On the Potential for Pareto Improving Social Security Reform with Second-Best Taxes

Kent Smetters
The Wharton School and NBER

Prepared for the Sixth Annual Conference of Retirement Research Consortium
"The Future of Social Security"
August 12-13, 2004
Washington, DC

The research reported herein was performed pursuant to a grant from the U.S. Social Security Administration. The opinions and conclusions are solely those of the author and should not be construed as representing the opinions or policy of the Social Security Administration or any agency of the Federal Government or the Retirement Research Consortium.

On the Potential for Pareto Improving Social Security Reform with Second-Best Taxes

Kent Smetters
The Wharton School and NBER

PRELIMINARY AND INCOMPLETE: PLEASE DO NOT CITE

1. INTRODUCTION

Reforming social security to include personal accounts is generally viewed as a zero-sum game between generations. This paper demonstrates that adopting personal accounts can produce efficiency gains in the presence of elastic labor supply even when only *second-best* labor income taxes are available. While the traditional analysis starting with Hicks (1943, 1944-5) and Harberger (1974) tends to focus on report welfare gains from policy changes in which agents are compensated using first-best taxes,[26] the focus on second-best taxes is also relevant since those are the instruments that are actually available to the government.

However, care must be taken in the design of the introduction of personal accounts: Some standard methods of introducing personal accounts are either neutral or even efficiency reducing. For example, an actuarially-fair "carve out" is neutral whereas shutting down Social Security by distributing "recognition bonds" to current participants could actually decrease welfare by increasing the effective tax rate on labor income. Instead, the shutdown must be designed correctly in order to achieve the gains.

[26] Feldstein and Liebman (*), for example, discuss the deadweight costs of Social Security relative to first-best taxes.

1.1. Importance of Dimensionality of the Model

Obtaining a Pareto improvement with labor income taxes is not possible inside a standard two-period framework in which agents work the first period and retire the second period. In that setting, shutting down Social Security would require compensating second-period agents for their loss of benefits by taxing first-period agents using a labor income tax, thereby undoing the privatization. In a model with three or more periods, however, a Pareto improvement can be obtained through an implicit lump-sum tax (wealth levy) on benefits already accrued by middle-age workers alive at the time of the reform. This wealth levy is then implicitly used to afford them a higher return on their future contributions, thereby reducing the effective tax rate on their labor supply.

1.2. Intuition

To understand this result intuitively, consider the following thought experiment. Suppose that households work for two periods and retire the third period. The government suddenly decides that it would no longer recognize the Social Security benefits that the second-period period workers accrued during their first period of life. In exchange for this implicit wealth levy, however, the government would give these workers the capital market rate of return on a portion of their second-period contributions (i.e., would privatize a portion of their second-period contributions). The portion of second-period contributions invested in the capital market would be set to produce a revenue loss equal in present value to this implicit wealth levy at the pre-reform level of labor supply. Workers are clearly better off under this experiment because they are fully compensated for their losses in present value but they now face a lower effective tax rate

on new contributions since some of the contributions now earn the market rate of return. Moreover, the government's inter-temporal budget constraint is improved at post-reform levels of labor supply assuming standard preferences for leisure. To capture these gains, however, the privatization experiment must be designed correctly.

1.3. Outline

Section 2 outlines the pre-reform economy with a social security system. Section 3 discusses the meaning of an "actuarially-fair reform." Sections 4 and 5 demonstrates that two traditional methods of privatization – a standard "carve out" and a "shutdown" with recognition bonds – would not produce efficiency gains and could even lead to efficiency losses. Section 6 shows how to modify the shutdown to produce efficiency gains.

NOTE: In this draft, the remainder of these *NOTES* is terser and the notation is potentially confusing since I economized on some of the notation (I should have included more subscripts to more clearly distinguish between *time* and *ages*). A future draft will provide more supporting text as well as more complete notation.

2. PRE-REFORM ECONOMY

This section describes the economy before social security is reformed.

2.1. Demographics and Factor Prices

There are N_t first-period agents alive at time *t*. Population grows at time *t* grows at rate $n_t \equiv N_t / N_{t-1}$. The total wage base time *t* grows at a gross rate $G_t \equiv (1+g_t) \equiv$

$(1+n_t)(1+x)$, where x is the exogenous and constant rate of technological change between time periods. We assume that the population is stationary and so we drop the time subscripts for *N* and *G* in the subsequent discussion. Factor prices are also taken as fixed, as would be the case in a small open economy. The gross rate of return to risk-less capital is denoted as $R=(1+r)$ while wage rate each period is w.

2.2. Households

In these *NOTES*, it is sufficient to focus on budget constraints in order to make our basic point. Agent *j* is born at time *t* and lives for three periods. She works in the first two periods and retires at the beginning of the third period. Her exogenous levels of productivity (efficiency units) in the first and second periods are α_1^j and α_2^j, respectively, and she supplies l_1^j and l_2^j units of labor in each period, respectively.[27] The budget constraints in a stationary economy are as follows:

(1) $c_1^j + a_1^j = \alpha_1^j w l_1^j (1-\tau)$

(2) $c_2^j + a_2^j = \alpha_2^j (1+x) w l_2^j (1-\tau) + R a_1^j$

(3) $c_3^j = R a_2^j + b^j$

where *b* is the pay-as-you-go social security benefit received in the third period.

Over 150 countries have defined-benefit social security systems that are mostly pay-as-you-go financed. In most cases, benefits are proportional to the average of past earnings; in a few countries like the United States, benefits paid *relative* to previous earnings are decreasing in the value of average past earnings (although some of this

[27] Labor income varies by agent type but is deterministic and observable by the government. Components of labor income (i.e., productivity), however, need not be observable by the government.

redistribution might be offset by differences in mortality and spousal qualifications). So, in this section, we focus on proportional benefits where every agent gets the same gross "internal" rate of return, *G*, to their social security contributions:

(4) $$b^j = G^2 \alpha_1^j w l_1^j \tau + G \alpha_2^j (1+x) w l_2^j \tau$$

2.3. Age-Specific Marginal Net Tax Rates

The *net* social security tax paid by agent *j* varies by age and equals the value of taxes paid at a given age less the present value of future benefits received from that tax:

(5) $$\tilde{T}_1^j \equiv \tau \alpha_1^j w l_1^j - \left(G/R\right)^2 \tau \alpha_1^j w l_1^j = \left[1 - \left(G/R\right)^2\right] \tau \alpha_1^j w l_1^j$$

(6) $$\tilde{T}_2^j \equiv \tau \alpha_2^j (1+x) w l_2^j - \left(G/R\right) \tau \alpha_2^j (1+x) w l_2^j = \left[1 - \left(G/R\right)\right] \tau \alpha_2^j (1+x) w l_2^j$$

(7) $$\tilde{T}_3^j = -b$$

We focus on a dynamically efficient economy, which implies $G > R$.[28] Hence, the net taxes paid by agent *j* at ages 1 and 2 are positive but below the statutory taxes.

The *marginal* net tax rates on social security contributions by age indicates how much the net taxes increase with wage income earned at that age:

(8) $$\tilde{\tau}_1^j \equiv \partial \tilde{T}_1^j / \partial \left(\alpha_1^j w l_1^j\right) = \left[1 - \left(G/R\right)^2\right] \tau$$

(9) $$\tilde{\tau}_2^j \equiv \partial \tilde{T}_2^j / \partial \left(\alpha_2^j (1+x) w l_2^j\right) = \left[1 - \left(G/R\right)\right] \tau \quad .$$

(10) $$\tilde{\tau}_3^j \equiv 0$$

[28] While an OLG economy, in theory, need not be dynamically efficient, the empirical evidence strongly suggests that this condition holds in many industrial countries. For the United States, Abel, Mankiw, Summers and Zeckhauser (*) found that the condition held in every year in their dataset, from * - *.

The marginal net tax rate at age 3 is zero since benefits received in that period are unaffected by any wage income received in the third period.[29] Notice that $\tilde{\tau}_2^j < \tilde{\tau}_1^j$, i.e., the net marginal tax rates decrease in the agent *j*'s age, *s*, as the "compound effect" of the $\left(R/G\right)$ "wedge" is reduced. Net marginal tax rates by age and different demographic status for the U.S. Social Security system are reported in Samwick (*).

3. Focusing on Actuarially-Fair Reforms

Since this paper focuses on efficiency gains, we consider "actuarially-fair" reforms in these *NOTES*, that is, reforms that would not redistribute resources across generations at *pre-reform* levels of labor income.[30] This restriction is, in part, made to simplify the analysis in order to focus on the mechanisms that produce efficiency gains: reforms that are not actuarially-fair would generally require an additional "redistribution authority" that compensates those agents that lose from a reform. A future draft will consider an even richer set of reform options.

The government loses some tax revenue from the personal account reforms that we consider herein, requiring the government to float some new debt in order to meet its short-run pay-as-you-go obligations. However, its future unfunded obligations are reduced by the same amount in present value. Hence, at pre-reform levels of labor supply, the government's inter-temporal budget constraint is unaffected by an actuarially-

[29] In reality, many social security systems throughout the world contain provisions that produce a positive marginal net tax rate during retirement, including actuarially-unfair early retirement deductions; unfair delayed retirement credits; earnings test; and, taxes on benefits when other income exceeds a threshold. The purpose of this paper is to examine the potential for efficiency gains without appealing to these specific provisions as well as to other distortions such a spouse benefit in the United States system.

[30] It is important to distinguish between fair *reforms* and a fair *initial* system. It is well known that when $R > G$, a pay-as-you-go social security system is actuarially unfair since it produced a windfall gain for some previous generations. However, the reforms we consider are fair when judge against the market rate, *R*.

fair reform; its budget constraint is only affected by policy-induced changes in labor supply. So if, for example, a reform reduces the marginal net tax rates *and* induces a greater amount of labor supply then distorting tax rates (including those used to finance the rest of government consumption that we ignore herein) can be reduced even more. {TO DO: formalize}

4. Reform Option 1: A Carve-Out

One way of introducing personal accounts is to "carve out" a portion of an agent's social security payroll tax and to deposit the money into a personal account. Since the agent pays less money into the traditional defined-benefit social security system, her future social security benefit is reduced by the same amount in present value.

Suppose that the portion π of the payroll tax, τ, is carved out and placed into a personal account while the portion $(1-\pi)$ of τ continues to be paid into the social security system. The new budget constraints become

(1') $$c_1^j + a_1^j = \alpha_1^j w l_1^j \left(1 - \tau\pi - \tau\left[1-\pi\right]\right) = \alpha_1^j w l_1^j \left(1-\tau\right)$$

(2') $$c_2^j + a_2^j = \alpha_2^j \left(1+x\right) w l_2^j \left(1 - \tau\pi - \tau\left[1-\pi\right]\right) + R a_1^j = \alpha_2^j \left(1+x\right) w l_2^j \left(1-\tau\right) + R a_1^j$$

(3') $$c_3^j = R a_2^j + b^j ,$$

where b^j is now the third-period assets that are in the new carve-out personal account plus the reduced benefit that is received from the social security system:

(4')

$$b^j = \overbrace{\left[G^2 \alpha_1^j w l_1^j \tau - R^2 \alpha_1^j w l_1^j \tau\pi\right]}^{\text{1st-Period Accrued Benefit less Reduction}} + \overbrace{\left[G\alpha_2^j \left(1+x\right) w l_2^j \tau - R\alpha_2^j \left(1+x\right) w l_2^j \tau\pi\right]}^{\text{2nd-Period Accrued Benefit less Reduction}} + \overbrace{\left[R^2 \alpha_1^j w l_1^j \tau\pi + R\alpha_2^j \left(1+x\right) w l_2^j \tau\pi\right]}^{\text{Assets in new Carve-Out Personal Account}}$$
$$= G^2 \alpha_1^j w l_1^j \tau + G\alpha_2^j \left(1+x\right) w l_2^j \tau$$

Budget constraints (1) – (4) are identical to constraints (1') – (4'). As a result, the marginal net tax rates are unaffected by the introduction of an actuarially-fair carve out. Intuitively, on one hand, lowering the tax rate paid into the traditional pay-as-you-go social security system reduces marginal tax rate across all age groups. However, the concomitant benefit reduction has the effect of increasing the marginal tax rate on contributions, raising the marginal tax rates back to their pre-reform levels. Hence, the creation of a fair carve out is completely neutral even with elastic labor supply.

5. Reform Option 2: A Simple Shutdown

A seeming more drastic reform option would "shut down" the social security system altogether and deposit the payroll taxes directly into personal accounts. Recall, though, from equations (5) and (6) that when the economy is dynamically efficient ($R > G$), working agents face a positive net tax rate; simply eliminating that tax, therefore, would redistribute resources to them at the cost to some other generation (either the current elderly or future workers). In order to make the shutdown actuarially-fair to each generation, the government deposits new "shutdown bonds" into the new personal account for each worker that are equal to the present value of the benefit that would have been accrued during that age under the social security system. Since $R > G$, this amount is less than the actual payroll taxes paid by workers at that point in time. This difference – a "haircut" -- is used by the government to pay the new debt service (into perpetuity) that is exactly required to finance the benefits of the elderly at the time of the reform.[31]

[31] {Give proof in this footnote}

5.1. Young and Future Generations at the Time of the Reform

For young (first-period) and all future workers at the time of the reform, the new budget constraints implied by the new policy, therefore, are as follows:

(1") $c_1^j + a_1^j = \alpha_1^j w l_1^j (1-\tau) + b_1^j$

(2") $c_2^j + a_2^j = \alpha_2^j (1+x) w l_2^j (1-\tau) + R a_1^j + b_2^j$

(3") $c_3^j = R a_2^j$

where the new government "shutdown bonds" at ages 1 and 2 are equal to

(11) $b_1^j = \tau \alpha_1^j w l_1^j \left(G/R\right)^2$

(12) $b_2^j = \tau \alpha_2^j (1+x) w l_2^j \left(G/R\right)$

These new bonds earn the market interest rate, *R*. The value of these bonds in period 3, therefore, equals $R^2 b_1^j + R b_2^j = G^2 \tau \alpha_1^j w l_1^j + G \tau \alpha_2^j (1+x) w l_2^j = b^j$, which is the same value as the benefits received under social security. Since there are no borrowing constraints, agent *j*'s inter-temporal lifetime budget constraint is, therefore, the same after privatization as before, producing no change in marginal net tax rates or efficiency gains.

Intuitively, two competing factors are again at work. On one hand, contributions to personal accounts are invested at the higher rate of return *R*, thereby eliminating the marginal tax rate associated with these contributions. On other hand, the "haircut," which is the difference between the taxes actually paid and the shutdown bonds received, is a *pure* tax which is fully distorting. In fact, the amount of this pure tax at age 1 is equal to $\alpha_1^j w l_1^j \tau - b_1^j = \alpha_1^j w l_1^j \tau \left[1 - \left(G/R\right)^2\right] = \tilde{T}_1^j$, which is the value of the net taxes paid under the former social security system at age 1. Similarly, the pure tax at age 2 is equal

to $\alpha_2^j(1+x)wl_2^j\tau - b_2^j = \alpha_2^j(1+x)wl_2^j\tau\left[1-\left(G/R\right)\right] = \tilde{T}_2^j$. Hence, shutting down the social security system does not reduce the marginal net tax rates facing any generation.

5.2. Middle-Age Workers at the Time of the Reform

Special consideration, however, must be given to middle-age (second-period) workers at the time of the reform who already accrued benefits in the former social security system. Their budget constraints are as follows:

(13) $c_2^j + a_2^j = \alpha_2^j(1+x)wl_2^j(1-\tau) + Ra_1^j + b_2^{j,M}$

(14) $c_3^j = Ra_2^j$,

where the value of their "shutdown bond" received at age 2 equals

(15) $b_2^{j,M} = Rb_1^j + b_2^j$

This shutdown bond is simply equal to the value of the shutdown bond that they will receive for their second-period contributions shown in equation (12) plus the present value of benefits that they accrued during the first period. It is easy to show that the creation of personal accounts does not change the budget constraints for middle-age workers; hence, no efficiency gains are produced for middle-age workers either.

5.3. Example of an Efficiency-Reducing Shutdown

{Give an example of how a shutdown with recognition bonds could reduce efficiency by disconnecting future benefits from taxes.}

6. Reform Option 3: A "Cleverer" Shutdown

Let's now modify the shutdown experiment somewhat. Young and all future workers at the time of the reform receive the same shutdown bonds shown earlier in equations (11) and (12). We now, however, demonstrate how to produce efficiency gains by altering the construction of the shutdown bonds for middle-age workers.

The budget constraints for middle-age reform at the time of reform are still given by equations (13) – (14) along with a modified construction of their shutdown bond:

(15') $b_2^{j,M} = \Gamma\left(\alpha_2^j\left[1+x\right]wl_2^j\tau\right)Rb_1^j + b_2^j$

where $\Gamma'(\square) \geq 0$. Suppose that

(16) $\Gamma\left(\alpha_2^j\left[1+x\right]wl_2^j\tau\right) = \alpha_2^j(1+x)wl_2^j\tau \Big/ \overline{\alpha_2^j(1+x)wl_2^j\tau}$,

where $\overline{\alpha_2^j(1+x)wl_2^j\tau}$ is the level of payroll taxes that would have been paid by agent *j* under the previous social security system. This shutdown is then actuarially-fair since lifetime resources are unchanged for all agents at pre-reform levels of labor supply.

The net taxes paid by middle-age worker *j* is $\tilde{T}_2^{j,M} \equiv \tau\alpha_2^j(1+x)wl_2^j - b_2^{j,M}$. Differentiating this equation with respect to second-period wage income produces the following marginal net tax rate:

(17) $\tilde{\tau}_2^{j,M} \equiv \partial\tilde{T}_2^{j,M} \Big/ \partial\left(\alpha_2^j(1+x)wl_2^j\right) = \left[1-\left(G/R\right)\right]\tau - \Gamma'(\square)Rb_1^f$

Notice that $\tilde{\tau}_2^{j,M} < \tilde{\tau}_2^j$ when $\Gamma'(\square) > 0$, as in the example shown in equation (16). Hence, the shutdown considered in this section has reduced the marginal net tax rate facing middle-age agents at the time of the reform. Since this reform (i) has not affected the

benefits of those retired at the time of the reform or (ii) altered the budget constraints of young or future generations, it is clearly pareto improving with fixed factor prices.

Intuitively, this “cleverer” shutdown produces efficiency gains by implicitly levying a lump-sum tax on benefits accrued under the social security at the time of the reform and then using those proceeds to reduce the value of the marginal net tax rates facing middle-aged workers.