raised, people start to receive a pension from age 61, and the earnings test also starts from age 61. The earnings test's high marginal tax rate is abolished at 60, so people aged 60 face higher wages and work more. On the other hand, when the pension benefit is cut, people aged 60 face the earnings test's high marginal tax rate. Their earnings are low, leading to lower social welfare. Table 5 shows that the present discounted value of income under raising the pension eligibility age is higher.

I also analyzed the effect of raising the pension eligibility age on labor supply. One of the objectives of raising the pension eligibility age is to encourage the labor supply of elderly people and compensate for the shrinking labor force. Therefore, it is valuable to analyze the effect on labor supply behavior. When the pension eligibility age is raised by one year, the labor force participation rate's lifetime average is increased by 0.3%. The lifetime average of hours worked is increased by 9 hours annually. When the pension benefit is cut by 20% for all ages, the lifetime average of the labor force participation rate is increased by 1.3%. The lifetime average of hours worked is increased by 35 hours annually.

9 Conclusion

This paper makes three contributions. First, I derive the welfare-maximizing earnings test in the optimal income tax framework. I find that the social welfare under the current earnings test is substantially lower than that under the counterfactual earnings test with a linear tax rate. 0.72% of consumption equivalence is obtained without additional pension expenditure. Second, I apply the optimal income tax theory to the design of the earnings test. The theory states that the optimal tax rates on low-earnings are negative if the elasticity of an extensive margin relative to an intensive margin is large. After estimating an extensive margin's elasticity, I find that the earnings test with the negative marginal rates on low-earnings achieves higher social welfare than that without negative tax rates. Finally, I analyze the effect of other social security reforms. I estimate the life cycle model of elderly people with Japanese data by the methods of simulated moments. Although Japanese society is rapidly aging, the effect of social security reform on labor supply has never been estimated with the life cycle model.

As a policy recommendation, my simulation exercise shows that the optimal earnings test improves social welfare without additional pension expenditure. Other social security reforms, including increasing pension eligibility age, can reduce government expenditure and reduce the welfare of the elderly. These policies are unpopular among voters. Since the share of elderly people among voters is high in an aging society, there are political costs to implementing such a social security reform. In contrast, the reform of the earnings test avoids the trade-off, so it is politically costless. Politicians have incentives to implement the optimal design of the earnings test.

One area to examine in future work is extending the model to the general equilibrium model. The production side is not incorporated in the model, so the counterfactual policy's consumption equivalence is overestimated. I leave this exploration for future research.

References

[1] Altonji, Joseph G., and Lewis M. Segal. "Small-sample bias in GMM estimation of covariance structures." Journal of Business and Economic Statistics 14.3 (1996): 353-366.

[2] Attanasio, Orazio P., and Guglielmo Weber. "Is consumption growth consistent with intertemporal optimization? Evidence from the consumer expenditure survey." Journal of Political Economy 103.6 (1995): 1121-1157.

[3] Blundell, Richard, and Thomas MaCurdy. "Labor supply: A review of alternative approaches." Handbook of labor economics. Vol. 3. Elsevier, 1999. 1559-1695.

[4] Blundell, Richard, et al. "Optimal income taxation of lone mothers: an empirical comparison of the UK and Germany." The Economic Journal 119.535 (2009): F101-F121.

[5] Blundell, Richard, and Andrew Shephard. "Employment, hours of work and the optimal taxation of low-income families." The Review of Economic Studies 79.2 (2011): 481-510.

[6] Börsch-Supan, Axel, et al. "Dangerous flexibility–retirement reforms reconsidered." Economic Policy 33.94 (2018)a: 315-355.

[7] Börsch-Supan, Axel, Klaus Härtl, and Duarte N. Leite. "Earnings test, non-actuarial adjustments and flexible retirement." Economics Letters 173 (2018)b: 78-83.

[8] Braun, R. Anton, Karen A. Kopecky, and Tatyana Koreshkova. "Old, sick, alone, and poor: A welfare analysis of old-age social insurance programmes." The Review of Economic Studies 84.2 (2017): 580-612.

[9] Braun, R. Anton, Karen A. Kopecky, and Tatyana Koreshkova. "Old, frail, and uninsured: accounting for features of the US Long-Term care insurance market." Econometrica 87.3 (2019): 981-1019.

[10] Brinch, Christian N., Dennis Fredriksen, and Ola L. Vestad. "Life expectancy and claiming behavior in a flexible pension system." The Scandinavian Journal of Economics 120.4 (2018): 979-1010.

[11] De Nardi, Mariacristina, Eric French, and John B. Jones. "Why do the elderly save? The role of medical expenses." Journal of political economy 118.1 (2010): 39-75.

[12] Diamond, Peter A. "Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates." American Economic Review (1998): 83-95.

[13] Disney, Richard, and Sarah Smith. "The labour supply effect of the abolition of the earnings rule for older workers in the United Kingdom." The Economic Journal 112.478 (2002): C136-C152.

[14] French, Eric. "The effects of health, wealth, and wages on labour supply and retirement behaviour." The Review of Economic Studies 72.2 (2005): 395-427.

[15] French, Eric, and John Bailey Jones. "The effects of health insurance and self-insurance on retirement behavior." Econometrica 79.3 (2011): 693-732.

[16] Friedberg, Leora. "The labor supply effects of the social security earnings test." Review of Economics and Statistics 82.1 (2000): 48-63.

[17] Fukai, Taiyo, et al. "Medical Expenditures over the Life Cycle: Persistent Risks and Insurance." (2021).

[18] Geyer, Johannes, et al. "Labor market and distributional effects of an increase in the retirement age." Labour Economics (2020): 101817.

[19] Gourinchas, Pierre-Olivier, and Jonathan A. Parker. "Consumption over the life cycle." Econometrica 70.1 (2002): 47-89.

[20] Gruber, Jonathan, and David A. Wise, eds. Social Security and Retirement Around the World. National Bureau of Economic Research. Conference Report. University of Chicago Press, 1999.

[21] Gustman, Alan L., and Thomas L. Steinmeier. "A structural retirement model." Econometrica 54.3 (1986): 555–584.

[22] Heckman, James J. "A life-cycle model of earnings, learning, and consumption." Journal of political economy 84.4, Part 2 (1976): S9-S44.

[23] Ichimura, Hidehiko, Hideki Hashimoto, and Satoshi Shimizutani. "Japanese study of aging and retirement." JSTAR First Results (2009).

[24] Immervoll, Herwig, et al. "Welfare reform in European countries: a microsimulation analysis." The Economic Journal 117.516 (2007): 1-44.

[25] İmrohoroğlu, Selahattin, and Sagiri Kitao. "Labor supply elasticity and social security reform." Journal of Public Economics 93.7-8 (2009): 867-878.

[26] Imrohoroğlu, Selahattin, and Sagiri Kitao. "Social security reforms: Benefit claiming, labor force participation, and long-run sustainability." American Economic Journal: Macroeconomics 4.3 (2012): 96-127.

[27] Laroque, Guy. "Income maintenance and labor force participation." Econometrica 73.2 (2005): 341-376.

[28] Laun, Tobias, and Johanna Wallenius. "Social insurance and retirement: a cross-country perspective." Review of economic dynamics 22 (2016): 72-92.

[29] Lucas Jr, Robert E. "Macroeconomic priorities." American economic review 93.1 (2003): 1-14.

[30] Kitao, Sagiri. "Fiscal cost of demographic transition in Japan." Journal of Economic Dynamics and Control 54 (2015): 37-58.

[31] Kondo, Ayako "*Does the Elderly Employment Security Act increase the employment of elderly people? : The policy evaluation of the Elderly Employment Security Act*" [Translated from Japanese.] Journal of Japanese Labor Research 56.1 (2014): 13-22.

[32] Kondo, Ayako, and Hitoshi Shigeoka. "The effectiveness of demand-side government intervention to promote elderly employment: evidence from Japan." ILR Review 70.4 (2017): 1008-1036.

[33] Kopecky, Karen A., and Tatyana Koreshkova. "The impact of medical and nursing home expenses on savings." American Economic Journal: Macroeconomics 6.3 (2014): 29-72.

[34] Krusell, Per, et al. "Revisiting the welfare effects of eliminating business cycles." Review of Economic Dynamics 12.3 (2009): 393-404.

[35] Kuroda, Sachiko, and Isamu Yamamoto. "Estimating Frisch labor supply elasticity in Japan." Journal of the Japanese and International Economies 22.4 (2008): 566-585.

[36] MaCurdy, Thomas E. "An empirical model of labor supply in a life-cycle setting." Journal of political Economy 89.6 (1981): 1059-1085.

[37] Mirrlees, James A. "An exploration in the theory of optimum income taxation." The Review of Economic Studies 38.2 (1971): 175-208.

[38] Oecd. Pensions at a Glance. OECD, 2016.

[39] Pakes, Ariel, and David Pollard. "Simulation and the asymptotics of optimization estimators." Econometrica: Journal of the Econometric Society (1989): 1027-1057.

[40] Powell, James L. "Estimation of semiparametric models." Handbook of econometrics 4 (1994): 2443-2521.

[41] Rogerson, Richard, and Johanna Wallenius. "Nonconvexities, retirement, and the elasticity of labor supply." American Economic Review 103.4 (2013): 1445-62.

[42] Saez, Emmanuel. "Optimal income transfer programs: intensive versus extensive labor supply responses." The Quarterly Journal of Economics 117.3 (2002): 1039-1073.

[43] Sefton, James, Justin Van De Ven, and Martin Weale. "Means testing retirement benefits: fostering equity or discouraging savings?." The Economic Journal 118.528 (2008): 556-590.

[44] Song, Jae G., and Joyce Manchester. "New evidence on earnings and benefit claims following changes in the retirement earnings test in 2000." Journal of Public Economics 91.3-4 (2007): 669-700.

[45] Tran, Chung, and Alan Woodland. "Trade-offs in means tested pension design." Journal of Economic Dynamics and Control 47 (2014): 72-93.

[46] Vigtel, Trond Christian. "The retirement age and the hiring of senior workers." Labour Economics 51 (2018): 247-270.

Appendix A Pension system in Japan

A.1 Earnings Test in Japan

In Japan, earnings test is targeted to people above 60.Earnings test which people aged between 60 and 64 years old face is the following. I define monthly earnings as annual salaries plus annual bonus divided by 12. Total income is the sum of monthly earnings and monthly pension. Note that 100 yen $\approx$ 1 USD. There are five cases.

1. If the total income is below 280000 yen, pension is not deducted.

$$\text{deduction} = 0$$

2. The case when monthly pension is below 280000 yen and the total income is below 470000 yen

$$\text{deduction} = (\text{monthly earnings} + \text{monthly pension} - 280000) \times 0.5 \times 12$$

3. When monthly pension below 280000 yen and the total income is above 470000 yen

$$\text{deduction} = (470000 + \text{monthly pension} - 280000) \times 0.5 + (\text{monthly earnings} - 470000) \times 12$$

4. When monthly pension above 280000 yen and the total income is below 470000 yen

$$\text{deduction} = \text{monthly earnings} \times 0.5 \times 12$$

5. When monthly pension above 280000 yen and the total income is above 470000 yen

$$\text{deduction} = 470000 \times 0.5 + (\text{monthly earnings} - 470000) \times 12$$

This amount of deduction is deducted from the pension benefit. Marginal tax rates are 0% or 50% or 100%. In the case 1, marginal tax rate is 0%. In the case 2 and 4, it is 50%. In the case 3 and 5, it is 100%. According to data set, almost all people receive monthly pension below 280000 yen. 99 percentile of pension benefit is 250000 yen, so the case 4 and 5 is excluded in my model.

Figure 11 describes the income deducted by the earnings test under different pension benefit. The horizontal axis is labor income. The vertical axis is the pension benefit deducted by the earnings test. When income is low, people get full amount of pension benefit. When the income is higher than a specified threshold, the marginal tax rate is 50%. When the income is higher than another threshold, they receive zero pension benefit and they face 0% marginal tax rate.

People aged above 65 years old face the following system.

1. If the total income is below 470000 yen, pension is not deducted.

$$\text{deduction} = 0$$

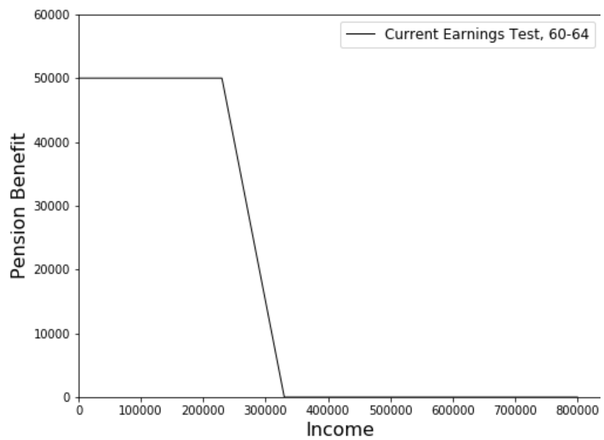
2. If the total income is above 470000 yen,

$$\text{deduction} = (\text{monthly earnings} + \text{monthly pension - } 470000) \times 0.5 \times 12$$

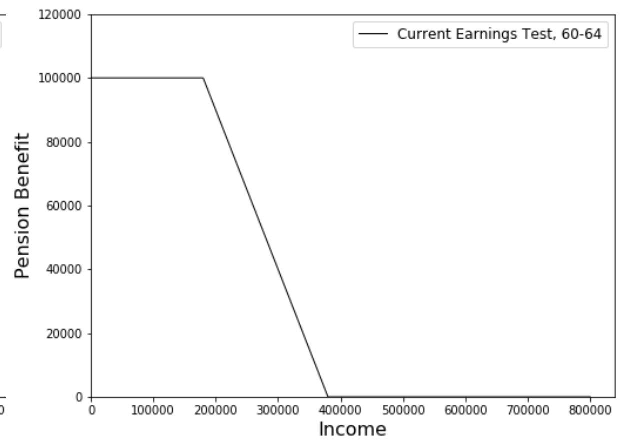
Figure 12 shows the income deducted by the earnings test under different pension benefit. People above 65 face lower marginal tax rates than people aged 60-64. The area with zero marginal tax rates is broader than earnings test targeted to 60-64. Labor supply for people 60-64 is more discouraged.

A.2 Other Aspects of Pension System

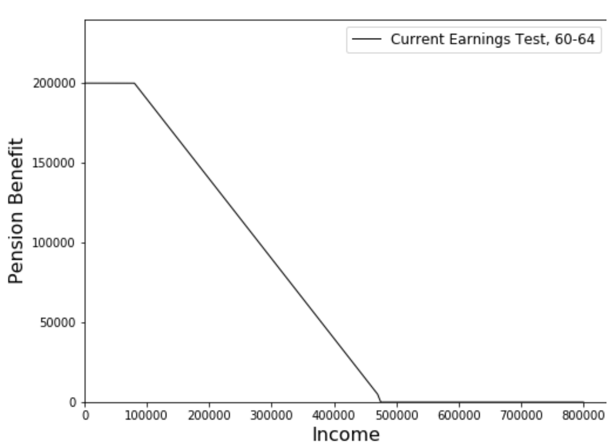
First, I describe how the pension benefit before deduction that simulated individuals receive is decided in the model. Pension benefit is denoted as Pb_t in the budget constraint. In the Japanese system, the amount of pension benefit is the sum of basic part, which does not depend on past earnings, and the proportional part, which is proportional to past earnings. Since it is too complicated to incorporate the full pension system into the model, I simplify the pension system so that the distribution of pension benefits in data is approximately equal to the distribution pension benefits that simulated individuals receive in models.



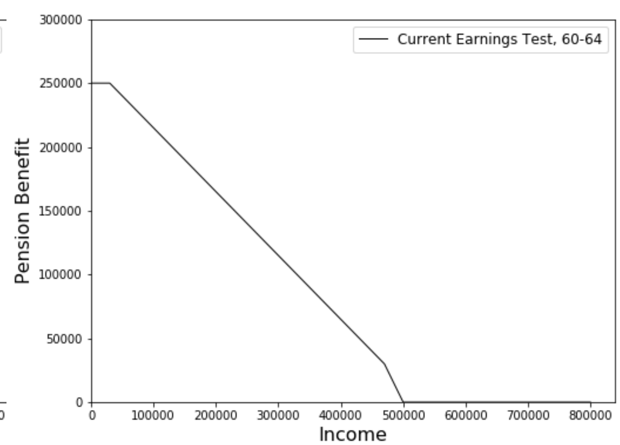
(a) Pension benefit is 50,000 yen



(b) Pension benefit is 100,000 yen

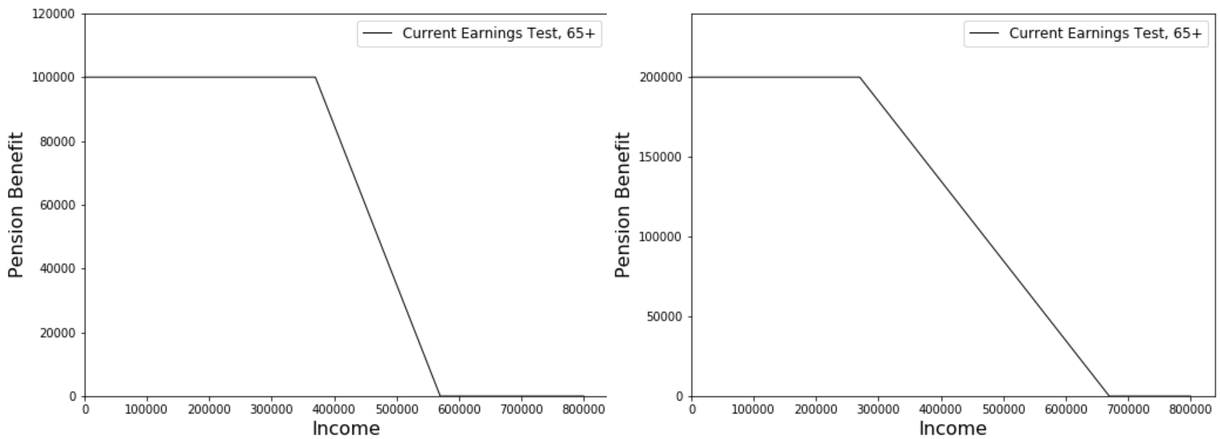


(c) Pension benefit is 200,000 yen



(d) Pension benefit is 250,000 yen

Figure 11: Earnings Test Targeted to 60-64.



(a) Pension benefit is 100,000 yen

(b) Pension benefit is 200,000 yen

Figure 12: Earnings Test Targeted to 65+.

In a simulated model, annual pension benefits people receive take 12 values that are equally distributed from 250,000 yen (≈ 2,500 US dollars) to 3,000,000 yen (≈ 30,000 US dollars). The lower bound of 250,000 yen is considered as the fixed part of pension benefits. The reason why 3,000,000 yen is set as the upper bound is that 3,000,000 yen is 99 percentile of pension benefit in data. In a model, annual pension benefits people receive take 12 values that are equally distributed from 250,000 yen (≈ 2,500 US dollars) to 3,000,000 yen (≈ 30,000 US dollars). The lower bound of 250,000 yen is considered as the fixed part of pension benefits. 17% of people in data receive pension benefits less than 250,000 yen. 3,000,000 yen is set as the upper bound because 3,000,000 yen is 99 percentile of pension benefit in data. Although 3,000,000 yen seems to be high, only 3% of people receive between 2,750,000 yen and 3,000,000 yen in data. Therefore, it is not quantitatively significant, but it is included for robustness.

Table 6: Distribution of pension benefits in model.

The amount of pension benefits	Ratio
250,000 yen	0.17
500,000 yen	0.09
750,000 yen	0.15
1,000,000 yen	0.15
1,250,000 yen	0.12
1,500,000 yen	0.04
1,750,000 yen	0.05
2,000,000 yen	0.05
2,250,000 yen	0.06
2,500,000 yen	0.03
2,750,000 yen	0.03
3,000,000 yen	0.04

Table 6 shows the distribution of pension pension benefits people receive in data. For example, 17% of people receive 250,000 yen of pension benefit in my model. This ratio is approximately equal to the ratio in my data set. In the same way, this distribution of pension benefits in simulated model is approximately equal to the distribution of pension benefits in data.

In the model, there is heterogeneity in the amount of pension benefits. Who receives higher pension benefits and who receives lower pension benefits in the model? The way to determine the individuals' amount of pension benefits is based on by the total amount of wages they earned between 50 and 60 years old. In this sense, pension benefit is proportional to the past earnings. For instance, people whose amount of wages they earned between 50 and 60 years old is in the 10th percentile in this model receive the 10th percentile of pension distribution in the data, which is 250,000 yen. The pension benefit before deduction is constant over lifecycle though the earnings test is age-dependent.

Since the pension benefit is completely decided by exogenous wages in the model, individuals in this model will not increase their hours worked to receive more pension in the future. Once pension benefit is decided at 60 years old, it is constant over the life cycle in my model.

I discuss when people start to receive pension in my model. In Japan, people can choose the pension eligibility age under the current pension system. The eligibility age can be from 60 to 70 years old. In contrast, in my model, it is assumed that all people receive a pension from 60. The current pension system and my research objectives justify this assumption. My goal is finding the welfare improving earnings test. In Japan, even if working elderly people prolong to claim the pension benefit, pension is deducted by the earnings test. If they work, the amount of benefit, which they will receive in the future is decreased. Individuals cannot take advantage of delaying the claim of pension. For example, suppose there is a person A aged 60 who earns 400,000 yen monthly and can receive 300,000 yen of pension benefits monthly. Under the earnings test, 200,000 yen is deducted from pension benefits monthly if they receive pension benefits at 60. Importantly, *even if they postpone their pension, 200,000 yen is deducted from pension benefits after they start to receive pension.* Suppose a person A postpones pension benefits at 60 and start to receive them at 61. His pension benefit is 100,000 yen (=300,000-200,000) at 61. In this sense, when people start to receive pension does not matter if this paper focuses on the earnings test. Therefore, it is justified to assume that all people receive a pension from 60 in my model. Also, in my data set, 68% of people start to receive pension from 60. 11% of people start to receive pension from 61. 10% of people of people start to receive pension between 62 and 64. 5% of people start to receive pension from 65. Given this data, I believe that the assumption that all people start to receive pension from 60 can be justified.

Online Appendix: Note For Publication

Appendix B Data on Health Condition and Survival Probability

I calibrate the transition matrix of the health condition. Self-rated health status is used, defined as follows. questionnaires ask people to evaluate their overall health status according to five levels, which are "excellent," "good," "fair," "bad," "terrible." In my model, health status takes ’good’ or ’bad’. Good health is "excellent," "good," and "fair". Bad health is"bad" and "terrible." Figure 13 is the life cycle profile of the ratio of good health.

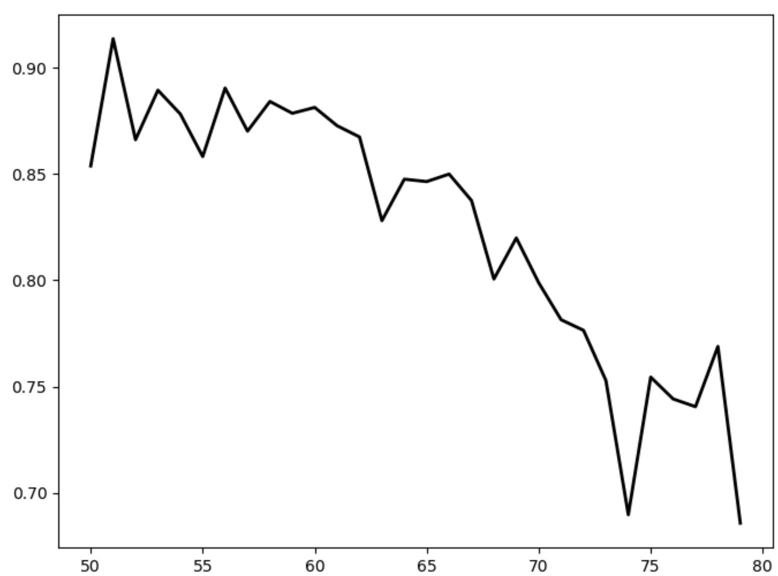


Figure 13: Life Cycle Profile of Good Health Ratio.

Health status follows a Markov process. The health transition probabilities (prob$(M_{t+1}|M_t, \text{age})$ in equation (7)) are separately estimated in three age groups, which are 50–59, 60–69, and 70-79 years. Table 7 shows a result.

Table 7: Health Transition Matrix

Age		Good in the next year	Bad in the next year
50-59	Good	0.927	0.073
	Bad	0.608	0.392
60-69	Good	0.909	0.091
	Bad	0.593	0.407
70-79	Good	0.849	0.151
	Bad	0.477	0.523

Source: JSTAR. Samples include both sex and self-employed.

The survival productivity (s_t in equation (7)) is calibrated using another source. JSTAR does not have rich information about mortality rate, so the survival ratio is cited from the "Simple Life Expectancy Table" generated by the Ministry of Health, Labor and Welfare in Japan in 2017. It covers the entire Japanese population as their sample. Figure 14 shows survival probability over lifecycle. For instance, people die with a probability 0.01105 at age 70, and they die with a probability 0.0348 at age 80.

Appendix C The Procedure of Simulation

The procedure of obtaining the profile of decision variables is as follows.

1. Policy function is obtained given estimated parameters in by backward induction. Choice variables are consumption and hour worked, and state variables are asset, health condition, and wage. Policy function reveals what level of consumption and hour worked households choose given state variables.

2. Households enter market at 50 years old with state variables. Given them, households choose

[20]Source: "Simple Life Expectancy Table" generated by the Ministry of Health, Labor and Welfare in Japan in 2017. The horizontal axis is age. The vertical axis is the survival probability. 0.98 implies that 98% of people survive and 2% of people die in the year.

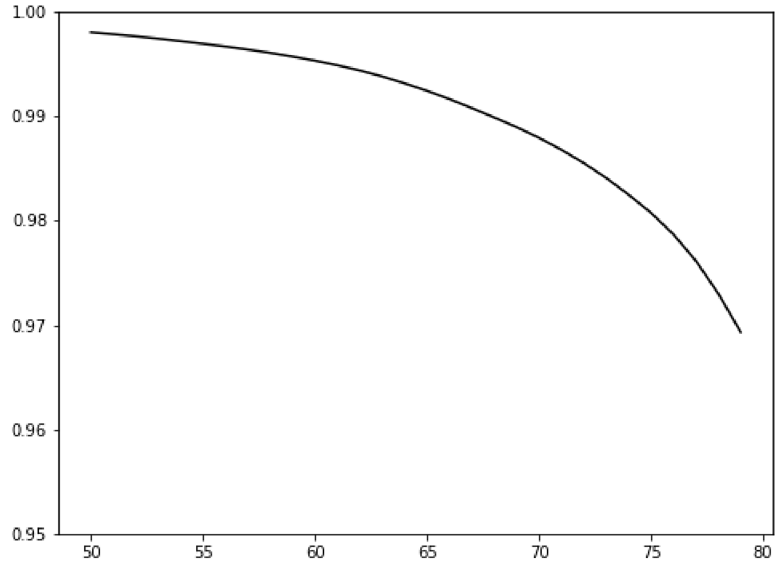


Figure 14: Life Cycle Profile of Survival Probability[20].

consumption and hour worked, following policy function obtained in step 1. Then, asset at 51 years old is determined from budget constraint.

3. Wage and health condition at 51 years old are stochastically determined based on those variables at 50 years old and Markov process, so all state variables at 51 years old are determined.

4. Choice variables at 51 years old are determined based on policy function and state variables. This procedure continues by 82 years old.

5. Whole procedure is repeated by 20000 times. For each variable and each age, there are 20000 individuals, and average is taken.

When computing the social welfare, the lifetime profiles of consumption, hours worked, and assets are substituted into the lifetime utility. I take average social welfare over simulated individuals. This is social welfare.

When computing the consumption equivalence of the welfare maximizing earnings test, first, I compute two social welfare: the social welfare under the current system and the social welfare under the welfare maximizing earnings test. Second, I multiply consumption under current policy by $\nu\%$ for all age t. I derive ν that makes the social welfare under the current policy equal to the social welfare under the optimal policy. The ν is consumption equivalence.

Appendix D Identification

In this section, I discuss the identification of preference parameters. The fixed cost of working (θ), which determines the elasticity of an extensive margin, is identified by the distribution of hours worked. The larger fixed cost of working makes more people choose the unemployment instead of part-time worker. Due to the high fixed cost of working, individuals who work for a few hours obtain a few earnings but lose much leisure. In contrast, when the fixed cost of working is small, people choose part-time work instead of unemployment. The share of part-time workers and the unemployed identifies the fixed cost of working.

The level of earnings identifies time endowment (L). When individuals have more time endowment, they can work longer and earn more. When they have less time endowment, their earnings are low. Additionally, time endowment is identified by the share of part-time workers and full-time workers. Part-time workers are defined as the people who work between 1 and 36 hours per week. If people have more time endowment, fewer people are part-time workers[21].

I simulate how changes in key parameters affect the decision variables in my model. Figures 15-18 show the profiles for the decision variables when one parameter is changed, holding all other parameters at their baseline values. There are the three cases of parameter changes; a decrease by 20%, a a baseline, and an increase by 20%.

Figure 15 shows the decision variables when the coefficient of risk aversion (ν) is changed. Figure 15 reveals that ν is identified by the asset profile. The asset profile strongly responds to the coefficient of risk aversion. The higher the coefficient of risk aversion is, the more individuals save.

[21]Time endowment cannot be normalized as one. The rule of the earnings tests includes the level of earnings and pension benefit. The matching moment should include the level of earnings. If time endowment is normalized, the level of earnings cannot be matched.

Labor supply does not respond to the change in the coefficient of risk aversion[22].

Figure 16 shows the change in consumption weight (γ). Consumption weight is identified by the profile of savings and the distribution of hours worked. Higher consumption weight leads to fewer savings and more hours worked. Notably, γ changes the share of not working and full-time, but the share of part-time is stable. γ is identified by the distribution of hours worked.

Figure 17 shows the change in time endowment (L). Time endowment is identified by the level of earnings. Higher time endowment leads to a higher level of earnings. Change in the level of earnings is larger than changes in other figures. Moreover, time endowment is identified by the share of part-time and full-time. As time endowment increases, there are more full-time workers and fewer part-time workers. A negative correlation between the share of part-time workers and full-time workers is observed in only figure 17.

Figure 18 shows the change in the fixed cost of working (θ). The fixed cost of working is identified by the share of unemployed and part-time workers. The change in the share of unemployed is large when θ is changed. The higher the fixed cost of working is, the larger the elasticity of an extensive margin. There are more unemployed when the fixed cost of working is large. On the other hand, when the fixed cost of working is small, more people choose part-time workers instead of unemployment.

[22] Given the level of consumption, the first-order condition with respect to labor supply is

$$H_t = L - \theta - \frac{1-\gamma}{\gamma}\frac{1}{w_t}C_t.$$

Under the utility function (1), the labor supply equation does not include the coefficient of risk aversion (ν). Therefore, the distribution of hours worked does not change when the coefficient of risk aversion changes.

[i] These figures show the profiles for the decision variables under the different parameter values. The top figure is the profile of savings. Its vertical axis is age. The middle figure is the distribution of hours worked. It shows the share of unemployed, part-time, and full-time workers. Part-time workers are those who work between 1-36 hours per week. Full-time workers are those who work more than 36 hours per week. The bottom figure is the distribution earnings. 20th, 40th, 60th, and 80th percentile of earnings are described.

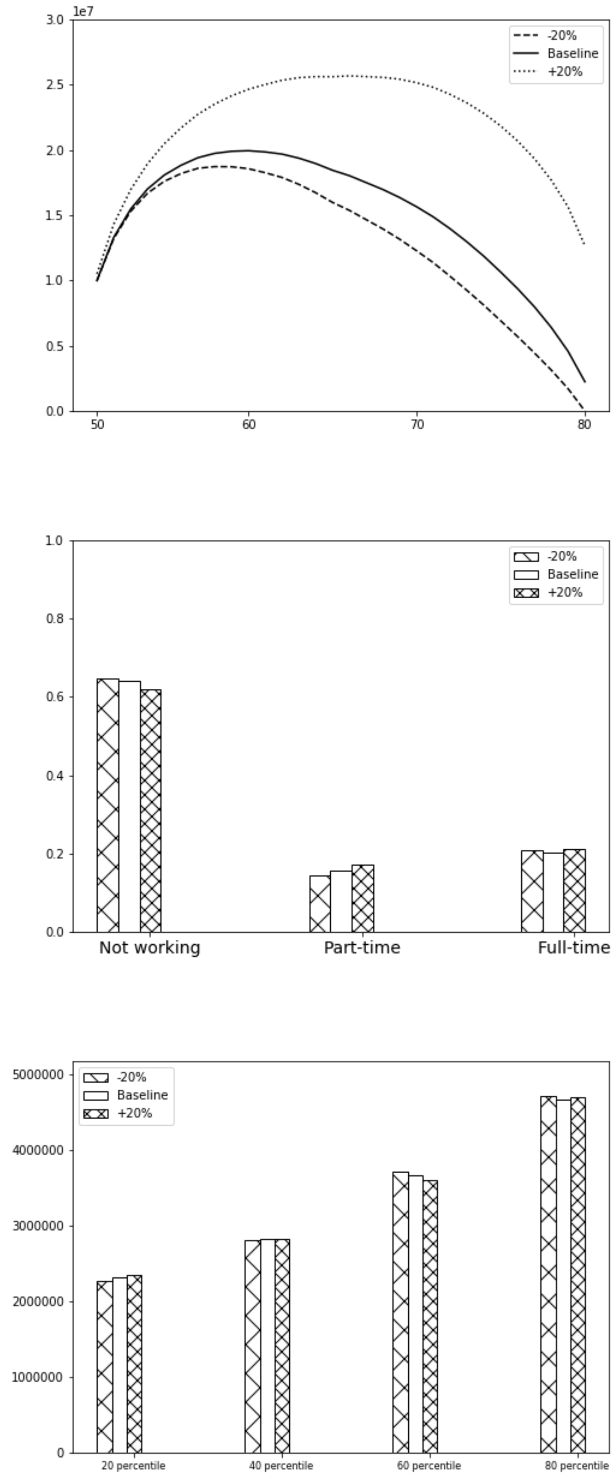


Figure 15: Change in the Coefficient of Risk Aversion (ν) [i].

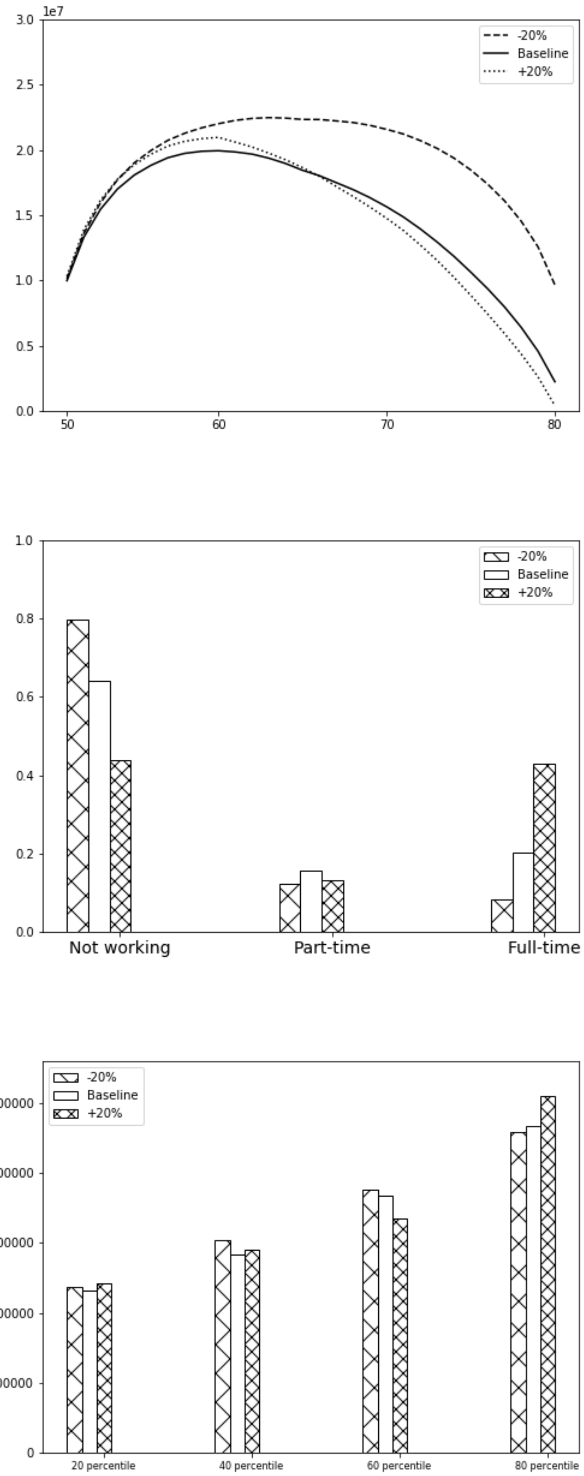


Figure 16: Change in Consumption Weight (γ)[i].

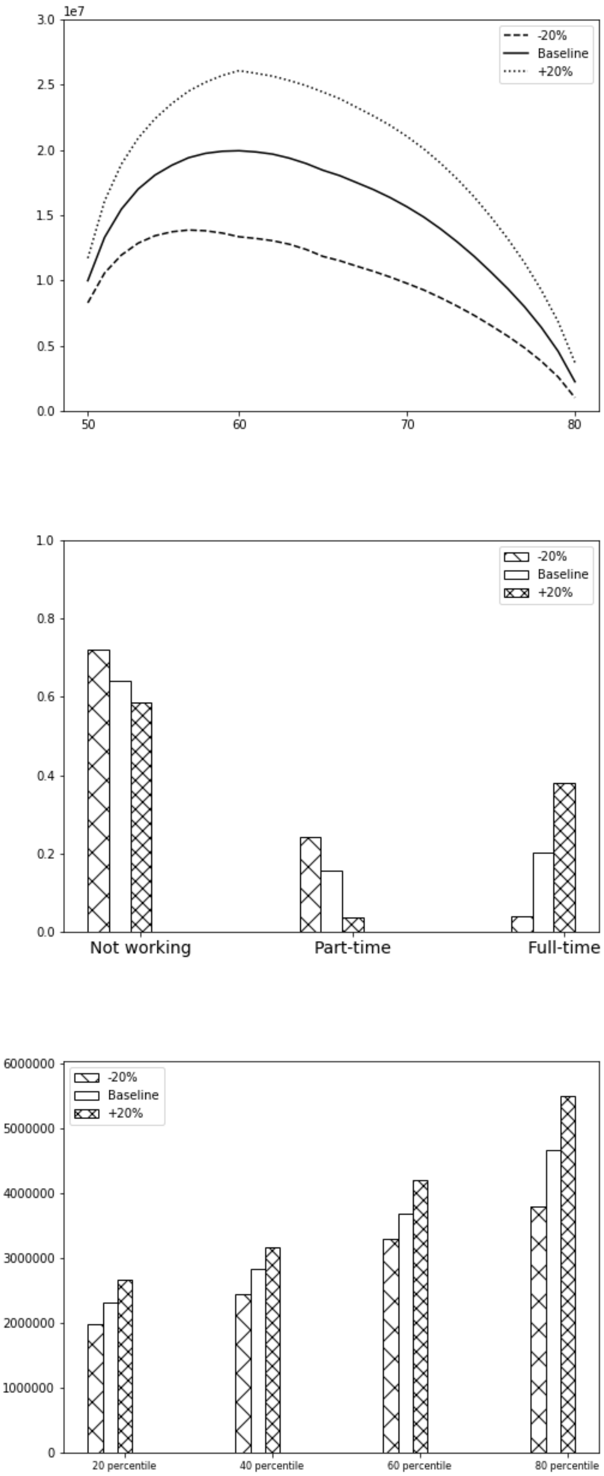


Figure 17: Change in time Endowment (L) [i].

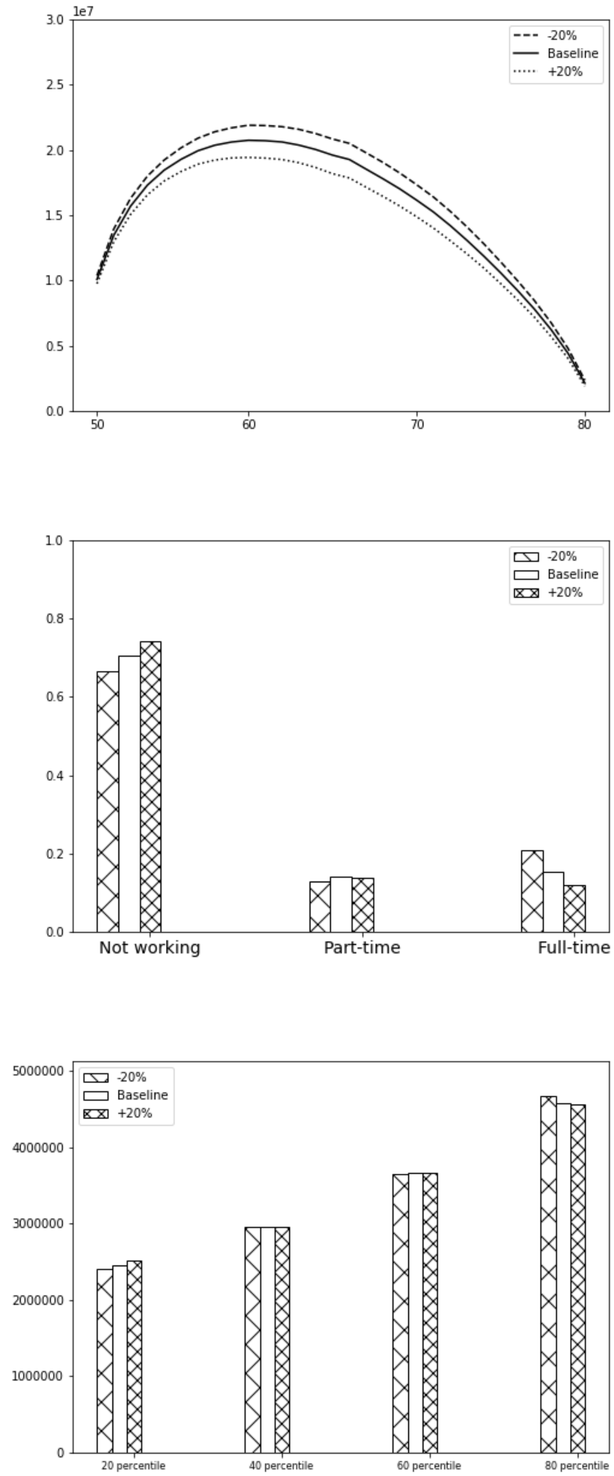


Figure 18: Change in Fixed Cost of Working (θ) [i].

Appendix E The Asymptotic Distribution of Parameter Estimates

In this section, I discuss the asymptotic distribution of parameter estimates. Suppose I have a data set of I independent individuals who are each observed for T periods. Let $\phi(\theta;\chi_0)$ denote the moment conditions, and $\hat{\phi}(\theta;\chi_0)$ is its sample analog. $\hat{\mathbf{W}}_{\mathbf{I}}$ is a weighting matrix, and it is $M \times M$ matrix, where M is the number of moments. The MSM estimator $\hat{\theta}$ is given by

$$\underset{\theta}{argmin} \frac{I}{1+\tau} \hat{\phi}'_I(\theta,\chi_0)\hat{\mathbf{W}}_{\mathbf{I}}\hat{\phi}_I(\theta,\chi_0) \tag{21}$$

where τ is the ratio of the number of observations to the number of simulated observations. Under the regularity condition described in Pakes and Pollard (1989), the MSM estimator is consistent and asymptotically normally distributed:

$$\sqrt{I}(\hat{\theta}-\theta_0) \rightsquigarrow N(0,\mathbf{V}) \tag{22}$$

The variance-covariance matrix $\mathbf{V}$ is given by

$$\mathbf{V} = (1+\tau)(\mathbf{D}'\mathbf{W}\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}\mathbf{S}\mathbf{W}\mathbf{D}(\mathbf{D}'\mathbf{W}\mathbf{D})^{-1} \tag{23}$$

where $\mathbf{S}$ is the variance-covariance matrix of data, and $M \times M$ matrix, where M is the number of moments. $\mathbf{D}$ is the Jacobian matrix of the population moment vector and is the $M \times N$ matrix, where N is the number of parameters to be estimated.

$$\mathbf{D} = \frac{\partial\phi(\theta,\chi_0)}{\partial\theta'}|_{\theta=\theta_0} \tag{24}$$

$\mathbf{W} = plim_{I\to\infty}\{\hat{\mathbf{W}}_{\mathbf{I}}\}$.

The asymptotically efficient weighting matrix is $\mathbf{S}^{-1}$ that is the inverse of the variance-covariance matrix of the data. When $\mathbf{W} = \mathbf{S}^{-1}$ holds, $\mathbf{V}$ is simplified to $(1+\tau)(\mathbf{D}'\mathbf{S}^{-1}\mathbf{D})^{-1}$. However, the optimal weighting matrix can bias estimators in small samples (Altonji and Segal (1996)). I use a diagonal weighting matrix. The diagonal weighting matrix uses the inverse of the matrix that is the same as $\mathbf{S}$ along the diagonal and has zeros off the diagonal of the matrix.

I estimate $\mathbf{S}$, $\mathbf{D}$, and $\mathbf{W}$ with their sample analogs. I use bootstrap to estimate the variance-covariance matrix of data, $\mathbf{S}$. The procedure is the following. First, JSTAR includes 7,268 of independent individuals who are observed at least one period. 7,268 of individuals are withdrawn from JSTAR allowing for replacement. Using this data set, each moment is calculated. Second, the first step is repeated for 3,000 times. 3,000 observations of each moment is obtained. Third, the variance-covariance matrix is calculated using the data from step 2. When computing $\mathbf{D}$, I merely take numerical derivatives of $\phi_I(\hat{\theta}, \chi_0)$.

Appendix F Decomposition of welfare gain

Figure 19 shows the decomposition of social welfare. The vertical axis shows the transfer (tf) and linear tax rate (τ) of the counterfactual earnings test. The horizontal axis is the consumption equivalence of the counterfactual earnings test. The triangle shows the consumption equivalence of the earnings test with both linear tax (τ) and transfer (tf) shown in figure .The empty bar chart shows the consumption equivalence of earnings test that has only transfer and does not have a linear tax rate. For instance, when counterfactual earnings test has 300000 yen of transfer and 0% of linear tax rate, its consumption equivalence is +4.5%. The hashed bar chart shows the consumption equivalence of earnings test that has only linear tax rate and does not have transfer. For instance, when counterfactual earnings test has zero transfer and 16% of linear tax rate, its consumption equivalence is -1.7%. It always reduces welfare compared to the current earnings test.

Unexpectedly, the consumption equivalence of earnings test that have both transfer and linear tax rate is markedly lower than the sum of two consumption equivalence. This is because only workers can receive transfer. If there is only transfer, it greatly encourages non-employed people to participate in labor market. However, if there are both transfer and a linear tax rate, their labor is less encouraged. Hours worked are higher than some threshold due to the fixed cost of working. The benefit for non-employed people to participate in labor market is reduced by linear tax rate and hours worked which are higher than some threshold.

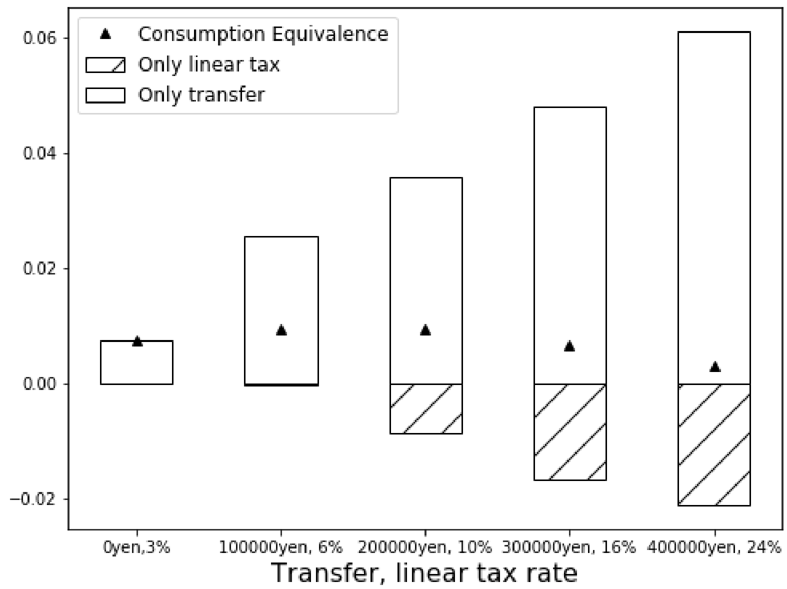


Figure 19: Decomposition of social welfare.

Appendix G Sensitivity of Results

G.1 Saving profile

I did additional exercise to check whether my result is robust to parameters that make people save less. I changed the value of coefficient of risk aversion (ν) and time endowment (L) to fit the simulated profile of savings to the data profile. In summary, the main result does not change under new parameters.

A new set of parameters that make people save less is shown in table 8. The coefficient of risk aversion and time endowment are lower than baseline case.

Table 8: Preference Parameter Estimates

Parameters	Baseline	Low savings case	French (2005)
ν: Coefficient of risk aversion, utility	2.9	2	3.1
γ : Consumption weight	0.49	0.49	0.533
L : Time endowment	5700	4700	3900
θ : Fixed cost of working	1000	1000	335
ϕ : Fixed cost of bad health	500	500	196
θ_b : Parameters of bequest	0.01	0.01	1.70
Frisch elasticity, age 55	0.86	1.16	–
Frisch elasticity, age 60	1.15	1.69	1.10
Coefficient of relative risk aversion	1.93	1.49	2.17

Figure 20 shows the profile of savings under (1) simulated profile with baseline parameter (2) simulated profile with new parameter (low savings case) (3) data profile. Under the new parameters, individuals save less because they are less risk-averse.

How the main result changes with there new parameters? The main result of this paper is that the optimal earnings test has a negative marginal tax rate. Figure 21 shows the replication of figure 6 in the main manuscript. It shows that a negative marginal tax rate is still welfare-improving.

Why is this the case under new parameters? As section 2 in the main manuscript stated, the large elasticity of extensive margin makes a negative marginal tax rate welfare-improving. Figure 22 shows that the ratios of "not working" are similar in all three cases. This implies that new parameters do not lower the size of the elasticity of extensive margin. Therefore, a negative marginal tax rate is still welfare-improving. The main conclusion of this paper holds in a case where people save less.

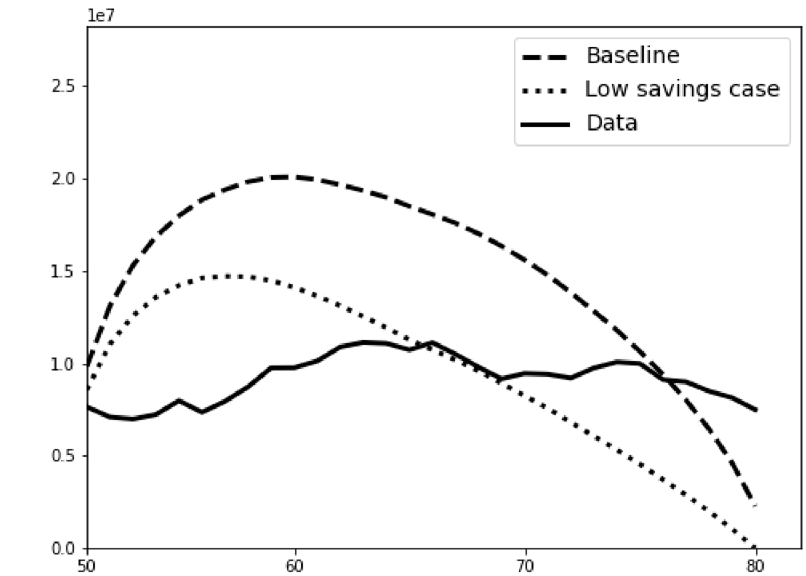


Figure 20: Profile of savings.

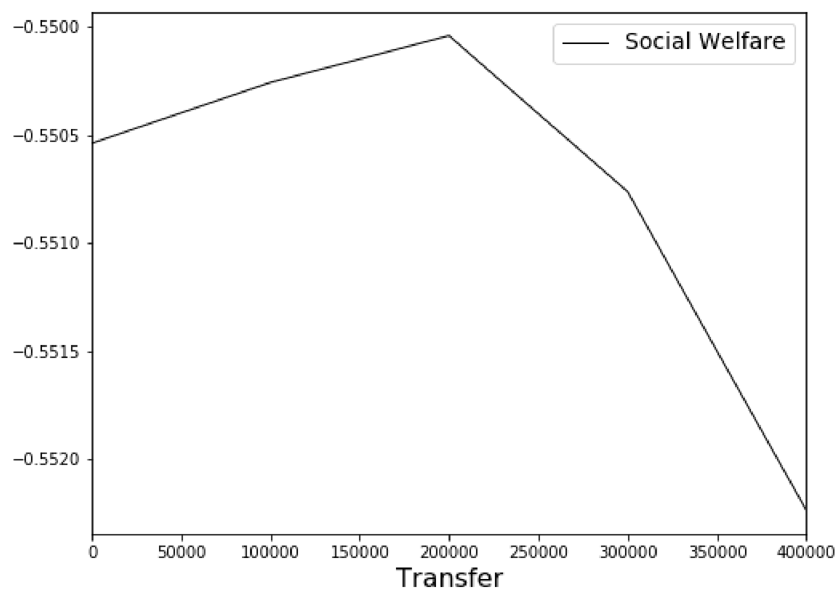


Figure 21: Social Welfare Function.

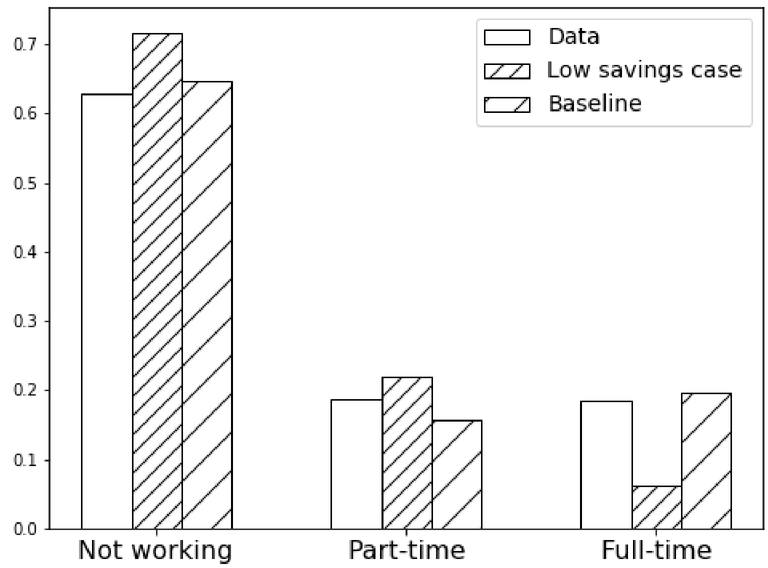


Figure 22: Distribution of hours worked.

G.2 Life cycle starting at 20 years old

One of controversial assumptions in my model is that the life cycle start at 50 years old, instead of younger age. This assumption comes from the data constraint. In order to start life cycle at younger age, exogenous variables to be given into the model, including wages and health condition, and the target moments, including labor supply and savings, are necessary in data. However, data for people younger than 50 years old is not available in JSTAR. Therefore, I cannot model people younger than 50 years old. How this assumption affects my results?

First, this assumption is justified if I consider that the government suddenly enacts new policies without announcing them beforehand. The individuals cannot adjust their asset because the counterfactual policies are suddenly enacted. The assumption that the distribution of asset at 50 is fixed is justified.

Second, if I consider the case where the government announces the policies before hand and individuals can adjust their assets before the implementation, what problems does the assumption cause? When life cycle starts at 50 years old, the initial wealth, with which individuals enter the market is fixed even when the counterfactual policy is enacted. In my model, people enter the market at 50 with some wealth. When counterfactual policy is enacted, in the reality, it should affect the savings at 50. Counterfactual policy makes people change their capital accumulation behavior before 50. However, in my model, it is assumed that initial wealth at 50 is fixed even when counterfactual policies are enacted.

My assumption that individuals enter the market at 50 overestimates their savings at 50 under the welfare improving earnings test. In other words, if the life cycle start at younger age, like 20, the average savings at 50 under the welfare improving earnings test is lower than the average savings under the current earnings test. Under the current earnings test, the marginal tax rates are substantially high, and it discourages the labor supply. Individuals need to accumulate savings more because they cannot earn much after they get older. In addition, under the welfare improving earnings test, the low-earnings people are supported by the lump-sum transfer, so their labor earnings after 50 is higher than the earnings under the current earnings test. They do not need to accumulate their savings for life after retirement, so their savings at 50 get lower.

To confirm that under the welfare improving earnings, the savings at 50 get lower, I simulate the profile of savings with the life cycle model that start at 20. Since there are no data about wages before 50, I create realistic hump-shape profile of wage. Figure 23 shows the profile of savings under the current earnings test, welfare-improving earnings test, and no earnings test. It shows that, savings under welfare improving earnings test at 50 years old is higher than savings under current earnings test by 10%.

In summary, when I assume that the life cycle start at 50, the savings at 50 are overestimated when the welfare improving earnings test is enacted. How this fact affects my results? There are two main results in this paper. First, the welfare improving earnings test substantially improves social welfare. Second, the earnings test that maximizes social welfare have negative marginal tax rates on low earnings people.

As for the first result, the assumption that the life cycle starts at 50 underestimate the consumption equivalence of welfare maximizing earnings test. When initial wealth at 50 is fixed when the counterfactual earnings test is enacted, the savings are overestimated. When individuals have more savings, they reduce their hours worked after 50. Since the welfare gain of the counterfactual policy comes from the labor supply, less hours worked reduce the social welfare under the counterfactual earnings test. My first result states that social welfare under the counterfactual earnings test is

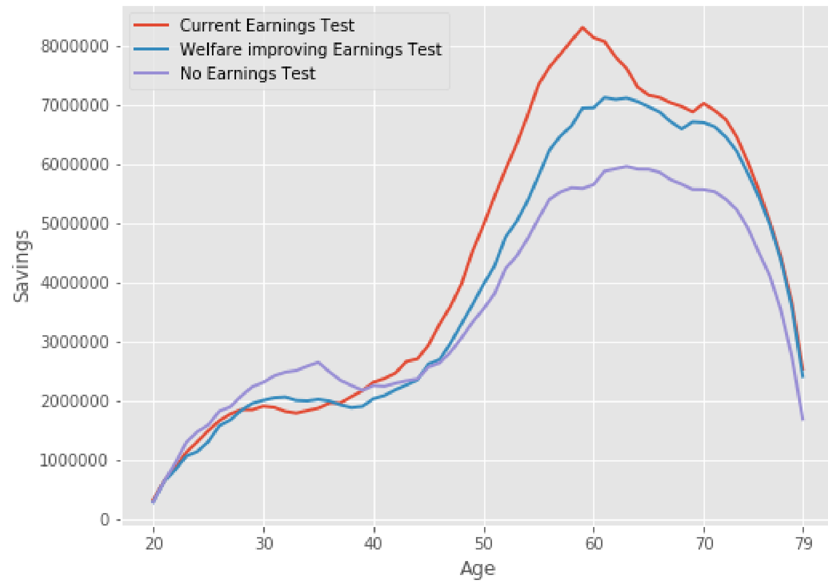


Figure 23: The Simulated Profile of Savings.

much higher than that under the current system. My assumption reduces the social welfare under the counterfactual earnings test. Consequently, I can safely argue that the welfare gain from the counterfactual earnings test is substantial.

Next, how the assumption affects the second result about the negative marginal tax rates? There are two reasons why negative marginal tax rates are optimal; the large elasticity of extensive margin, and the ratio of poor people in society. When there are more poor people in society, it is more likely that it is better for the government to impose negative marginal tax rates on low-earnings people. My assumption about the initial savings is related to the ratio of poor people. The assumption reduces the ratio of individuals with fewer savings under the counterfactual policy. As figure 23 shows, when it is assumed that people enter the market at 50, it overestimates the savings at 50. It reduces the ratio of poor people. In summary, it is safe to claim the negative marginal tax rates even when people enter the market at 50.

G.3 Long-term care risk

The model already incorporates health risks, but not long-term care risks. When people are in bad health, the fixed amount of leisure time (ϕ) is reduced from their leisure time as stated in equation (1). I believe this fixed cost of bad health in the utility function partially answers the referee's question.

However, this model does not incorporate long-term care risk. To consider the long-term care risk in additional exercises, I did the following two exercises.

Case 1. The probability of being bad health (prob$(M_{t+1} = \text{bad}|M_t, \text{age})$ in equation (7)) gets higher.

Case 2. The probability of being bad health gets higher, and the cost of bad health (ϕ in equation (1)) is increased.

Then, I analyze how these exercises change the main result. In these additional exercises, the fixed cost, ϕ, can be interpreted as the cost of bad health *and long-term care*. The increased probability of being bad health implies people are more likely to have fewer leisure time due to health issue *or long-term care risk*.

More specifically, case 1 uses the counterfactual transition of a matrix of health status shown in table 10 (Baseline transition matrix used in the main manuscript is shown in table 9). Under the counterfactual transition matrix, bad health is more likely. In case 2, the cost of bad health (ϕ) is 1000, whereas the baseline cost is $\phi = 500$. Additionally, the counterfactual transition of a matrix of health status is used in case 2.

Table 9: Baseline Health Transition Matrix

Age		Good in the next year	Bad in the next year
50-59	Good	0.92	0.08
	Bad	0.60	0.40
60-69	Good	0.90	0.09
	Bad	0.60	0.40
70-79	Good	0.85	0.15
	Bad	0.47	0.53

Table 10: Counterfactual Health Transition Matrix

Age		Good in the next year	Bad in the next year
50-59	Good	0.92	0.08
	Bad	0.60	0.40
60-69	Good	0.81	0.19
	Bad	0.53	0.47
70-79	Good	0.76	0.24
	Bad	0.42	0.58

Figure 24 shows the profile of savings worked under baseline, case1, case 2. To make comparison easier, the baseline savings are set to be 1 at each age, and the savings under case 1 and 2 divided by baseline savings are shown. As DeNardi et al. (2010), people aged 50 to 70 accumulate assets when they face severe health and long-term care risk.

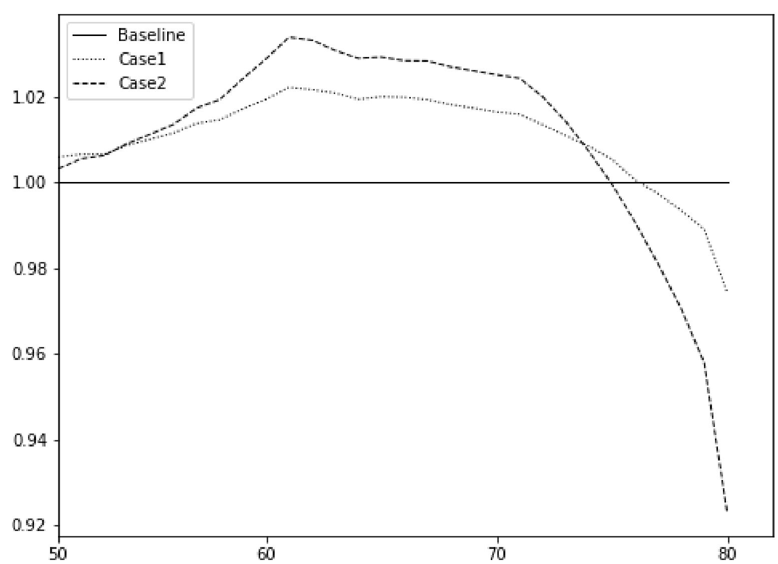


Figure 24: Profile of savings.

Figure 25 and 26 show the profile of hours worked and labor force participation rate under baseline, case1 and case2. Profiles are normalized by the baseline profile. They show that people work more than baseline case when they are aged between 50 ad 60, and they work less than baseline after that.

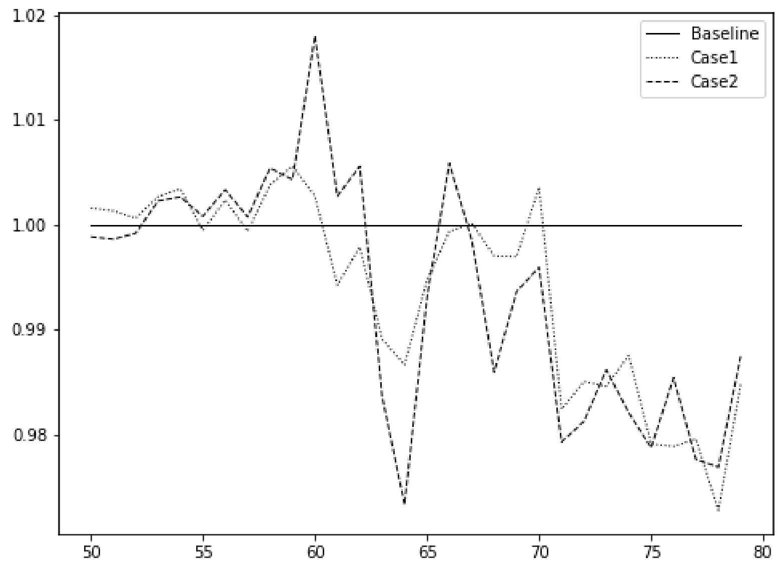


Figure 25: Profile of hours worked

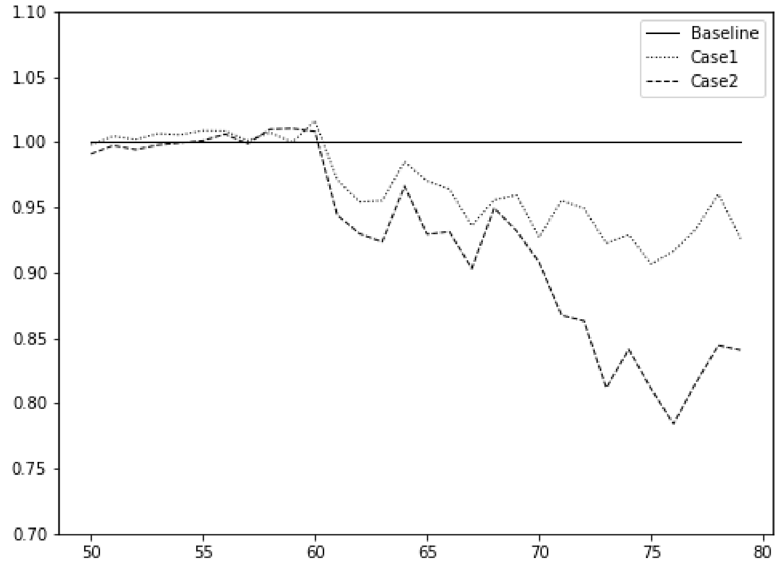


Figure 26: Profile of labor force participation rate.

Figure 27 and 28 show the main result of this paper under case1 and case3. They show that a negative marginal tax rate is welfare-improving. Under both cases, higher long-term care risk does not lower the elasticity of an extensive margin. As shown in figure 25 and 26, change in labor force participation rate is at most 20%, but change in hours worked is at most 2%. Since the elasticity of an extensive margin is large, the negative marginal tax rate is still welfare-improving as explained in section 2.

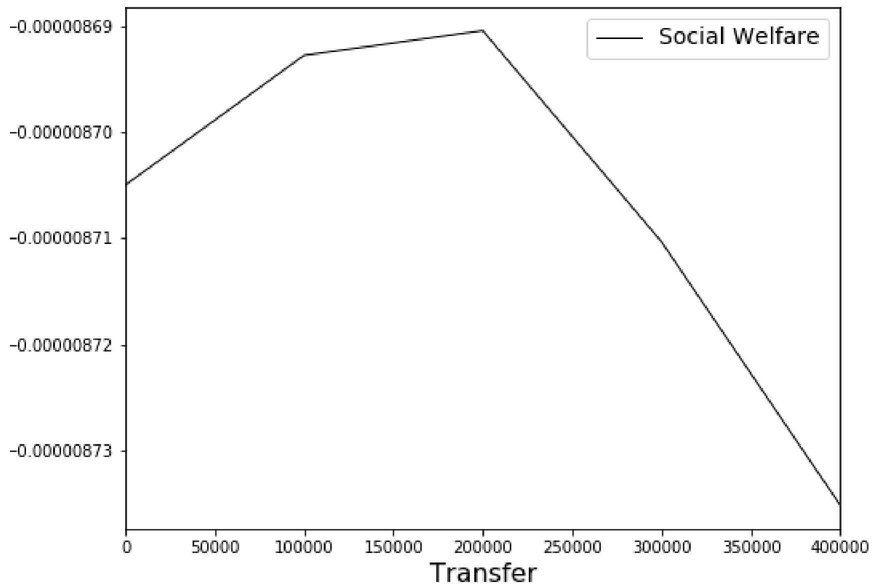


Figure 27: Social welfare function under case 1.

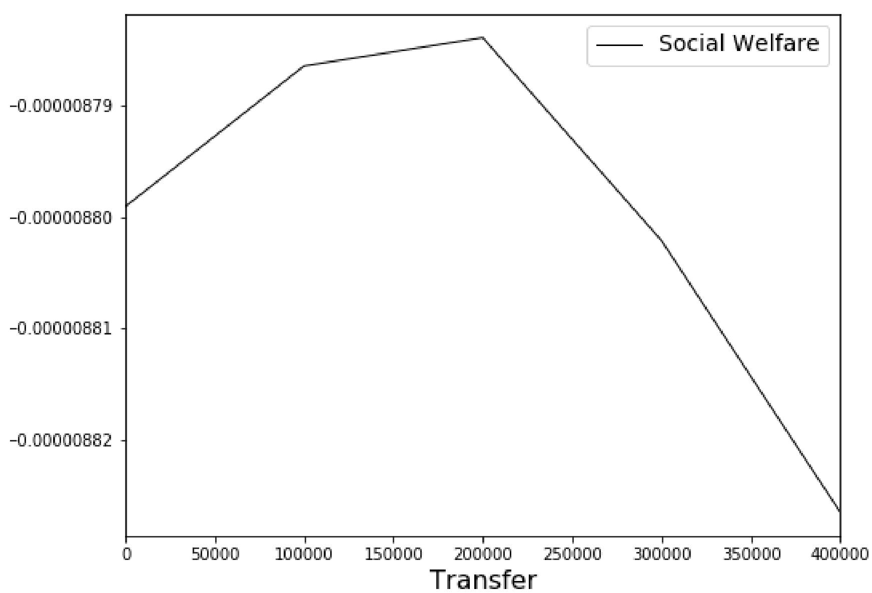


Figure 28: Social welfare function under case 2.