Markovian Social Security in Unequal Societies*

Kaiji Chen† Zheng Song‡

This version: January 28, 2013

Abstract

Can social security systems be supported by a majority population composed of self-interested workers? To address this long-standing issue, we develop a dynamic political-economic theory of social security. We analytically characterize a Markov perfect equilibrium and find that the interaction between Markovian tax policy and tax distortion on private investment in human capital shapes an intertemporal policy rule linking taxes positively over time. By allowing current taxpayers to influence their own future social security benefit, the positive intertemporal tax linkage provides the political support for social security. Moreover, we find that a larger wage inequality weakens the intertemporal tax linkage and, thus, reduces the inter-generational redistributive benefit. This may lead to a negative correlation between wage inequality and the size of a nation's social security system, consistent with the empirical pattern observed across OECD countries.

JEL Classification: E60 H55 P16

Keywords: Intertemporal Tax Linkage, Markov Perfect Equilibrium, Political Economy, Redistribution, Social Security, Wage Inequality.

---

*This paper is based on the fourth chapter of the second author's dissertation at IIES, Stockholm University. We am deeply indebted to Fabrizio Zilibotti for his guidance and encouragement. We thank the editor, Richard Friberg, and three anonymous referees for comments and suggestions. We also thank John Hassler, Dirk Niepelt, Torsten Persson and Kjetil Storesletten for helpful discussions.

†Emory University, Department of Economics, Atlanta, GA 30322. Email: kaiji.chen@emory.edu.

‡University of Chicago, Booth School of Business. 5807 South Woodlawn Ave., Chicago, IL 60637. Email: zheng.michael.song@gmail.com.

1 Introduction

For most OECD countries, unfunded social security programs redistribute resources both across generations (from workers to retirees) and within generations (from the rich to the poor).[1] Since inter- and intra-generational redistributive elements are intertwined through a single policy instrument, the political choice of social security necessarily involves an interaction of redistributive considerations along both dimensions. Exploring this interaction will not only provide new insights on the political-economic mechanism of social security, but also address the puzzling negative cross-country relationship between inequality and the size of welfare states, which has been established by many empirical studies.

This paper, therefore, poses a twofold question: First, how is a nation's social security system sustained? More precisely, why does a majority population comprised of (self-interested) workers contribute to the system, which redistributes to the old at any point in time? Second, how does social inequality affect the size of social security, and can political decisions be reconciled with the observed puzzling correlation between inequality and the size of social security across countries? While some theories have been proposed to address the first question, most of them have either been silent on the second or have delivered predictions opposite to the above empirical observation. In this paper, we offer an answer to the second question as a natural check of the empirical relevance of our proposed theory.

We construct a dynamic political economy in which self-interested forward-looking citizens vote repeatedly for a social security tax in the absence of commitment, reputation mechanism and electoral uncertainty (e.g., probabilistic voting). Two key results emerge from our economy. First, the impact of a current tax on the future decisive voters' redistributive benefit shapes an intertemporal tax linkage that implies positively correlated social security tax rates across time. By allowing the current median voters (taxpayers) to influence their own future social security benefit, the positive intertemporal tax linkage provides incentives for them to support social security. Second, we find that a larger wage inequality weakens the intertemporal tax linkage and, thus, reduces the inter-generational redistributive benefit. This may lead to a negative correlation between wage inequality and the size of a nation's social security program, consistent with the empirical pattern observed across OECD countries.

[1] For example, the original Social Security Act of 1935 in the United States embodied two principles that still guide benefit payment today: Benefits depend on the work history for covered employment, and replace a higher proportion of earnings for low earners.

Our workhorse model is a three-period overlapping-generations economy. Individuals work in the first two periods of their lives, youth and middle-age, and retire when old. They can invest in human capital at birth to increase productivity and, thus, wage income during both working periods. All individuals have linear utility on consumption, and human-capital investment involves a quadratic loss for the young. To incorporate intra-generational social security redistribution, we introduce ex-ante within-cohort heterogeneity by assuming that each individual is born with either high or low ability and, therefore, receives either a high or a low wage rate for each unit of human capital. Taxes are imposed on wage income of the young and the middle-aged to finance social security benefits of the old. During each period, the social security benefit is uniformly distributed across different types of the old, reflecting the intra-generational redistributive feature of social security in unequal societies.

We solve for differentiable Markov perfect equilibria in which the policy rule is a differentiable function of payoff-relevant state variables. In the class of equilibria we consider, the middle-aged agents with a low wage, henceforth referred to as "the middle poor", are always the median voter and are decisive for the political choice on the social security tax level. Since optimal human capital accumulation is linear in the wage rate, the distribution of human capital boils down to a single payoff-relevant state variable, the human-capital stock of the middle poor. As in standard theories, the human-capital stock of the middle poor negatively influences their future social security benefit, from both the inter- and intra-generational redistribution, and, therefore, their tax choice. This relationship establishes a Markovian policy rule. The interaction between the Markovian tax rule and tax distortion on human capital investment creates a positive intertemporal tax linkage: an increase in today's social security tax rate discourages the young's investment in their human-capital and leads to a higher tax rate tomorrow. Rationally perceiving this linkage, the current median voter understands that the more taxes she pays today, the more social security benefit she will receive tomorrow. As a result, she will vote for a positive tax to trade off her current tax burden against her future redistributive benefit.

The intertemporal tax linkage reveals a novel channel through which the intra-generational inequality affects the inter-generational redistributive benefit and, thus, the size of a nation's social security program. In a society with larger wage inequality, the relatively lower wage rate of the poor dampens the impact of the current tax on the human-capital investment of the young poor - the median voter in the next period. Consequently, the future tax rate becomes less responsive to the current rate. The weakened intertemporal tax linkage implies less inter-generational redistributive benefit for the current median voter. Anticipating this, the current median voter will choose a lower tax. We label this effect of wage inequality the "strategic effect." In addition, our model incorporates a standard effect of wage inequality: a larger wage inequality leads to a higher tax since it lowers the human-capital stock of the poor relative to the rich. We refer to this effect as the "redistributive effect."

The long-run impact of wage inequality on the size of social security embodies both strategic and redistributive effects. We find that a more unequal society may actually have a smaller steady-state size of social security if wage inequality is large. This is because the strategic effect dominates as wage inequality becomes sufficiently large. Only after inequality falls below some critical level does the redistributive effect start to overshadow the strategic effect, rendering the correlation positive, as the standard theory implies. Using data on 20 OECD countries and, for each country, computing the average earnings inequality and social security size between 1980 and 2000, we find that countries with smaller earnings inequality have, on average, larger social security expenditures as a percentage of GDP. Our model can thus help to explain the puzzling negative correlation between inequality and social security, which is hard to reconcile in the standard politico-economic models (e.g., Chapter 6, Persson and Tabellini, 2000).

Our paper contributes to the literature on the political sustainability of social security. A key issue in this literature is the temporal separation of contribution and benefit. The current literature addresses the issue by two approaches. The first circumvents the temporal separation problem by assuming that the welfare of benefit recipients weighs somehow in the preference of policymakers.[2] The second approach attempts to construct equilibria in which self-interested taxpayers support the system. However, the literature often restricts the set of possible policy functions to be constant policy functions.[3] This restriction makes the intertemporal tax linkage invariant to important politico-economic factors such as income inequality. In our model, the positive intertemporal tax linkage arises endogenously from the interaction between private decisions and political choices. Aside from its theoretical appeal, the positive intertemporal tax linkage has also empirical relevance: The associated strategic effect implies a negative correlation between wage inequality and social security. By contrast, the restriction to constant policy rules shuts down the strategic effect and, therefore, always implies a positive correlation.

This paper also contributes to the discussion on inequality and the welfare state. Empirical

[2] Examples of this sort include altruism, probabilistic voting (Gonzalez-Eires and Niepelt, 2008, Song, 2011) or gerontocracy (Mulligan and Sala-i-Martin, 1999).

[3] For example, in "once-and-for-all-voting," the initial median voter expects future generations to commit fully to his choice of the tax for at least his lifetime (See, among others, Browning, 1975 and Conesa and Krueger, 1999). In the "trigger strategy," though expectation of future policy choice is based on a system of rewards and punishments in an infinite dynamic game, the choice of future generations is confined to either approving or rejecting the tax rate chosen in the initial period (See, among others, Cooley and Soares, 1999 and Boldrin and Rustichini, 1999).

studies (e.g., Lindert, 1996 and Rodriguez, 1998) have found a negative correlation between income inequality and the size of welfare states across the OECD.[4] Our paper provides additional support for this negative correlation from the perspective of social security taxation and payment, the largest component of government transfers. As a complement to the literature, our theory shows that in the context of social security, the negative correlation can well be explained by the strategic effect through which intra-generational inequality may affect the inter-generational redistributive benefit.[5]

Our work is part of a growing literature on dynamic politico-economic equilibrium, in which current voting may change future political-economic fundamentals and, hence, affect future policy outcomes.[6] The methodology used in this paper is closely related to Hassler, Rodriguez Mora, Storesletten, and Zilibotti (2003, henceforth HRSZ), which analyzes the dynamics of the welfare state in Markov perfect equilibria by allowing closed-form solutions. In their paper, the current median voter can vote strategically to influence the identity of the median voter in the next period. This gives rise to an intertemporal policy linkage similar to that in our model. However, their model economy does not feature a temporal separation of the redistributive contribution and benefit because, by construction, the decisive voter must be the transfer recipient (the old). Therefore, unlike our model, theirs exhibits no strategic effects.

Song (2011) also adopts Markov perfect equilibrium approach to study the evolution of social security. That paper focuses on the dynamic interaction between wealth inequality and social security. Our paper has a fundamentally different goal: the sustainability of a welfare state featuring the temporal separation of redistributive contributions and benefits. In contrast, the sustainability of social security in Song (2011) is essentially guaranteed by probabilistic voting. Moreover, in his paper the absence of the temporal separation of redistributive contributions and benefits naturally shuts down the strategic effect highlighted in this paper. Accordingly, wealth inequality and the size of social security programs are always positively correlated due to the standard redistributive effect. While Song (2011) attempts to explain the simultaneous increase in wealth inequality and the size of social security along the time dimension, in our model the strategic effect is crucial in explaining the observed cross-country

[4] One exception is Tabellini (2000), which, by applying cross-country regression for more than 40 countries, finds that the size of social security systems is positively correlated with income inequality. Persson and Tabellini (2000, Chapter 6), however, note that the measure of inequality is bound to be imperfect for such a large sample of countries.

[5] Benabou (2000), Moene and Wallerstein (2001) focus on roles other than redistribution−say, social insurance under incomplete markets. Koethenbuerger et al. (2008) highlight the "efficiency-redistribution" channel in a static setup.

[6] See, among others, Krusell and Rios-Rull (1999), Azzimonti Renzo (2011), Song (2012) and Song, Storesletten, and Zilibotti (2012).

negative correlation between inequality and the size of social security programs.

The paper is organized as follows. Section 2 describes the economic environment. The political equilibrium is characterized in Section 3. Section 4 examines the impact of wage inequality on social security and discusses the empirical evidence on the relationship between inequality and social security. Section 5 concludes. The appendix contains proofs of the propositions and lemmas.

2 The Model Economy

Consider a small open economy inhabited by an infinite sequence of overlapping generations. Each generation lives three periods, youth, middle-age, and old. An individual works in the first two periods of her life and retires in the last. Labor supply in each of the first two periods is inelastic and normalized to unity. The young can make human-capital investments to increase their labor productivity.

There is heterogeneity within each cohort. Individuals are born with either high or low ability. High- (low-) ability individuals receive a high (low) wage rate per unit of human capital, denoted as w^s (w^u). For notational convenience, agents with high (low) ability will be referred to as the rich (the poor). Let h_t^j be the human-capital investment of a young individual born at time t with type j, $j = s, u$. Human capital and wage income at both working ages equal h_t^j and $w^j h_t^j$, respectively.[7]

We consider a pay-as-you-go social security system. The flat-rate payroll tax rate τ_t is determined through a political process that will be specified below. This flat-rate payroll tax is imposed on working generations to finance the social security benefits. In reality, social security systems contain both actuarial and redistributive components. For analytical convenience, in this dynamic model we take the degree of actuarial fairness as exogenously given, rather than as a political choice. This assumption also captures the idea that, in a social security system the degree of actuarial fairness in reality is more stable over time than the contribution rate, which may well adjust annually.[8] In addition to inter-generational transfers, our pay-as-you-go social security system also bears intra-generational redistributive elements. More specifically, following Conesa and Krueger (1999) and many others, we assume that the social security benefit is evenly distributed among old individuals (i.e. it is "non-actuarially fair").

[7] We could assume that human capital depreciates over time. Then, wage incomes for the middle-aged would be equal to $\delta w^j h_t^j$, where $1-\delta$ is the depreciation rate. The main results would not change under this extension.

[8] See Conde-Ruiz and Profeta (2007) for a static model with political choices on both the contribution rate and the degree of actuarial fairness.

Then, the lifetime wealth A_t^j follows

$$A_t^j = (1-\tau_t)\, w^j h_t^j + \frac{(1-\tau_{t+1})\, w^j h_t^j}{R} + \frac{p_{t+2}}{R^2}, \tag{1}$$

where p_{t+2} stands for the social security benefit per retiree born at time t, and R is the gross interest rate.

To obtain closed-form solutions, we assume agents to have a linear-quadratic preference over lifetime wealth and the costs of human-capital investment:

$$\max_{h_t^j} A_t^j - \frac{1}{2}\left(h_t^j\right)^2, \tag{2}$$

subject to (1). Solving (2) yields

$$h_t^j = \left(1-\tau_t + \frac{1-\tau_{t+1}}{R}\right) w^j. \tag{3}$$

For each type j, human-capital investment increases as a function of the wage rate and decreases as a function of the tax rate. By assuming the linear preference, we actually shut down private saving and therefore the interaction between savings and social security. However, as will be shown below, the human-capital investment in our model plays a similar role as savings in the determination of the size of social security in political equilibria: less human-capital investment (or savings) today leads to more social security benefits tomorrow. Extending the present model by incorporating the intertemporal choice on consumption, therefore, will only complicate the analysis and add no major new insight.[9]

The proportion of the poor is a constant λ in each cohort. We assume $\lambda \geq 1/2$ so that the poor comprise the majority of the population. The weighted average wage incomes for the cohort born at time t, denoted by $\overline{w}^t$, are equal to

$$\overline{w}_t^t = \overline{w}_{t+1}^t = \lambda w^u h_t^u + (1-\lambda)\, w^s h_t^s. \tag{4}$$

The first equality is due to the fact that the middle-aged have the same productivity as when they were young.

Assume that the gross population growth rate is a constant $n > 1$. Plugging (3) into (4), we obtain the output per retiree, which will be useful in the analysis below:

$$\begin{aligned} y_t &= n\overline{w}_t^{t-1} + n^2 \overline{w}_t^t \\ &= n\left(\frac{h_{t-1}^u}{w^u} + n\left(1-\tau_t + \frac{1-\tau_{t+1}}{R}\right)\right). \end{aligned} \tag{5}$$

[9] In an earlier version of this paper, we show that the mechanism emphasized here also applies in a model with private savings, which gives qualitatively the same results as below.

Here, we normalize $\lambda (w^u)^2 + (1-\lambda)(w^s)^2$ to unity so that wage inequality has no first-order effect on the tax base.[10] We also use the fact that $h^s_{t-1}/h^u_{t-1} = w^s/w^u$, as implied by (3). The output per retiree y_t is the current tax base for paying social security benefits. It is also convenient to have the future tax base y_{t+1}, which is determined by h^j_t and h^j_{t+1} and, hence, τ_t, τ_{t+1} and τ_{t+2}.

$$y_{t+1} = \Pi - n\left(\tau_t + \left(n + \frac{1}{R}\right)\tau_{t+1} + \frac{n}{R}\tau_{t+2}\right), \tag{6}$$

where $\Pi \equiv n(1+n)\left(1+\frac{1}{R}\right)$. Note that y_{t+1} is independent of the middle-aged human-capital stock h^j_{t-1} in the current-period. In addition, as will be shown later, the current tax rate τ_t distorts the future tax base, both directly and indirectly via its impact on future tax rates.

We assume that the budget of the social security system must balance in each period. This implies that in each period, the total benefits paid to the old equal the total contributions collected from the working generations:

$$p_t = \tau_t y_t. \tag{7}$$

The size of the social security system, as measured by the total social security benefit as a fraction total output, p_t/y_t, is equal to τ_t in this simple model. Hence, the flip side of the system's sustainability and the size of a country's social security program is the sustainability of a positive τ_t and its level.

The indirect utility functions of the young, the middle-aged, and the old of type j, denoted by $v^{y,j}$, $v^{m,j}$ and $v^{o,j}$, can be written as

$$v^{y,j}(\tau_t, \tau_{t+1}, \tau_{t+2}, \tau_{t+3}) = \frac{\left(\left(1 - \tau_t + \frac{1-\tau_{t+1}}{R}\right) w^j\right)^2}{2} + \frac{\tau_{t+2}\left(\Pi - n\left(\tau_{t+1} + \left(n+\frac{1}{R}\right)\tau_{t+2} + \frac{n}{R}\tau_{t+3}\right)\right)}{R^2} \tag{8}$$

$$v^{m,j}\left(h^j_{t-1}, \tau_t, \tau_{t+1}, \tau_{t+2}\right) = (1-\tau_t) w^j h^j_{t-1} + \frac{\tau_{t+1}\left(\Pi - n\left(\tau_t + \left(n+\frac{1}{R}\right)\tau_{t+1} + \frac{n}{R}\tau_{t+2}\right)\right)}{R}, \tag{9}$$

$$v^{o,j}\left(h^u_{t-1}, \tau_t, \tau_{t+1}\right) = n\tau_t\left(\frac{h^u_{t-1}}{w^u} + n\left(1 - \tau_t + \frac{1-\tau_{t+1}}{R}\right)\right). \tag{10}$$

Note that $v^{o,s} = v^{o,u}$ since social security benefit p_{t+1} are evenly distributed across retirees.

3 Political Equilibrium

In this section we characterize the political equilibrium, in which the social security tax rate τ_t is determined by some political decision process at any time t. The set of equilibria, characterized as the equilibria of a dynamic game played among successive generations of voters, is

[10] Alternatively, we can normalize the average wage rate to unity by letting $w^s = (1-\lambda w^u)/(1-\lambda)$. This leads to no major change in the following results.

potentially large. We restrict attention to Markov perfect equilibria, where τ_t follows a policy rule T contingent on payoff-relevant state variables. For analytical convenience, we further assume T to be non-constant, continuous, and differentiable. In principle, the strategic interaction between private intertemporal choice and policy decisions can switch the identity of the median voter over time. The dynamic political-economic equilibrium is, hence, very hard to characterize (the exceptions are Hassler et al., 2003 and Hassler et al., 2007). The differentiability helps to rule out the time-varying identity of the median voter and, hence, makes the analysis substantially simpler.[11] Restricting T to be a non-constant function rules out the trivial equilibrium $T(\cdot) = 0$ for all $h^u \in \left[\underline{h}, \overline{h}\right]$. The corresponding equilibrium is referred to as the differentiable Markov perfect equilibrium (DMPE henceforth).[12]

3.1 Differentiable Markov Perfect Equilibrium

There are two state variables at time t, human-capital stock h^s_{t-1} and h^u_{t-1}. Since $h^s_{t-1}/h^u_{t-1} = w^s/w^u$, we can, without loss of generality, confine the payoff-relevant state variable to h^u_{t-1}. The Markovian policy rule T can, thus, be written as as a function of h^u_{t-1} only:

$$\tau = T\left(h^u_{-1}\right), \tag{11}$$

where $T : \left[\underline{h}, \overline{h}\right] \to [\underline{\tau}, \bar{\tau}]$, $\underline{h} \equiv w^u\left(1 + 1/R\right)\left(1 - \bar{\tau}\right)$ and $\overline{h} \equiv w^u\left(1 + 1/R\right)\left(1 - \underline{\tau}\right)$ are the lower and upper bound of h^u_{-1}, respectively. We drop the time subscript when there is no source of confusion. We assume that τ cannot exceed 1, and is non-negative. So, $\bar{\tau} \leq 1$ and $\underline{\tau} = 0$. Plugging (11) into (3) with $j = u$, we have

$$h^u = w^u\left(1 - \tau + \frac{1 - T\left(h^u\right)}{R}\right). \tag{12}$$

Given T, equation (12) solves the human capital investment decision of the young poor $H : [\underline{\tau}, \bar{\tau}] \to \left[\underline{h}, \overline{h}\right]$.

A combination of T and H yields

$$\tau' = T \circ H\left(\tau\right) \equiv B\left(\tau\right), \tag{13}$$

[11] Song (2005, chapter 4) proves that the identity of the median voter is constant over time in DMPE. The intuition is as follows. When the identity of the median voter changes over time, we will observe different political regimes in equilibrium. This leads to a discontinuous or non-differentiable policy rule T. Therefore, the technical restriction that the policy function must be differentiable actually amounts to imposing a constant identity of the decisive voters over time.

[12] The assumption of differentiability has often been adopted in recent studies on social security (e.g., Azariadis and Galasso, 2002, Forni, 2005) and fiscal policy (e.g., Klein, Krusell, and Rios-Rull, 2008, Azzimonti Renzo, 2011). Krusell and Smith (2003) show that there could be, in principle, an infinitely large number of non-differentiable Markov equilibria. Markov equilibria with non-differentiable strategies may exist in our model. We will, however, not attempt to characterize such equilibria.

where $B : [\underline{\tau}, \overline{\tau}] \to [\underline{\tau}, \overline{\tau}]$, representing an intertemporal policy rule that links the future tax choice to the current one. The intertemporal tax linkage, endogenously obtained by the interaction between the Markovian tax rule and the tax distortion on human capital investment, will serve as the cornerstone of our analysis below.

Despite no commitment to future policy outcomes, the presence of the intertemporal tax linkage allows the current policy decision maker to indirectly influence future political decisions. Specifically, the derivative of B captures the degree to which the future tax responds to the current one:

$$B'(\tau) = T'(H(\tau))\, H'(\tau)\,. \tag{14}$$

We refer to $B'(\cdot)$ as the magnitude of the intertemporal tax linkage. Clearly, the larger is $B'(\cdot)$, the more easily the current tax can influence the future one.

Now we specify the political decision process. Consider a Downsian electoral competition in which there are two candidates (or parties) with the aim of winning the election.[13] The political choice on τ solves

$$\tau = \arg \max_{\tau \in [\underline{\tau}, \overline{\tau}]} v^{dec}, \tag{15}$$

subject to $\tau' = B(\tau)$. The term v^{dec} is the indirect utility function of the decisive voter.

Equations (9) and (10) show that the human capital of the middle poor, h^u_{-1}, affects the political choice of τ by both the middle-aged and the old.[14] Hence, without loss of generality, we define $\tau = \tilde{T}\left(h^u_{-1}\right)$ as the solution of equation (15). T is an equilibrium policy rule if and only if $T = \tilde{T}$. The definition of the equilibrium is given by

Definition 1 *A differentiable Markov perfect political equilibrium (DMPE) is a pair of differentiable functions $\langle T, H \rangle$, where $T : \left[\underline{h}, \overline{h}\right] \to [\underline{\tau}, \overline{\tau}]$ is the policy rule for the social security tax rate and $H : [\underline{\tau}, \overline{\tau}] \to \left[\underline{h}, \overline{h}\right]$ is a private decision rule for human-capital investment. T and H solve the following functional equations:*

(1) $T\left(h^u_{-1}\right) = \arg\max_{\tau \in [\underline{\tau}, \overline{\tau}]} v^{dev}$, subject to $\tau' = (T \circ H)(\tau)$.
(2) $H(\tau) = w^u \left(1 - \tau + \frac{1 - T \circ H(\tau)}{R}\right)$.

3.2 The Median Voter

This subsection shows that under certain conditions, the middle poor are always the median voter in our model economy. We first argue that the middle poor are decisive, and then check the validity of this argument.

[13] See, for example, Chapter 2 of Persson and Tabellini (2000).
[14] For the young, there is no payoff-relevant state variable for their political choice of τ.

We will focus on economies in which the population of the old and middle poor is always larger than the population of the young and middle rich. This implies $1+\lambda n \geq (1-\lambda)\, n+n^2$ or $\lambda \geq (n\,(1+n)-1)\,/2n$. Notice that in an economy with a stationary population (i.e., $n=1$), the condition reduces to $\lambda \geq 1/2$. Let $\tau^{y,j}$, $\tau^{m,j}$, and τ^{o} be the preferred tax rate of the type-j young, middle, and the old, respectively, under the expectation that the middle poor will be decisive in the future. Given that the middle-aged comprise the majority of voters and the poor comprises the majority of each cohort, the following condition is sufficient for the middle poor to be the median voter for any $h_{-1}^{u} \in \left[\underline{h}, \overline{h}\right]$:

$$\tau^{y,u} \leq \tau^{m,s} \leq \tau^{m,u} \leq \tau^{o}. \tag{16}$$

The second inequality is straightforward. Since the middle rich receive the same social security benefit as the middle poor while paying more taxes, they always prefer a lower tax rate than the middle poor.[15] So, we only need to check the first and the third inequality in (16). We will show in Section 3.4 below that for a wide range of parameter values, both $\tau^{m,u} \leq \tau^{o}$ and $\tau^{y,u} = \underline{\tau}$ will hold. The intuition is simple. As taxpayers, the middle-aged prefer a lower tax rate than the old, who bear no tax burden.[16] Meanwhile, the marginal cost of taxation for the young tends to be higher than for the middle-aged. This is because increasing taxes for the young reduces not only their after-tax wage income, but also their before-tax income by distorting human capital investment, both directly and indirectly via τ'. By contrast, since the human capital stock of the middle-aged is fixed, increasing taxes only lowers their after-tax income.

3.3 Political Choice

We now investigate how the tax rate is chosen by the middle poor. For notational convenience, we write the future tax base (6) in a recursive formulation:

$$y' = Y(\tau) \equiv \Pi - n\left(\tau + \left(n + \frac{1}{R}\right) B(\tau) + \frac{n}{R} B(B(\tau))\right). \tag{17}$$

Differentiating $Y(\tau)$ yields that $Y'(\tau) = -n\,(1 + (n+1/R)\, B'(\tau) + nB'(B(\tau))\, B'(\tau)\,/R) < 0$, irrespective of the value of $B'(\tau)$. That is to say, increasing τ always reduces the future tax base for the social security benefit payment.

The tax choice of the middle poor can, thus, be expressed as

$$\max_{\tau \in [\underline{\tau}, \bar{\tau}]} (1-\tau)\, w^{u} h_{-1}^{u} + \frac{1}{R} B(\tau)\, Y(\tau). \tag{18}$$

[15] For a similar reason, we have $\tau^{y,s} \leq \tau^{y,u}$.

[16] Moreover, the current tax base y for redistribution towards the old is less elastic with respect to τ than the future tax base y' for redistribution towards the current middle-aged.

The first term in the objective function is the after-tax wage income, and the second term represents the discounted future benefit from social security. The first-order condition is

$$w^u h_{-1}^u = \frac{1}{R}\left(B'(\tau) Y(\tau) + B(\tau) Y'(\tau)\right), \tag{19}$$

where the left-hand side (right-hand side) of (19) captures the marginal cost (benefit) of taxation for the middle poor. According to (19), the middle poor choose a tax rate that achieves a trade-off between their current tax burden and their future redistributive benefit. We suppress two multipliers associated with the constraints $\tau \geq \underline{\tau}$ and $\tau \leq \bar{\tau}$. As we show in the next section, inner solutions are obtained under some parameter restriction.

Equation (19) reveals the necessary conditions for the sustainability of social security. First, as $Y'(\cdot) < 0$, one can see from (19) that $B'(\cdot) > 0$ is necessary for $\tau > \underline{\tau}$. As taxpayers, the middle poor would vote for imposing the lowest tax rate on themselves if their future benefit were independent of or even negatively correlated with their contribution. Hence, a positive intertemporal tax linkage is a prerequisite for sustaining social security.

3.4 Equilibrium Policy Rules

In this section, we provide sufficient conditions for the existence of DMPE and then characterize the equilibrium. Thanks to the quasi-linear preference, closed-form solutions for equilibrium policy rules can easily be obtained.

Proposition 1 *Assume that*

$$\phi_0 + \phi_1 \overline{h} > \underline{\tau}, \tag{20}$$

$$\phi_0 + \phi_1 \underline{h} < \bar{\tau}, \tag{21}$$

$$\phi_0 + \phi_1 \underline{h} \leq \frac{1 + (1 - b_0)/R + \underline{h}/(n w^u)}{2(1 + b_1/R)}, \tag{22}$$

and

$$\begin{aligned} &\left(1 - \overline{\tau} + \frac{1 - b_0 - b_1 \overline{\tau}}{R}\right)(w^u)^2 \left(1 + \frac{b_1}{R}\right) \\ > \quad &\left[(\pi_0 + \pi_1 b_0) b_1^2 + (b_0 + b_1 b_0)\pi_1 b_1 + 2\pi_1 b_1^3 \tau_t\right]/R^2, \end{aligned} \tag{23}$$

where $b_1 > 0$*,* $\pi_0 > 0$*,* $\pi_1 < 0$*,* $\phi_0 \equiv -(b_1\pi_0 + b_0\pi_1)/(2b_1\pi_1) > 0$ *and* $\phi_1 \equiv R w^u/(2b_1\pi_1) < 0$ *(see the Appendix 6.1 for the definition of* b_0*,* b_1*,* π_0 *and* π_1*). Then, there exists a DMPE such* *that the median voter is the middle poor and*

$$T\left(h_{-1}^{u}\right) = \phi_0 + \phi_1 h_{-1}^{u}, \quad (24)$$

$$H(\tau) = h_0 + h_1 \tau, \quad (25)$$

$$B(\tau) = b_0 + b_1 \tau, \quad (26)$$

$$Y(\tau) = \pi_0 + \pi_1 \tau, \quad (27)$$

where $h_0 = \frac{w^u}{1+w^u\phi_1/R}\left(1+\frac{1-\phi_0}{R}\right) > 0$, $h_1 = -\frac{w^u}{1+w^u\phi_1/R} < 0$.

Proof. See the appendix. ■

The first part of the proposition describes technical assumptions for the existence of a DMPE, in which the middle poor are always the median voter. Equations (20) and (21) are assumptions made to ensure interior solutions, i.e., $T\left(h_{-1}^{u}\right) \in (0, \bar{\tau})$ for any $h_{-1}^{u} \in \left[\underline{h}, \overline{h}\right]$.[17] Equation (22) is a sufficient condition for the third inequality in (16) to be satisfied, i.e., the old always prefer a higher tax rate than the middle poor. Equation (23) is a sufficient condition for $\tau^{y,u} = 0$ and, thus, the first inequality in (16) to be satisfied. In other words, (22) and (23) ensure that the middle poor are always the decisive voter.[18] Therefore, an economy satisfying these assumptions can sustain a social security system in a DMPE.

Conditions (20), (21), (22), and (23) can be met under a wide variety of combinations of n and w^u. A numerical example is plotted in Figure 1, where we set $R = 1.025^{30}$, $\underline{\tau} = 0$, and $\bar{\tau} = 0.9$. Condition (20) is satisfied for all w^u and n above the solid line. Condition (21) fails to hold only for very large values of n (e.g., $n > 20$, which corresponds to an annual population growth rate of 16 percent if each period in our model corresponds to 20 years). The dotted and the dashed lines in the graph plot the threshold condition of n implied by (22) and (23), respectively. For any n above the two lines, the corresponding threshold condition is satisfied. Hence, the set of n and w^u allowing the middle poor to be the median voter is captured by the shaded area in Figure 1. It is immediately apparent that under a large set of n and w^u the middle poor is the median voter. Note that social security can be sustained in a dynamic efficient economy ($R > n$), as long as wage inequality is not too low, i.e., w^u is not too high. Intuitively, the political sustainability of social security originates from both inter-generational and intra-generational redistributive incentives. In a more unequal society, the poor are more willing to support social security for the purpose of intra-generational redistribution.

[17] Since $T'(\cdot) < 0$, the minimum (maximum) value of τ_t is located at $h_{t-1} = \overline{h}$ ($\underline{h}$). Accordingly, given (20), the middle poor would like to choose tax rates higher than $\underline{\tau}$ for any $h_{t-1} \in \left[\underline{h}, \overline{h}\right]$. Also, (21) ensures that $\tau_t \leq \bar{\tau}$ is not binding.

[18] Note that (22) is more likely to hold by a higher $\underline{h}$ (or a lower $\overline{\tau}$). This is because $\tau^{m,u}$ decreases in h_{-1}, while the old tend to impose a higher tax rate with a larger inelastic human capital stock h_{-1}.

The second part of the proposition characterizes our DMPE. First of all, both ϕ_1 and h_1 are negative. The negative values of h_1 or $H'(\cdot)$ are straightforward: a distortionary tax discourages human-capital investment. Also, the negative values of ϕ_1 or $T'(\cdot)$ simply result from the fact that the human-capital stock of the middle poor negatively affects their future social security benefit and, therefore, their tax choice in the current period.

The combination of $H'(\cdot) < 0$ and $T'(\cdot) < 0$ establishes a positive intertemporal tax linkage, the key to sustaining social security in our model economy. To see the underlying political-economic mechanism more clearly, consider an increase in today's tax rate. $H'(\cdot) < 0$ leads the young to make a lower human-capital investment. Then $T'(\cdot) < 0$ implies a higher tax rate in the next period due to more redistributive benefits for the next period's decisive voters (the current young). Rationally perceiving the linkage, the current middle poor understand that the more taxes they pay today, the more social security benefits they will receive tomorrow. This provides the incentive for the current middle poor to choose a positive tax rate.

The following corollary characterizes the dynamics of the size of the social security system.

Corollary 1 *Assume equations (20), (21) and (22) hold. In the DMPE, we have* $b_0 > 0$ *and* $b_0/(1-b_1) \in (0, \overline{\tau})$. *The social security tax rates monotonically converge to the steady state* $b_0/(1-b_1)$.

Proof. See the appendix. ■

Suppose that the political process that decides on social security occurs unexpectedly at time 0. Then, the above corollary predicts an increasing sequence of social security tax rates over time, which converge to the steady state rate $b_0/(1-b_1)$. To see this, note that the middle poor at time 0 invested more in human capital than the middle poor in subsequent periods, due to an expectation of zero taxes before time 0. Consequently, the initial tax rate chosen by the initial middle poor is actually the lowest in the dynamics. Moreover, the initial positive tax rate distorts the human-capital investment of the second-period middle poor, with the result that they end up with less human capital than the initial-period middle poor. This encourages the second-period middle poor to choose a tax rate higher than the initial one. A similar argument applies to the political choice afterwards and, thus, explains why tax rates in later periods increase monotonically.

The reason why the social security tax rate converges to a steady state is essentially that the marginal future redistributive benefit, the slope of the Laffer curve, decreases with τ.[19] Accordingly, when τ becomes sufficiently large along the transition, the marginal cost of taxation

[19] Algebraically, the marginal effect of τ on the slope of Laffer curve is $\frac{\partial \tau' Y'}{\partial \tau^2} = 2b_1\pi_1 < 0$.

tends to outweigh the marginal future benefit. Technically, this is guaranteed by assumption (21), which implies that at $\tau = \overline{\tau}$, for tomorrow's median voter, the marginal future tax burden $w^u H(\overline{\tau})$ is higher than the marginal future redistributive benefit if he chose $\tau' = \overline{\tau}$. As a consequence, tomorrow's median voter will choose a tax rate smaller than the current one, i.e., $\tau' = B(\overline{\tau}) < \overline{\tau}$. The differentiability of policy rules then implies that as τ becomes sufficiently large, the future median voter has no incentive to further increase taxes, which ensures that τ does not diverge.

4 Effects of Wage Inequality

Now, we will address this paper's second key question: how wage inequality affects the size of social security. The answer to this question provides testable implications, which can be confronted with empirical observations. There are two steps in this analysis. First, we show that wage inequality has two opposite effects on the size of social security. We then ask how through these two effects wage inequality influences the tax choice at the steady state.

4.1 The Strategic Effect and the Redistributive Effect

It is useful to define a benchmark social security tax rate:

$$\hat{\tau} \equiv \phi_0 + \phi_1 \hat{h}^u, \tag{28}$$

where $\hat{h}^u = w^u (1 + 1/R)$ is the first-best human capital stock of the middle poor. We refer to $\hat{\tau}$ as the "baseline tax rate," which is the minimum tax rate under the Markovian tax rule T. Later on, we will show that the actual tax rate in the steady state is equal to the product of the baseline tax rate and a multiplier.

The impact of wage inequality on the baseline tax rate can be expressed as

$$\frac{\partial \hat{\tau}}{\partial w^u} = \underbrace{\frac{\partial \phi_0}{\partial w^u} + \frac{\partial \phi_1}{\partial w^u} \widehat{h}^u}_{\text{the strategic effect}} + \underbrace{\phi_1 \frac{\partial \widehat{h}^u}{\partial w^u}}_{\text{the redistributive effect}}. \tag{29}$$

Equation (29) shows that wage inequality affects the baseline tax rate $\hat{\tau}$ through two channels, as represented by the two arguments on the right-hand side of (29). The first channel is via ϕ_0 and ϕ_1 in the Markovian tax rule T, while the second is via its effect on human capital stock.

We first explore the second channel. It is straightforward that $\widehat{h}^u$ is increasing in w^u. Since $\phi_1 < 0$, the positive impact of w^u on $\widehat{h}^u$ implies that an increase in w^u tends to reduce the baseline tax rate. Intuitively, a lower w^u yields a lower first-best human-capital investment, which leads to more redistributive benefit and, therefore, a higher preferred tax rate for tomorrow's decisive voters. This is referred to as the "redistributive effect," which has been extensively investigated in the standard political-economic theory.

The effect of w^u through ϕ_0 and ϕ_1, two parameters governing the Markovian tax rule, is a novel feature of our model. This channel actually implies that wage inequality has a negative effect on the size of social security program. To see this, we first explore the effect of w^u on b_1, a key determinant for ϕ_0 and ϕ_1 and, thus, the inter-generational redistributive benefit. The following lemma reveals a negative correlation between wage inequality and the magnitude of the intertemporal tax linkage b_1.

Lemma 1 *Assume that equation (20), (21) and (22) hold. Then,* $\partial b_1/\partial w^u > 0$.

Proof. See the appendix. ■

A larger wage inequality lowers the absolute value of h_1, the marginal impact of the current tax on the young poor's human-capital investment. This is the first-order effect of w^u on b_1, resulting in a weakened intertemporal tax linkage. Therefore, Lemma 1 establishes a negative effect of wage inequality on b_1 and, hence, on the inter-generational redistributive benefit. This feature helps to explain how the Markovian tax rule T, captured by ϕ_0 and ϕ_1, responds to a change of w^u.

Proposition 2 *Assume that equation (20), (21) and (22) hold. Then,* $\partial \phi_0/\partial w^u > 0$ *and* $\partial \phi_1/\partial w^u > 0$.

Proof. See the appendix. ■

Proposition 2, together with equation (29), shows that given human-capital stock h^u_{-1}, wage inequality has a negative effect on the choice of τ. This is because an increase in wage inequality can actually reduce the inter-generational redistributive benefits for the current policy decision makers by weakening the intertemporal tax linkage. Anticipating this, the current policy decision maker will choose a lower tax rate. We label this negative impact of wage inequality on τ via equilibrium policy rules, captured by ϕ_0 and ϕ_1 in (29), as the "strategic effect."

The overall effect of wage inequality on the baseline tax rate, as captured by the sign of $\partial\hat{\tau}/\partial w^u$, depends on which of the above two effects dominates. As Panel A of Figure 2 shows, $\hat{\tau}$ is hump-shaped in w^u. In other words, when wage inequality is large, there exists a negative relationship between wage inequality and the baseline tax rate. Only after inequality falls below some critical level does the correlation between inequality and the baseline tax rate become positive. Panel B of Figure 2 plots the relative magnitude of the two effects for different levels of wage inequality. We see that when w^u is small, the magnitude of the strategic effect is large and dominates the redistributive effect. As w^u keeps increasing, the strategic effect becomes smaller. This is because the current tax becomes more distortionary due to the larger b_1: an increase of today's tax leads to a larger increase of tomorrow's tax, which distorts further the current human-capital investment. Consequently, the redistributive effect starts to dominate when w^u becomes sufficiently large.

4.2 The Steady-State Size of Social Security

We now extend the analysis to the impact of wage inequality on the steady-state payroll tax rate. The linear intertemporal tax linkage (26) delivers a steady-state tax rate of

$$\tau^* \equiv \frac{b_0}{1-b_1}. \tag{30}$$

Note that b_0 represents the component of the next-period tax that is independent of the current tax. Equation (44) in the appendix shows that

$$b_0 = \hat{\tau}\left(1+\frac{b_1}{R}\right). \tag{31}$$

The first component on the right-hand side of (31) is the baseline tax rate given in (28). The second component is a multiplier of the baseline tax rate, representing the repercussions of a future tax on itself via human-capital investment. Specifically, an increase in future taxation distorts human-capital investment today, which translates into a lower human-capital stock and, thus, a higher tax rate tomorrow.

Substituting equation (31) into equation (30), we obtain

$$\tau^* = \hat{\tau}\frac{(1+b_1/R)}{1-b_1}. \tag{32}$$

Similar to b_0, τ^* can be decomposed into two components. The first component on the right-hand side of (32) is, again, the baseline tax rate. The second component is a multiplier of the baseline tax rate, representing the repercussions of the current tax on τ^*. Specifically, with a positive b_1, a marginal increase in the current tax leads to a higher future tax rate by discouraging human-capital investment today. The higher future taxation discourages human-capital investment further and, thus, feedbacks into a higher steady-state tax rate.

The overall effects of wage inequality can be decomposed as follows:

$$\frac{\partial \tau^*}{\partial w^u} = \frac{(1+b_1/R)}{1-b_1}\frac{\partial \hat{\tau}}{\partial w^u} + \frac{\hat{\tau}\,(1+1/R)}{(1-b_1)^2}\frac{\partial b_1}{\partial w^u}. \tag{33}$$

The first term on the right-hand side of (33) captures the effect of w^u on the baseline tax rate, and the second term is the effect of w^u via the multiplier. The second term is always positive, while the sign of the first term is ambiguous due to the two opposite effects of w^u on $\hat{\tau}$. In particular, we find that in our numerical experiments, τ^* is always non-monotonically related to w^u. Figure 3 shows an inverted-U shaped correlation between w^u and τ^*. The hump shape is due primarily to the hump shape of the baseline tax rate depicted in Figure 2.

Finally, it is worth mentioning how the demographic structure affects the size of social security in our model. Following the same procedure as in the proof of Lemma 1, one can easily establish that $\partial b_1/\partial n < 0$. Although similar to a lower w^u, a larger n weakens the intertemporal tax linkage, it directly increases the inter-generational redistributive benefit. Our numerical simulation suggests that the direct effect always dominates the strategic effect via b_1. Therefore, our model delivers results that are qualitatively similar to those in the standard theory: a high dependency ratio leads to a larger social security program.[20]

4.3 Empirical Evidence

Can the above-mentioned theoretical predictions be consistent with the empirical observations? In this section, we provide cross-country evidence on the correlation between wage inequality and the size of a nation's social security program. There are a number of empirical studies addressing the question of whether higher distributional skewness makes the constituency favor more redistributive legislation, following the work pioneered by Meltzer and Richard (1981). Our perspective is slightly different from the literature, with a focus on the correlation between the size of social security and earnings inequality, a close proxy of wage dispersion.

To deal with problems of data quality and comparability, we limit our cross-country analysis to a set of 20 OECD countries with reliable data on both earnings inequality and social security transfers. OECD national account statistics provide the ratio of the median earnings to the upper limit of earnings of employees in the bottom decile of the earnings distribution, denoted by D5/D1, in a total of 21 countries.[21] Social security transfers (SST), measured as a percentage share of GDP, are available from the OECD's Economic Outlook.[22] A combination

[20] Moreover, following the same procedure as in the proof of Lemma 1, one can easily establish that $\partial b_1/\partial R > 0$. However, a larger R reduces the present value of future redistributive benefits and, thus, tends to reduce the tax rate. Our numerical results suggest that unless wage inequality is very large (w^u very small), the latter effect dominates so that an increase R reduces the steady state τ.

[21] Earnings are measured on a gross basis, i.e., before deducting income taxes and social security contributions paid by workers. Earnings include base wages and salaries, overtime payments, bonuses and gratuities, extra monthly payments, and regular and irregular allowances, but may exclude elements of the remuneration package of managers and other executives, such as stock options.

[22] The data for social security expenditure as a percentage of GDP until 1969 are from OECD Historical Statistics, various years. From 1970 onwards, the data are from OECD, National Accounts Statistics, Historical

of the two data sources gives a set of 20 OECD countries.[23] We then compute the average earnings inequality and the average social security transfers for each country between 1980 and 2000, the period to which most of the observations belong.

Consistent with the literature, our evidence suggests a negative correlation between wage inequality and the size of social security. As shown by Figure 4, the social security expenditure in countries that have smaller earnings inequality constitutes, on average, a larger percentage of GDP than in countries having larger earnings inequality, with a correlation coefficient of -0.40.[24] Such a pattern, while it is hard to explain by the standard theory, is in line with our model's prediction.

5 Conclusion

This paper develops a dynamic political-economic theory of social security to address two questions. First, how is social security sustained by a majority population comprised of (self-interested) working-age generations? Second, can political decisions be reconciled with the puzzling negative cross-country correlation between inequality and the size of social security program? To this end, we introduce wage inequality in an environment that features the absence of altruism, commitment, reputation mechanism, and electoral uncertainty. And we analytically characterize a Markov perfect equilibrium with the median voter being the middle poor.

Our major finding is that the interaction between Markovian tax policy and tax distortion on human capital investment shapes an intertemporal policy rule that links social security taxes positively over time. This intertemporal policy linkage not only serves as the key element in the political sustainability of social security, but also reveals a novel channel through which the intra-generational inequality negatively affects the inter-generational redistributive benefits and, thus, the size of social security. Thanks to the presence of this novel channel, our theoretical predictions are in line with the empirical pattern across OECD countries. Our work, therefore, contrasts sharply with most previous studies, which imply opposite predictions by resorting to constant policy rules for sustaining social security in similar environments.

Furthermore, the mechanism described in this paper has the potential to explain other social programs that are characterized by the temporal separation of contributions and benefits.

Statistics - Structure or composition of certain economic aggregates, Vol. 2002, release 01.

[23] The sample countries include Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, the United Kingdom and the United States.

[24] In a linear regression of the size of social security regressed against earnings inequality, the estimated coefficient is negative and statistically significant at 10 percent.

One example is the government-sponsored health insurance program. An investigation into the sustainability and the evolution of such programs might be fruitful as they are growing fast and continue to become an increasingly large portion of government expenditures. Another equally important policy issue is government debt. A key question is what prevents current generations from exploiting future unborn generations by issuing debt (see, e.g., Song, Storesletten, and Zilibotti, 2012)? It will be interesting for future researchers to incorporate government borrowing into the current setup to see how public debt interacts with social security.

References

[1] Azzimonti, Marina (2011), "Barriers to Investment in Polarized Societies," *American Economic Review*, 101, 2182-2204.

[2] Azariadis, Costas and Vincenzo Galasso (2002), "Fiscal Constitutions," *Journal of Economic Theory*, 103, 255-281.

[3] Benabou, Roland (1996), "Inequality and Growth," in Ben S. Bernanke and Julio J. Rotemberg, eds., *NBER Macro Annual*, Vol. 11, 11-74, Cambridge, MA: The MIT Press.

[4] Benabou, Roland (2000), "Unequal Societies: Income Distribution and the Social Contract," *American Economic Review*, Vol. 90, 96-129.

[5] Boldrin, Michele and Aldo Rustichini (2000), "Political Equilibria with Social Security," *Review of Economic Dynamics*, 3, 41-78.

[6] Conde-Ruiz J. Ignacio and Paola Profeta (2007), "The Redistributive Design of Social Security System," *Economic Journal*, 117, 686-712.

[7] Conesa, Juan C. and Dirk Krueger (1999), "Social Security Reform with Heterogeneous Agents," *Review of Economic Dynamics*, 2, 757-795.

[8] Cooley, Thomas F. and Jorge Soares (1999), "A Positive Theory of Social Security Based on Reputation," *Journal of Political Economy*, 107, 135-160.

[9] Cremer, Helmuth, Philippe De Donder, Dario Maldonado and Pierre Pestieau (2007), "Voting over Type and Size of a Pension System When Some Individuals are Myopic," *Journal of Public Economics*, 91, 2041-2061.

[10] Forni, Lorenzo (2005), "Social Security as Markov Equilibrium in OLG Models," *Review of Economic Dynamics*, 8, 178-194.

[11] Gonzalez-Eiras, Martin and Dirk Niepelt (2008), "The Future of Social Security," *Journal of Monetary Economics*, 55(2), 197-218.

[12] Hassler, John, Jose V. Rodriguez Mora, Kjetil Storesletten and Fabrizio Zilibotti (2003), "The Survival of the Welfare State," *American Economic Review*, 93, 87-112.

[13] Hassler, John, Kjetil Storesletten and Fabrizio Zilibotti (2007), "Democratic Public Good Provision," *Journal of Economic Theory*, 133, 127-151.

[14] Klein, Paul, Per Krusell and Jose-Victor Rios-Rull (2008), "Time-Consistent Public Policy," *Review of Economic Studies*, 75, 789-808.

[15] Koethenbuerger, Marko, Panu Poutvaara and Paola Profeta (2008), "Why Are More Redistributive Social Security Systems Smaller? A Median Voter Approach," *Oxford Economic Papers*, 60, 275-292.

[16] Krusell, Per and Jose-Victor Rios-Rull (1999), "On the Size of US Government: Political Economy in the Neoclassical Growth Model," *American Economic Review*, 80, 1156-1181.

[17] Krusell, Per and Anthony A. Smith, Jr. (2003), "Consumption-Savings Decisions with Quasi-Geometric Discounting," *Econometrica*, 71, 365-375.

[18] Lindert, Peter H. (1996), "What Limits Social Spending," *Explorations in Economic History*, 33, 1-34.

[19] Meltzer, Allan and Scott Richard (1981), "A Rational Theory of the Size of Government," *Journal of Political Economy*, 89, 914-927.

[20] Moene, Karl O. and Michael Wallerstein (2001), "Inequality, Social Insurance, and Redistribution," *American Political Science Review*, Vol. 95, 859-974

[21] Mulligan, Casey B. and Xavier Sala-i-Martin (1999), "Gerontocracy, Retirement, and Social Security," *NBER Working Paper*, No. 7117.

[22] Perotti, Roberto (1996), "Growth, Income Distribution, and Democracy: What the Data Say," *Journal of Economic Growth*, 1, 149-87.

[23] Persson, Torsten and Guido Tabellini (2000), *Political Economics: Explaining Economic Policy*, MIT Press, Cambridge, MA.

[24] Rodriguez, Francisco (1998), "Essays on Redistribution, Development and the State," Ph.D. Dissertation, Harvard University.

[25] Song, Zheng (2005), "Essays on Dynamic Political Economy," PhD Dissertation, IIES, Stockholm University.

[26] Song, Zheng (2011), "The Dynamics of Inequality and Social Security in General Equilibrium," *Review of Economic Dynamics*, 14, 613-635.

[27] Song, Zheng (2012), "Persistent Ideology and the Determination of Public Policies over Time," *International Economic Review*, 53, 175-202.

[28] Song, Zheng, Kjetil Storesletten and Fabrizio Zilibotti (2012), "Rotten Parents and Disciplined Children: A Dynamic Politico-Economic Theory of Public Expenditure and Debt," 80(6), *Econometrica*, 2785-2803.

[29] Storesletten, Kjetil, Christopher I. Telmer and Amir Yaron (2004), "Consumption and Risk Sharing over the Life Cycle," *Journal of Monetary Economics*, 51, 609-633.

[30] Tabellini, Guido (2000), "A Positive Theory of Social Security," *Scandinavian Journal of Economics*, 102, 523-545.

[31] Wolfinger, Raymond E. and Steven J. Rosenstone (1980), *Who Votes?*, New Haven, CT: Yale University Press.

6 Appendix

6.1 The Definition of Parameters in Proposition ??

Let $q \equiv \left(2 - 27n^2 R (w^u)^2 / 2n\right) / 27n^3$. b_0, b_1, π_0, π_1 are defined as follows

$$\begin{aligned}
b_1 &\equiv \sqrt[3]{-q + \sqrt{q^2 - \frac{1}{729n^6}}} + \sqrt[3]{-q - \sqrt{q^2 - \frac{1}{729n^6}}} - \frac{1}{3n}, && (34)\\
\pi_1 &\equiv -n\left(1 + \left(n + \frac{1}{R}\right) b_1 + \frac{n}{R} b_1^2\right), && (35)\\
b_0 &\equiv \frac{\Pi b_1 - (w^u)^2 (1+R)}{n b_1 (n + 1/R + n/R (1 + b_1)) - \pi_1 (1 + 2b_1) - (w^u)^2}, && (36)\\
\pi_0 &\equiv \Pi - n\left(\left(n + \frac{1}{R}\right) b_0 + \frac{n}{R} b_0 (1 + b_1)\right). && (37)
\end{aligned}$$

6.2 Proof of Proposition 1

To prove Proposition 1, we first suppose that the median voter is the middle poor and solve for the equilibrium policy rules. We then derive the sufficient conditions for the middle poor to be the median voter at each period.

Due to the linear-quadratic preference, it would be natural to guess that the policy rule T is linear

$$T\left(h_{-1}^u\right) = \phi_0 + \phi_1 h_{-1}^u, \tag{38}$$

where ϕ_0 and ϕ_1 are two undetermined coefficients. Substituting (38) into (12), we get

$$H(\tau) = \left(1 + \frac{1 - \phi_0}{R} - \tau\right) \frac{w^u}{1 + w^u \phi_1 / R}. \tag{39}$$

Combining (38) and (39), we obtain a linear social contract B

$$B(\tau) = b_0 + b_1 \tau, \tag{40}$$

where

$$\begin{aligned}
b_1 &\equiv -R w^u \phi_1 / (R + w^u \phi_1) && (41)\\
b_0 &\equiv \phi_0 + w^u \phi_1 (1 + R - \phi_0) / (R + w^u \phi_1) \\
&= \phi_0 (1 + b_1/R) - b_1 (1 + 1/R). && (42)
\end{aligned}$$

Note that equation (41) implies

$$b_1 = -w^u \phi_1 (1 + b_1/R). \tag{43}$$

To obtain b_0 as the product of baseline tax rate and a multiplier, we plug (43) into (42) and get

$$\begin{aligned} b_0 &= \left(\phi_0 + \phi_1 w^u \left(1 + 1/R\right)\right)\left(1 + b_1/R\right) \\ &= \left(\phi_0 + \phi_1 \hat{h}^u\right)\left(1 + b_1/R\right) \\ &= \hat{\tau}\left(1 + \frac{b_1}{R}\right), \end{aligned} \tag{44}$$

where $\hat{h}^u \equiv w^u \left(1 + 1/R\right)$ is the first-best human-capital investment, and $\hat{\tau} \equiv \left(\phi_0 + \phi_1 \hat{h}^u\right)$ is the baseline tax rate.

Plugging (40) into (17), we can express the future tax base as

$$Y\left(\tau\right) = \pi_0 + \pi_1 \tau, \tag{45}$$

where π_0 and π_1 are defined by (37) and (35), respectively. Then, the first-order condition (19) yields

$$\begin{aligned} \tau &= -\frac{b_1 \pi_0 + b_0 \pi_1}{2 b_1 \pi_1} + \frac{R w^u}{2 b_1 \pi_1} h^u_{-1}. \\ &= \widehat{\phi}_0 + \widehat{\phi}_1 h^u_{-1} \end{aligned} \tag{46}$$

Our definition of Markov Perfect Equilibrium pins down ϕ_0 and ϕ_1 in (38) as a fixed point

$$\phi_0 = \widehat{\phi}_0 = -\frac{b_1 \pi_0 + b_0 \pi_1}{2 b_1 \pi_1}, \tag{47}$$

$$\phi_1 = \widehat{\phi}_1 = \frac{R w^u}{2 b_1 \pi_1}. \tag{48}$$

To solve for the above fixed point, we reorder (43) and (42) as

$$\phi_1 = -\frac{b_1}{w^u \left(1 + b_1/R\right)} \tag{49}$$

$$\phi_0 = \frac{b_0 + b_1 \left(1 + 1/R\right)}{1 + b_1/R}. \tag{50}$$

Plugging (49) and (35) into (48), we obtain a four-order polynomial of b_1:

$$2n \frac{n}{R} b_1^4 + 2n\left(n + \frac{1}{R}\right) b_1^3 + 2n b_1^2 - \left(w^u\right)^2 b_1 - R\left(w^u\right)^2 = 0. \tag{51}$$

Factorizing (51), one root of b_1 equals $-R$, which should be omitted by the necessary condition $b_1 > 0$ for the sustainability of social security. The other three roots solve

$$n b_1^3 + b_1^2 - \Psi = 0, \tag{52}$$

where $\Psi \equiv R(w^u)^2/2n$. Rearrange (52):

$$b_1^2 = \frac{\Psi}{1+nb_1}. \tag{53}$$

It is straightforward to see that the left-hand side and the right-hand side of (53) have a unique cross with $b_1 > 0$, which gives the only real root of b_1, i.e., (34). The other undetermined coefficient, b_0, can then be solved by plugging (50) and (37) into (47).

Then we need to check whether $T\left(h_{-1}^u\right) \in (0, \overline{\tau})$ for all $h_{-1}^u \in \left[\underline{h}, \overline{h}\right]$. This gives the existence conditions (20) and (21).

Having solved the equilibrium policy rules, we now derive the sufficient conditions for the middle poor to be the median voter or, equivalently, $\tau^{m,u} \leq \tau^o$ and $\tau^{y,u} \leq \tau^{m,s}$ in (16). We first establish the sufficient condition for $\tau^{m,u} \leq \tau^o$. Note that the left-hand side of (22) is the highest tax rate for which the middle poor would vote. Given the equilibrium social contract B, the current tax base y_t can be written as

$$y = Y_c\left(h_{-1}^u, \tau\right) = n\left(\frac{h_{-1}^u}{w^u} + n\left(1 - \tau + \frac{1 - B(\tau)}{R}\right)\right). \tag{54}$$

Maximizing the indirect utility of the old, (15) solves $\tau^o = \min\left\{\bar{\tau}, -Y_c\left(h_{-1}^u, \tau\right) / \left(\partial Y_c\left(h_{-1}^u, \tau\right)/\partial\tau\right)\right\}$. In words, the old choose τ^o to attain the top of the Laffer curve. Substituting the linear tax policy rule B into (54), we have

$$\tau^o = \min\left\{\bar{\tau}, \frac{1 + 1/R - b_0/R + h_{-1}^u/nw^u}{2\left(1 + b_1/R\right)}\right\}, \tag{55}$$

where b_0 and b_1 are defined in (36) and (34), respectively. Since τ^o increases in h_{-1}^u, the minimum τ^o locates at $h_{-1}^u = \underline{h}$, which gives the right-hand side of (22). Therefore, (22) says that the lowest tax rate for which the old would vote is larger than the highest tax for which the middle poor would vote.

To make sure $\tau^{y,u} \leq \tau^{m,s}$, we derive the sufficient condition for $\tau^{y,u} = \underline{\tau} = 0$. Differentiating the first component on the right-hand size of (8) yields the marginal cost of increasing τ_t:

$$\left(1 - \tau_t + \frac{1 - b_0 - b_1\tau_t}{R}\right)\left(w^j\right)^2\left(1 + \frac{b_1}{R}\right).$$

Here we use the fact that $\tau_{t+1} = b_0 + b_1\tau_t$. Differentiating the second component on the right-hand size of (8) yields the marginal benefits of increasing τ_t:

$$\frac{\partial p_{t+2}}{\partial \tau_t} = \frac{1}{R^2}\left[\left(\pi_0 + \pi_1 b_0\right)b_1^2 + \left(b_0 + b_1 b_0\right)\pi_1 b_1 + 2\pi_1 b_1^3\tau_t\right].$$

where $\tau_{t+2} = b_0\left(1 + b_1\right) + b_1^2\tau_t$ is used to substitute out τ_{t+2}. Note that the marginal cost of taxation is decreasing in τ_t. Intuitively, an increase in τ_t lead to a lower human capital investment, both directly and indirectly via its impact on τ_{t+1}. As a result, the before-tax income is smaller. Moreover, the marginal benefit is decreasing in τ_t, since an increase in τ_t reduces the future tax base $Y\left(B\left(\tau_t\right)\right)$ by increasing τ_{t+1}. Hence, the marginal cost (benefit) reaches its lower- (upper-) bound at $\tau_t = \overline{\tau}$ ($\tau_t = \underline{\tau} = 0$) and $w^j = w^u$. The young would always vote for zero tax rate if the lower-bound of marginal cost is above the upper-bound of marginal benefit. This gives (23) in Proposition 1.

6.3 Proof of Corollary 1

Note that $T\left(\overline{h}\right) > \underline{\tau} = 0$ requires $\phi_0 + \phi_1 \overline{h} > 0$, which implies $\phi_0 > -\phi_1 w^u (1+R)/R$. By the definition of b_0 in (44), it is easy to show that $b_0 > 0$. On the other hand, $T\left(\underline{h}\right) < \overline{\tau}$ implies that $b_0 + b_1\overline{\tau} = T\left(H\left(\overline{\tau}\right)\right) \leq T\left(\underline{h}\right) < \overline{\tau}$. Together with $b_0 > 0$, we have $b_1 < 1$ and $b_0/\left(1-b_1\right) \in \left(0, \overline{\tau}\right)$. $\square$

6.4 Proof of Lemma 1

Differentiating (52) with respect to w^u, we have

$$\frac{\partial b_1}{\partial w^u} = \frac{\partial \Psi / \partial w^u}{3nb_1^2 + 2b_1}.$$

Since $b_1 > 0$, sgn$(\partial b_1/\partial w^u)$ =sgn$(\partial \Psi/\partial w^u)$. It immediately follows that $\partial \Psi/\partial w^u > 0$. $\square$

6.5 Proof of Proposition 2

We first prove that $\frac{\partial \phi_0}{\partial w^u} > 0$. From (47), we can write

$$\phi_0 = \frac{b_1 n(1+n)(1+1/R) - \begin{pmatrix} b_1 n\left(n + \frac{1}{R} + \frac{n}{R}\left(1+b_1\right)\right) \\ +n\left(1+nb_1\right)\left(1+b_1/R\right) \end{pmatrix} b_0}{2b_1 n\left(1+nb_1\right)\left(1+b_1/R\right)}. \tag{56}$$

Using the fact that $b_0 = \phi_0\left(1+b_1/R\right) - b_1\left(1+1/R\right)$, (56) becomes

$$\phi_0 = \frac{\begin{array}{c} b_1 n(1+n)(1+1/R) \\ -\begin{pmatrix} b_1 n\left(n + \frac{1}{R} + \frac{n}{R}\left(1+b_1\right)\right) \\ +n\left(1+nb_1\right)\left(1+b_1/R\right) \end{pmatrix}\left(\phi_0\left(1+b_1/R\right) - b_1\left(1+1/R\right)\right) \end{array}}{2b_1 n\left(1+nb_1\right)\left(1+b_1/R\right)}. \tag{57}$$

Equation (57) delivers an implicit function for ϕ_0, with

$$\phi_0 = f\left(\phi_0, b_1\right), \tag{58}$$

where f is defined as the right-hand side of (57). Therefore, to establish $\frac{\partial \phi_0}{\partial w^u} > 0$, by Lemma 1, we only need to prove

$$\frac{d\phi_0}{db_1} = \frac{f_{b_1}}{1 - f_{\phi_0}} > 0,$$

where f_x denotes the partial derivative of f to x. It is easy to show that

$$f_{\phi_0} = -\frac{1}{2}\left[\frac{n+1/R+\frac{n}{R}(1+b_1)}{1+nb_1}+\frac{1+b_1/R}{b_1}\right] < 0.$$

Therefore, $\frac{d\phi_0}{db_1} > 0$ iff $f_{b_1} > 0$. Taking derivatives, we find that

$$f_{b_1} = \frac{\left(n^2(2+1/R)\right)b_1^2+2nb_1+1}{(1+nb_1)^2 b_1^2}\frac{\phi_0}{2} - \frac{1+1/R}{2}\frac{n^2}{(1+nb_1)^2}. \tag{59}$$

So it is equivalent to prove

$$\phi_0 > \frac{(1+1/R)\,n^2 b_1^2}{n^2(2+1/R)\,b_1^2+2nb_1+1}. \tag{60}$$

Using (56), (60) can be rewritten as

$$\frac{1+n+Q}{(1+b_1/R)(2b_1(1+nb_1)+Q)} > \frac{n^2 b_1}{n^2(2+1/R)\,b_1^2+2nb_1+1}. \tag{61}$$

where

$$Q \equiv 2\left(1+(n+1/R)\,b_1+nb_1^2/R\right)+nb_1/R-1.$$

Rearranging (61), we get

$$(1+n)\left(n^2(2+1/R)\,b_1^2+2nb_1+1\right)+\left(2n^2b_1^2+2nb_1+1\right)Q > n^2b_1(1+b_1/R)\,2b_1(1+nb_1)+n^2b_1Q. \tag{62}$$

Note that the right-hand side of the inequality (62) is equal to

$$\begin{aligned}
& \left(2n^2b_1^2+2n^2b_1^3/R\right)(1+nb_1)+n^2b_1\left(1+(2n+2/R+n/R)\,b_1+2nb_1^2/R\right) \\
= \ & 2n^2b_1^2+2n^2b_1^3/R+2n^3b_1^3+2n^3b_1^4/R+n^2b_1+n^2(2n+2/R+n/R)\,b_1^2+2n^3b_1^4/R \\
= \ & 4n^3b_1^4/R+\left(2n^2/R+2n^3\right)b_1^3+n^2(2+2n+2/R+n/R)\,b_1^2+n^2b_1.
\end{aligned}$$

The left-hand side of the inequality (62) is equal to

$$\begin{aligned}
& 4n^3b_1^4/R+2n^2(2n+2/R+n/R)\,b_1^3 \\
& +\left(\begin{array}{c} 2n^2+2n^3+n^2/R+n^3/R+2n^2 \\ +2n(2n+2/R+n/R)+2n/R+4n^2/R+2n/R \end{array}\right)b_1^2 \\
& +\left(2n+2n^2+2n+2n+2/R+n/R\right)b_1+2+n.
\end{aligned}$$

It is then immediate that the left-hand side of (62) is greater than the right-hand side of (62). Therefore, $f_{b_1} > 0$ and we have $\frac{\partial\phi_0}{\partial w^u} > 0$.

We next prove that $\frac{\partial\phi_0}{\partial w^u} > 0$. (48) gives

$$\phi_1 = \frac{1}{2n} \frac{R/\phi_1 + w^u}{1 + \left(n + \frac{1}{R}\right) b_1 + \frac{n}{R} b_1^2}. \tag{63}$$

Equation (63) delivers an implicit function for ϕ_1

$$\phi_1 = g\left(\phi_1, w^u\right), \tag{64}$$

where $g \equiv \frac{1}{2n} \frac{R/\phi_1 + w^u}{1+\left(n+\frac{1}{R}\right)b_1 + \frac{n}{R}b_1^2}$. Equation (64) implies

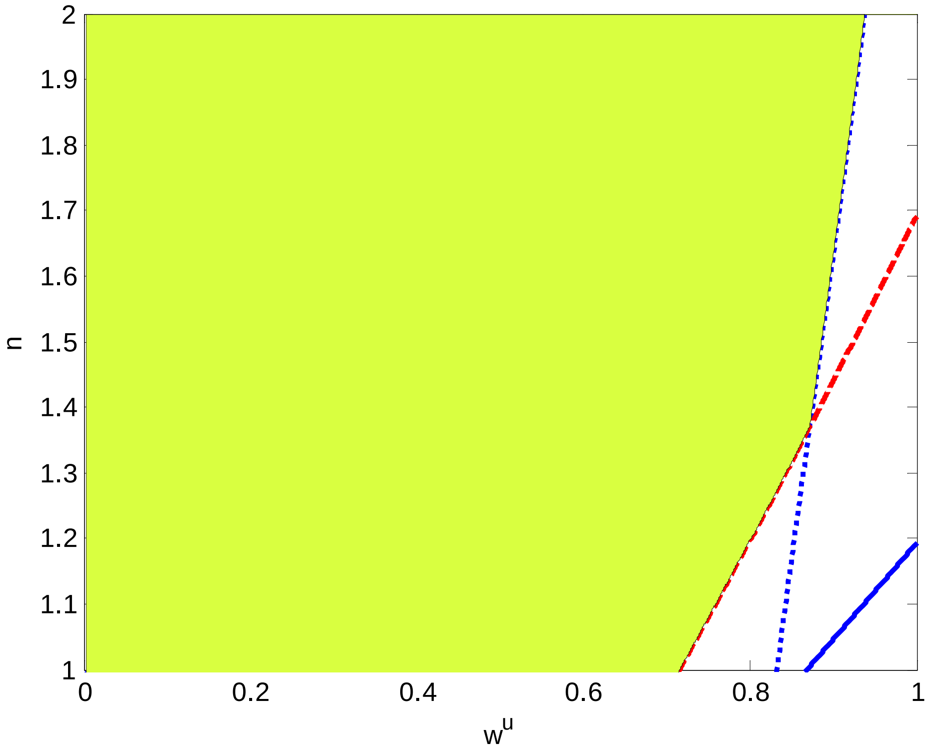
$$\frac{\partial \phi_1}{\partial w^u} = \frac{g_{w^u}}{1 - g_{\phi_1}}$$

where g_x denotes the partial derivative of g with respect to x. Since $g_{\phi_1} < 0$, it is straightforward that $\frac{\partial \phi_1}{\partial w^u} > 0$ iff $g_{w^u} > 0$. Note that

$$g_{w^u} = \frac{1}{2n} \left[\frac{1}{1 + \left(n + \frac{1}{R}\right) b_1 + \frac{n}{R} b_1^2} + \left(R/\phi_1 + w^u\right) \frac{\partial \frac{1}{1+\left(n+\frac{1}{R}\right)b_1+\frac{n}{R}b_1^2}}{\partial w^u} \right].$$

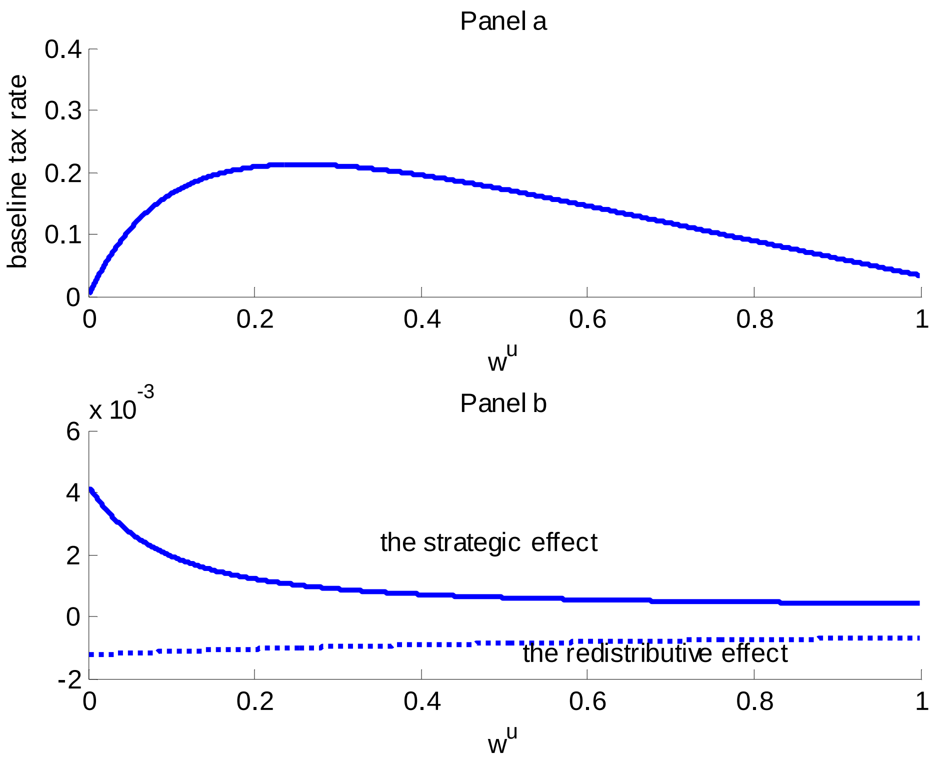
It is easy to see that $\partial \left(\frac{1}{1+\left(n+\frac{1}{R}\right)b_1+\frac{n}{R}b_1^2} \right) / \partial w^u < 0$, as $\partial b_1 / \partial w^u > 0$. Also, notice that (63) implies that $R/\phi_1 + w^u < 0$ as $\phi_1 < 0$. Therefore, the second argument in the bracket is positive. This, together with the positive sign of the first argument, establishes that $g_{w^u} > 0$. Therefore, we have $\frac{\partial \phi_1}{\partial w^u} > 0$.

Figure 1. Range of Population Growth Rate and Wage Inequality for Political Sustainability of Social Security



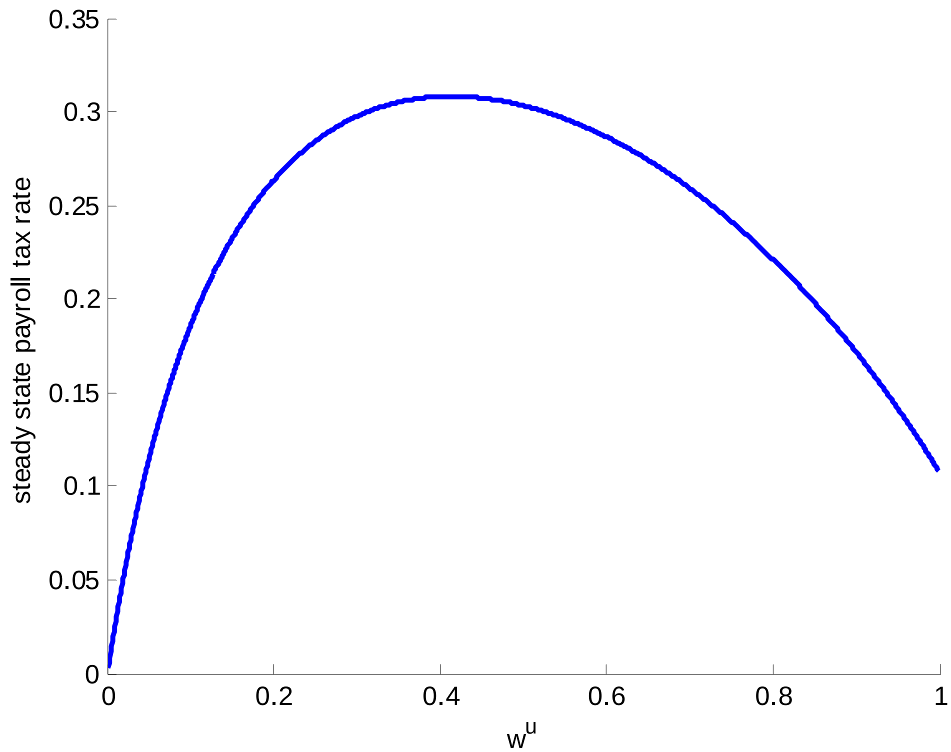
Note: This figure shows the combination of population growth rate, n and wage rate of the poor, w^u, that allows the middle-poor to be the median voter in a Markov perfect equilibrium. The dash line plots the threshold condition of n implied by (23). The dotted and the solid lines, represent the threshold conditions for (19) and (20) respectively. The set of n and w^u allowing the middle poor to be the median voter is captured by shaded area. $R = 1.025^{30}, \bar{\tau} = 0, \underline{\tau} = 0.9$.

Figure 2. The Strategic and the Redistributive Effects on the Baseline Tax Rate



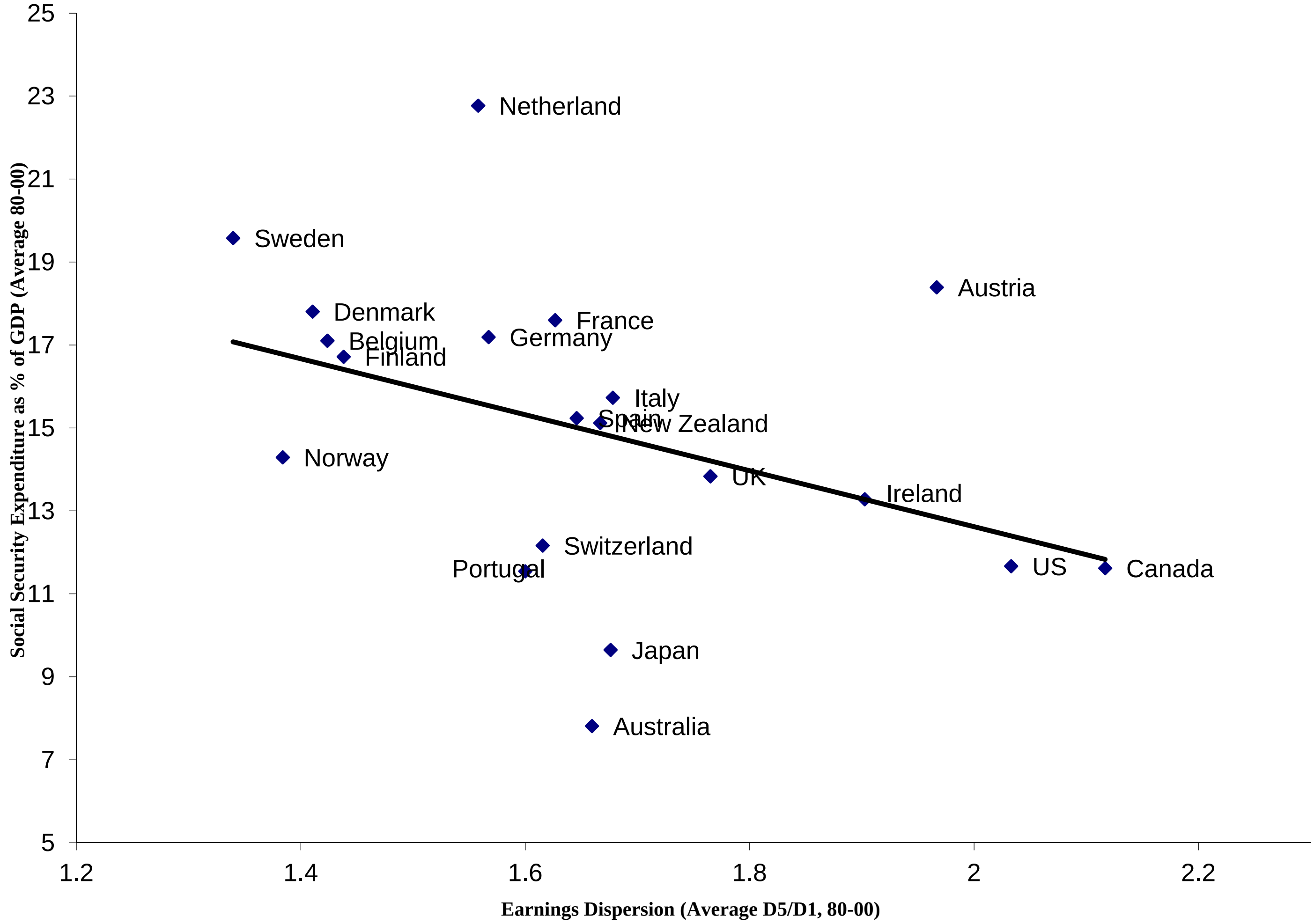
Note: Panel (a) shows the relationship between wage inequality, as captured by the wage rate of the poor, w^u, and the baseline tax rate, $\hat{\tau}$. Panel (b) shows the relationship between wage inequality and its two effects on the size of social security. $\mathrm{R} = 1.025^{30}, \mathrm{n} = 1.38, \bar{\tau} = 0, \underline{\tau} = 0.9$.

Figure 3. The Relationship between Wage Inequality and Steady-State Payroll Tax Rate



Note: This figure shows the relationship between wage inequality, as captured by the wage rate of the poor, w^u, and the steady-state tax rate on social security. $\mathrm{R} = 1.025^{30}$, $\mathrm{n} = 1.38$, $\bar{\tau} = 0$, $\underline{\tau} = 0.9$.

Figure 4. The Correlation between Earnings Inequality and The Size of Social Security Across OECD Countries



Note: This figure plots the relationship between earnings dispersion and the social security expenditure as percentage of GDP across OECD countries. Data from earnings inequality, as measured by the ratio of the median earnings to the upper limit of earnings of employees in the bottom decile of the earnings distribution, are from OECD national account statistics. Data for social security expenditure as a percentage share of GDP (SST), are from Economic Outlook of OECD.