

Welfare and Generational Equity in Sustainable Unfunded Pension Systems

Alan J. Auerbach
Economics and Law
University of California, Berkeley

Ronald Lee
Demography and Economics
University of California, Berkeley

August 2008

This research was supported by the U.S. Social Security Administration through grant #10-P-98363-1-02 to the National Bureau of Economic Research as part of the SSA Retirement Research Consortium and by NIA through the grant R37-AG025488. The findings and conclusions expressed are solely those of the authors and do not represent the views of SSA, any agency of the Federal Government, or the NBER. Ed Palmer and Ole Settergren gave us valuable comments and advice that helped us model the actual Swedish system. Alex Ludwig provided comments and very helpful information for modeling the German system. We also thank participants in the NBER Summer Institute, the NBER Public Economics Program meeting, including our discussant Peter Diamond, and a seminar at Stockholm University for comments on earlier drafts, and gratefully acknowledge the assistance of Anne Moore, Erin Metcalf, and Carl Boe.

Abstract

We evaluate several actual and hypothetical sustainable PAYGO pension structures, including: (1) versions of the US Social Security system with annual adjustments of taxes or benefits to maintain fiscal balance; (2) Sweden's Notional Defined Contribution system and several variants developed to improve fiscal stability; and (3) the German system, which also includes annual adjustments to maintain fiscal balance. For each system, we present descriptive measures of uncertainty in representative outcomes for a typical generation and across generations. We then estimate expected utility for generations based on simplifying assumptions and incorporate these expected utility calculations in an overall social welfare measure. Using a horizontal equity index, we also compare the different systems' performance in terms of how neighboring generations are treated.

While the actual Swedish system smoothes stochastic fluctuations more than any other and produces the highest degree of horizontal equity, it does so by accumulating a buffer stock of assets that alleviates the need for frequent adjustments. In terms of social welfare, this accumulation of assets leads to a lower average rate of return that more than offsets the benefits of risk reduction, leaving systems with more frequent adjustments that spread risks broadly among generations as those most preferred.

Alan J. Auerbach
Department of Economics
University of California
508-1 Evans Hall
Berkeley, CA 94720-3880
auerbach@econ.berkeley.edu

Ronald Lee
Department of Demography
University of California
2232 Piedmont Ave
Berkeley, CA 94720-2120
rlee@demog.berkeley.edu

Introduction

Actual and projected population aging threatens the financial viability of Pay-As-You-Go (PAYGO) pension programs in many countries. These programs promise a level of benefits that cannot be sustained given current tax rates, so deep structural reforms are expected and in some countries have already occurred. Beyond the problem of fiscal instability, most PAYGO programs are Defined Benefit and create incentives for early retirement (Gruber and Wise, 1999) and may distort labor supply decisions over the whole life cycle. Furthermore, because these programs create pension wealth that is not backed up by assets, they may weaken incentives to save and thereby lead to a lower capital stock.

Reforms that adjust the general level of taxes or benefits in PAYGO programs can address the problem of fiscal sustainability, but would have little effect on incentives for work or saving. A new kind of pension program, called Notional Defined Contribution or Non-financial Defined Contribution (NDC), is intended to address both fiscal stability and labor supply incentives. Sweden has developed and implemented an NDC system and some other countries have followed suit including Italy, Poland, Latvia, Mongolia and the Kyrgyz Republic. Germany has recently adopted pension reforms that reflect some of the NDC principles, and France is also considering doing so (Legros, 2003; Holtzmann and Palmer, 2005).

NDC plans are intended to mimic the structure and incentives of Defined Contribution plans in that individuals contribute to their own accounts which yield a specified rate of return and are converted into annuities yielding a specified rate of return at an age chosen by the individual but above some stated minimum age. The specified rate of return earned by the accounts and paid by the annuity is in principle linked to the growth rate of total wages, which is the sustainable rate of return in a steady state PAYGO system. The terms of the annuity reflect

current mortality conditions, so effectively benefits are indexed to life expectancy. These provisions should make NDC systems fiscally sustainable. Furthermore, the individual accounts based on individual contributions and explicit rates of return can reduce the distortion of work incentives if workers view their benefits and taxes as more closely linked than under traditional Defined Benefit systems. Because these NDC systems are PAYGO and do not hold significant assets, they are easier to implement than funded systems since the painful transitional costs, which can total one, two or three years worth of GDP, do not have to be borne. However, the other side of this coin is that NDC programs would still be expected to weaken the incentive to save and to reduce the aggregate capital stock as would other kinds of PAYGO programs.

Many actual NDC programs, including Sweden's, set the rate of return equal to the growth rate of the wage rate rather than the growth rate of total wages, so that it does not reflect the growth rate of the labor force which in many countries is expected to be negative. Demographic change and decline will lead such programs into fiscal problems, so they include additional mechanisms to achieve fiscal sustainability.

The costs of different reform plans will have different consequences for generations during the transition to the new system, resulting from the interaction of the reform plan with the particular initial demographic conditions of each country such as the baby boom and bust in the United States. In this paper, our goal is not to analyze generational risk sharing during this specific transitional period for any particular country. Rather we analyze the generational uncertainty and risk sharing in a more general context of economic and demographic uncertainty, with the goal of finding more general properties of the pension systems that do not result from some particular demographic circumstances.

Any pension plan must operate in an environment of demographic and economic uncertainty. Fiscally sustainable plans must make adjustments in light of this changing environment. Different plans will make different adjustments, leading to differences in the way that risks are shared among generations. These differences will influence the level of uncertainty faced by a typical generation, and also affect the extent to which welfare varies across generations. This paper modifies a stochastic forecasting model (Lee and Tuljapurkar, 1998, and Lee *et al.*, 2003) to investigate both these consequences of plan structure. An earlier paper (Auerbach and Lee, 2008) used the same setup to investigate the fiscal stability and sustainability of different PAYGO pension plans, and we will build on those results. Here we will restrict our attention to plans that are fiscally sustainable.

We will consider a number of actual and hypothetical PAYGO pension structures, including: (1) versions of the US Social Security system in which taxes, benefits, or taxes and benefits are adjusted annually to maintain fiscal balance with zero debt or assets in every period;¹ (2) the actual Swedish NDC system, together with several modifications of it developed in our earlier paper; and (3) the actual reformed German system, which also is adjusted to maintain fiscal balance in each period.² For each sustainable system, we will consider first some descriptive measures of uncertainty in outcomes for generations. We then estimate expected utility measures based on simplifying assumptions, and incorporate these expected utility calculations in an overall measure of social welfare. We also compare the performance of the

¹ We have also simulated the actual US Social Security system, but because it is not sustainable under current law there is little interest in comparisons to sustainable systems.

² Because it is convenient for technical reasons to carry out simulations of the Swedish systems beginning with a small initial asset balance, we start the US and German systems with the same positive asset balance in order to put the different systems on an equal footing. We discuss this point further below when reviewing the simulations.

different systems in terms of how neighboring generations are treated, using the measure of horizontal equity developed in Auerbach and Hassett (2002).

Varieties of PAYGO Pension Plans

Our simulation model is based on the average age profiles of tax payments and benefit receipts for surviving members of the population, which we might interpret as referring to the representative individual in each generation. Since we do not consider individual heterogeneity within a generation, many details of a pension plan are irrelevant. Nor do we model behavioral responses to differences in pension plans. Of course, these aspects of pension structure and behavioral response are very important in their own right, but in this paper we simplify drastically in order to focus on macro uncertainty and intergenerational differences. In our discussion of various pension plans below, we will therefore only discuss those aspects which are relevant for this undertaking.

One way in which we simplify is to impose certain common characteristics across the different pension systems. We scale all pension systems so that they have average contributions of the same scale as the current US system, and assume that all individuals work until age 67, the long-run normal retirement age under current US law, and that all individuals are retired thereafter. Modifications that these assumptions impose on the different schemes are discussed further below.

US Social Security

We will consider three plans derived from the current law US Social Security program. For the Old Age and Survivors (OASI) portion of this system, the current law tax rate is set at 10.6 percent of taxable payroll, and we set the tax rate at this level for all of the systems we

consider in this paper.³ The current law average benefit level for a generation is set as a certain percentage of the average real wage at age 60 with some upward adjustment that depends on the extent to which the generation continues to work past the early age at retirement at age 62. As the population ages and the ratio of retirees to workers rises, the balance of tax revenues and benefit costs will change so the system is not fiscally stable. We create fiscally sustainable systems by modifying this current law structure.

In the “Tax Adjust” program, the age schedule of taxes is adjusted by a multiplicative factor each period so as to produce the amount of revenue needed to cover that period’s benefits, so that the system is in perfect balance each period. In the “Benefit Adjust” program, the age schedule of benefits is similarly adjusted each period so that the costs of benefits exactly equal that period’s tax revenues. In the “50-50 Adjust” program, balance is achieved half by adjusting taxes and half by adjusting benefits. Further details will be provided when the simulations are discussed.

Swedish Notional Defined Contribution System

We describe the current Swedish system relatively briefly here. Our earlier paper provides a more detailed discussion and analysis of this system, and descriptions can also be found in Holtzmann and Palmer, eds. (2005).

The actual Swedish NDC program specifies that the rate of return earned on accounts in each year t equals the contemporaneous growth rate of the wage level, g_t . Because the sustainable rate of return in a steady state PAYGO system is $g+n$, where n is the growth rate of the labor force, this system may have problems with stability. Recognizing this, the program

³ For systems in which contributions fluctuate in order to maintain budget balance, as under the US “Tax Adjust” plan, we scale the system so that the average tax rate over all trajectories and cohorts is 10.6 percent.

also has a balancing mechanism, or “brake” (to be described shortly), that goes into effect when program assets reach a certain level, reducing the rate of return earned on accounts and annuities until asset levels are restored. In this way, demographic variations influence the system through the back door. Therefore the effects of economic and demographic variations are likely to be distributed across the generations quite differently than in the Social Security case.

At the time of retirement, which we assume to be at age 67, the individual’s account is converted into an annuity based on the account’s notional balance as of that date. The terms of the annuity reflect mortality conditions at the time of conversion and a rate of return which is set equal to a pre-specified expected rate of wage growth taken in the Swedish system to be .016, but in our simulations taken to be the actual growth rate at the time, which has a mean value of .011, the underlying average rate of labor productivity growth in our simulations. The benefit stream is then calculated to be constant over the individual’s (or in our case, the generation’s) remaining years, but it is adjusted for *ex post* deviations of wage growth from the assumed rate. In our simulations we also adjust the annuity level to reflect changes in mortality after retirement, although we found in our earlier paper that including this post-retirement updating had a minor impact on system stability.

Those who designed the Swedish system expected that with a rate of return equal to g and with the benefit actuarially reflecting mortality conditions at the time of retirement the system would be fiscally quite stable. Nonetheless they realized that it could still go into debt, for example if the population growth rate were negative for a sustained period, so they designed a brake, which would automatically reduce the rate of return under certain conditions, and thereby guarantee long term fiscal stability. To this end they defined a balance ratio b :

$$(1) \quad b = \frac{F + C}{NPW + P}$$

In this expression, F is the level of a financial asset, similar to the trust fund of US Social Security, but able to take negative values in case of debt. C is a new measure called the “Contribution Asset” which is defined to approximate the present value of future tax payments by participants. It equals the three-year average value of tax payments times the expected length of time between tax payments and receipt of benefits which is roughly thirty years. This measure would exactly equal this present value in steady state, with discounting at the rate $n+g$, and therefore would equal the system’s ability to meet future pension obligations through taxes (Settergren, 2005; Lee, 2005).

The denominator of the expression is the total pension liability of the system. The first term, NPW , is the present value of obligations to current workers. The second term, P , approximates the present value of benefits due to current retirees. This balance measure does not rely on explicit projections of demographic and economic variables and is therefore based on less than all the available information. On the other hand, the measure is a transparent one that is based on current cross sectional data and therefore less likely to be distorted by political pressures.

If the balance b falls below 1.0 then the brake is activated. When the brake is active, the gross rate of return $1+g_t$ that is used in computing both accruals of notional pension wealth and the growth of annuity payments is reduced by a factor of $b_t < 1.0$, that is the gross rate of return becomes $(1+g_t)b_t$. When the brake is not in place, the rate at which pension benefits grow in the actual Swedish system is $(1+g_t)/.016$, where .016 is the assumed central value of g . That is, if

wages grow more rapidly than .016 then the benefit is raised. Multiplication by b_t makes it more likely that the level of benefits will decline rather than rise.

The brake remains active until the product of the balance levels b_t for all the years since it is activated first exceeds 1.0, that is, it applies in year $s > t$ if $\prod_{v=t}^s b_v < 1.0$. The purpose of this condition for continuing activation of the brake is to ensure that the temporary application of the brake has no long term effect on the level of benefits. That is, there is a catch-up period while the brake is in effect, with $b > 1$ until the initial slowdown in the growth of notional pension wealth and annuity payments is reversed. The fiscal problem is thus resolved by reducing the level of benefits only temporarily while the brake is active, the saving coming from a temporary drop below the unrestrained benefits trajectory rather than a move to a lower benefits trajectory.

Note that the brake is asymmetric, and does not prevent the unlimited growth of the trust fund F . Note also that while it is mathematically possible for b to fall below zero, this could not meaningfully happen because it would entail more than complete confiscation of pension wealth and benefits.

We have simulated the Swedish system with the brake as just described, and with a rate of return of g . However, we have also defined alternative NDC systems that are closely modeled on the Swedish system but which have modified versions of the brake, or which have a rate of return of $n+g$, or both. We will now explain these modifications.

By setting the rate of return to $n+g$ rather than g , a system should track the varying demographic context and therefore have less need of the brake. We in fact found this to be so in our earlier paper; while even a system based on $n+g$ would on certain stochastic trajectories require some further intervention to preserve stability, the strength and frequency of these

interventions were reduced by incorporating labor force growth in the annual benefit adjustments.

We have noted that the Swedish brake is asymmetric. In considering alternatives to the existing Swedish system, we will focus on systems in which there is a symmetric brake, that is both a brake and an accelerator, such that the rate of return rises when $b > 1.0$ in addition to falling when $b < 1.0$. Because we impose the brake symmetrically, it is not necessary to incorporate the catch-up phase described above; we simply adjust benefits downward when $b < 1$ and upward when $b > 1$.

We have also developed systems with a somewhat different braking structure. Let r_t^a be the adjusted net rate of return. In the Swedish system this is $r_t^a = (1 + r_t)b_t - 1$, where $r = g$ or $n+g$; that is, it equals the adjusted gross rate of return minus 1. At low values of b , this takes a fierce bite out of the pension, and if $b=0$, which is possible, the net rate of return is -1. To soften the action of the brake we introduce a scaling factor A , between 0 and 1, as follows:

$$(2) \quad r_t^a = (1 + r_t)[1 + A(b_t - 1)] - 1$$

$A=1$ under the original Swedish brake. When A is less than 1, the adjusted rate of return varies less than in proportion to the b_t , and full confiscation does not occur until b falls to $1-1/A < 0$. If we set $A=0$ then there is no brake at all. In some of our simulations we use a brake with $A=.5$, which does not become fully confiscatory unless b falls to -1. In our simulations this almost never happens.

German System

A simplified description of the German pension system and our implementation of it follow. For more details, see Börsch-Supan and Wilke (2003), Börsch-Supan, Reil-Held and Wilke (2003), and Ludwig and Reiter (2008). Under this system, each pension beneficiary i receives a payment in year t equal to:

$$(3) \quad B_{t,i} = PV_t * EP_i * AA_i$$

where PV_t is the current pension value in year t , EP_i is the individual's "earning points" collected until retirement, and AA_i is an actuarial adjustment based on when the pensioner retired.

Earnings points are an increasing function of an individual's earnings, scaled to equal 1.0 for the average-wage individual for each year the individual worked. Given that we are ignoring intragenerational heterogeneity, we set EP_i constant at the assumed number of years of labor force participation. The actuarial adjustment AA_i scales benefits up or down according to the age at which the individual retires. We just assume retirement at 67 and set $AA_i = 1$. Thus, the benefits formula reduces to:

$$(3') \quad B_{t,i} = B_t = Y * PV_t$$

where Y is the number of years worked until retirement. (In the end, the value assumed for Y does not matter, because we scale the size of the system so that the average tax rate equals that of the US system.) Note that pensions are set up so that retirees of different ages get the same benefit, since earnings points are based on wages relative to the average wage when one worked, not on one's absolute wages.

The pension in year t evolves according to:

$$(4) \quad PV_t = PV_{t-1} * \frac{AGW_{t-1}(1-CR_{t-1})}{AGW_{t-2}(1-CR_{t-2})} * \left[1 - \alpha * \left(\frac{OA_{t-1} - OA_{t-2}}{OA_{t-2}} \right) \right]$$

where CR is the pension “contribution rate” (i.e., the social security payroll tax), AGW is the average gross earnings of employees, and OA_t is the old-age dependency ratio (OADR), defined as the ratio of population over 65 to population aged 15-64 in year t . The parameter α is set to 0.25, which we use in our analysis. In our terms, pension benefit growth is linked to g , but adjusted for fluctuations in CR and OA .

Substituting (3') into (4), we get:

$$(5) \quad B_t = B_{t-1} * \frac{AGW_{t-1}(1-CR_{t-1})}{AGW_{t-2}(1-CR_{t-2})} * \left[1 - \alpha * \left(\frac{OA_{t-1} - OA_{t-2}}{OA_{t-2}} \right) \right]$$

Expression (5) describes the calculation of benefits. Taxes are adjusted each year as a residual so that taxes and benefits are equal in the aggregate.

The German system differs from the others in that the pension benefit of all generations rises after retirement in proportion to preceding wage growth, g . However, benefits for a generation also vary over time in inverse proportion to the tax rate (“contribution rate”), so that beneficiaries share the pain or the gain of fiscal adjustment with taxpayers. Another factor reduces benefits if the old-age dependency ratio is rising.

Discussion of These Systems

We have simplified our representation of these systems to focus on their intergenerational risk sharing characteristics, but even so they remain complex, particularly the German and the NDC plans. Nonetheless, we can highlight one key difference among them. In the US Tax

Adjust system, the benefits are fixed and the burden of fiscal adjustment falls entirely on the taxpayers. In contrast, in the US Benefit Adjust and all the NDC systems, the tax rate is fixed and the burden of fiscal adjustment falls entirely on the beneficiaries. Finally, in the German system and in the US 50-50 Adjust system, the burden of adjustment is shared by taxpayers and beneficiaries.

The Stochastic Simulations

To evaluate and compare the risk-sharing characteristics of these public pension systems, we build upon a stochastic simulation model developed in earlier work by Lee and collaborators (see also the approaches in Alho *et al.*, 2005, and Alho *et al.*, in press).

In this earlier work, Lee and Tuljapurkar developed a model to generate stochastic long-term forecasts of the US Social Security system (Lee and Tuljapurkar, 1998; Lee *et al.*, 2003). In this model, the log of each age specific mortality rate is a linear function of a single mortality index that is in turn modeled as a random walk with drift, based on Lee and Carter (1992). Fertility is modeled as an ARIMA process with a pre-assigned long term mean of 1.95 births per woman (the period Total Fertility Rate) as assumed in the 2004 Social Security Trustees Report, henceforth TR04, as discussed in Lee (1993) and Lee and Tuljapurkar (1994). Immigration is taken as deterministically given by the intermediate assumption in the Social Security Trustees Report of 2004 (see Lee *et al.*, 2004 for treatment of immigration as stochastic in this framework). The fertility and mortality processes are fit to historical US data. Together these assumptions and models are the basis for stochastic simulations of the population of the United States (Lee and Tuljapurkar, 1994). Monte Carlo methods were used to generate a large number of stochastic sample paths, and from these the probability distribution of demographic outcomes

such as the OADR (population age 65 and over divided by population age 20 to 64)⁴ were calculated.

To give a sense of the demographic variations that result from this approach, Figure 1 plots the simulated OADR along 15 randomly chosen sample paths from among the 1000 we use for our analysis. Because the time series models for fertility and mortality were fit to US data, these simulated sample paths reflect randomly occurring low-frequency fluctuations something like the US baby boom and baby bust, with consequent effects on the OADR. Mortality variations have a less profound effect on the age distributions because the variance in mortality is far smaller than in fertility.

The Social Security model builds on the stochastic population using cross-sectionally estimated age profiles of labor earnings, tax payments and benefit receipts. The age profiles of labor earnings and tax payments are multiplicatively shifted from period to period by a time series of labor productivity growth. A time series model is fitted to the historical time series of productivity since 1950, purged of the influence of changes in the age composition of the labor force. A Monte Carlo simulation generates sample paths of productivity, and earnings and tax payments by age. The same simulation can also generate sample paths of age specific benefits levels, since these depend on average wages at age 60 for each generation, which are in turn determined by stochastic productivity.

The modified version of the US system we consider here has payroll taxes set equal to zero above age 67 and benefits set equal to zero below age 67, consistent with our modeling assumption that retirement under each system occurs at age 67. Benefits after age 67 are based

⁴ Note that this is a slightly different definition of the OADR than that used by the German system, discussed earlier, which starts from age 15 in computing the denominator.

on a simplified version of the actual US formula of providing a replacement rate of average indexed monthly earnings that is then indexed to the price level after retirement.

The trust fund for the US system is set at a small initial balance that is maintained constant as a share of payroll. For all systems, we accumulate balances using a time series model of the real interest rate earned on Social Security special issue bond holdings. The interest rate is modeled as a stationary series, as is the productivity growth rate (they are modeled jointly using VAR). The productivity growth rate has a positive mean of 1.1 percent (based on the Actuary's assumption in the 2004 Trustees' Report for the growth rate of covered real wage), so productivity, earnings, taxes and benefits are all trending upwards.⁵

For our purposes we wish to abstract from the particular demographic situation in the United States today so that we may derive results that are of more general applicability. We modify the model described above so it converges to an approximately stationary stochastic population distribution. This is accomplished by setting the mortality drift term to zero. The expected value of fertility is below replacement level, so in the long run the population converges to a level at which the deficit of births is just balanced by the deterministic inflow of migration. Because the level of mortality follows a random walk the population variance increases very slowly over time, and therefore this stochastic equilibrium is only approximately stationary, but in practice we have found that the nonstationarity of mortality is negligible. More important is the nonstationarity of productivity levels. Productivity growth is stationary, but the productivity levels to which it cumulates are not.

⁵ The Trustees Reports distinguish between the growth rate of productivity and of covered real wage, with the gap reflecting changes in hours worked and the proportion of compensation that is fringe benefits. We abstract from these issues and simply use the assumed growth rate of the covered real wage, .011, which we will henceforth refer to as the productivity growth rate or growth rate of the wage.

While this description of the model glossed over many details, it should convey a general idea of how we simulate the pension systems. For the actual simulations, we consider a sample of 1000 randomly drawn trajectories, using the same sample of 1000 for each of the pension systems considered. For each trajectory, we start with initial conditions based on long-run average values of different state variables and run the model for a “pre-sample” period of 100 years to generate histories needed for certain pension calculations. We then follow the paths for an additional 500 years, allowing us to examine the welfare of nearly 400 cohorts over their entire lives, which are assumed to extend for a maximum of 106 years.⁶

Previous Results on Fiscal Stability

Our earlier paper used the same stochastic simulations to consider the fiscal stability of the basic Swedish system and several variants, including some analyzed in this paper as well. We found that the current Swedish NDC system with its asymmetric brake effectively avoids accumulating large debt but sometimes does accumulate substantial assets which could otherwise have been returned as higher benefits for plan participants. We developed a symmetric version of the brake with a parameter regulating its severity of action, already described above. This symmetric brake insures two-sided fiscal stability for the NDC system, based like the existing Swedish system on the growth rate of the wage, g , even if it operates less severely than the current Swedish asymmetric brake. As discussed above, we also developed an NDC system in which the stipulated rate of return responds to demographic shocks as well as to variations in the productivity growth rate. This system is inherently more stable than the system based on the wage rate alone when the two systems are compared in operation without a brake. We also

⁶ Further details of this simulation methodology are provided in our earlier paper.

found that economic factors, as well as demographic ones, contribute significantly to the volatility of NDC pension systems.

Measures of Outcomes for Generations

Our earlier paper was largely limited to considering issues of fiscal stability in relation to system design. Here we will restrict our attention to systems that are fiscally stable and examine aspects of risk sharing across generations. This requires different kinds of measures, which we now discuss.

We will begin by considering two common measures of pension program outcomes. The first is the Implicit or Internal Rate of Return (*IRR*, or simply ρ in our equations) for each generation on each sample path as that constant value of the real discount rate $r_u = \rho$ for all u , for which the present value of taxes and benefits equals zero. We obtain a different $\rho_{t,i}$ for each generation and sample path, and can calculate the global average value. We know that in a deterministic steady state ρ equals $n+g$, but the IRR is a highly nonlinear measure so the average over the stochastic sample paths can be rather different than this.

In combining results for different generations it may be desirable to weight the internal rate of return by the size of the generation at birth. The structure of some pension plans may result in higher or lower benefits for larger generations, for example. A program that paid lower benefits to larger generations would have a larger global mean *IRR* without weighting by size, but the welfare implications would be unclear. For example, under the US Tax Adjust system it is advantageous to be in a large cohort because the tax burden is shared among more taxpayers and the tax rate will be lower, other things equal. Under the Benefit Adjust system it is better to be in a small cohort, other things equal, since for a given amount of tax revenues benefits will be shared among fewer and will be higher. For most systems it is not obvious whether large or

small generations will be favored, and some might be favored by population weighting for some measures and not for others.

A second summary measure of performance is the Net Present Value (*NPV*) of expected lifetime contributions or taxes minus benefits, expressed as a share of the expected present value of a cohort's lifetime earnings,

$$(6) \quad NPV_t = \frac{\sum_i \pi_{t,i} \sum_{s=t}^{t+T} \frac{1}{E_t \left[\prod_{u=t}^s (1+r_u) \right]} F_{s,i}}{\sum_i \pi_{t,i} \sum_{s=t}^{t+R} \frac{1}{E_t \left[\prod_{u=t}^s (1+r_u) \right]} Y_{s,i}},$$

where $F_{s,i}$ is the survival-weighted per capita benefit minus tax payment for this generation in year s and along trajectory i , $Y_{s,i}$ is survival-weighted income per capita for the generation in year s and along trajectory i , r_u is the interest rate in year u , T is the maximum lifespan (106 in our simulations) and R is the retirement age (67 in our simulations). When trajectories are weighted by population, the weights $\pi_{t,i}$ equal the cohort's initial population along trajectory i as a fraction of the sum of the cohort's initial populations along all trajectories. Otherwise, we set $\pi_{t,i} \equiv 1/N$, where N is the number of trajectories. As discussed in the Appendix below, the *NPV* as described in (6) measures the impact of the social security system on the cohort's expected utility (expressed as a share of expected lifetime earnings) for the case of risk-neutral preferences.

The dispersion of *IRR* and *NPV* about their means for any given generation describes the uncertainty faced by that generation, and these can be summarized by their variances which can be compared across pension systems, as we will do later. But it is also desirable to evaluate these different distributions of outcomes in terms of individual welfare. Our simulations cover

only the pension system and do not include saving, asset income, or non-pension taxes. A full expected utility calculation would need to take these other elements into account in a very complex dynamic programming problem. As an alternative, we derive a local approximation of the impact of different systems on expected utility using simplifying assumptions that relate the marginal utility of consumption along a trajectory to the level of risk aversion and the level of wages along that trajectory. The Appendix describes this methodology in detail. To take account of the fact that the welfare of transition generations may differ systematically according to the nature of a plan's method of adjustment to fiscal imbalances, we also compute a measure of social welfare based on our expected utility measures that incorporates the impact on transition generations.

To this point we have considered measures of the uncertainty of pension program outcomes and the tradeoff of this uncertainty against the mean return. The expected utility measure reflects this trade-off, and some might argue that the vector of expected utilities for different cohorts provides all the information needed to evaluate social welfare. But we may care in addition about how generations fare relative to other generations along a particular trajectory.⁷ One simple measure of performance in this regard is the variance of outcomes like the *IRR* and *NPV* among generations along a particular trajectory (as opposed to the variance across trajectories for a particular generation, already considered above), and we provide these below when comparing different systems. But we sense that such concerns about the relative well-being of different generations relate primarily to the treatment of individuals of similar ages, since these other generations form a likely reference group. For example, the “notch”

⁷ Consider, for example, the example of two generations and two outcomes, “good” and “bad,” which occur with equal probability. In case A, the two generations experience the same outcome (i.e., the combined outcomes for the two generations are good-good and bad-bad, each half the time); in case B, the generations experience opposite outcomes (good-bad and bad-good, each half the time). The expected utility for each generation is the same in case A and case B, but we might conceivably have a social preference for case A over case B.

generations in the United States experienced particularly sudden and large variations in their pension benefits in a way that struck many as unfair.

To reflect these concerns, we provide an additional set of performance measures for the various public pension schemes, based on the horizontal equity measure developed by Auerbach and Hassett (2002). This measure is derived from a social welfare function for which the degree of inequality aversion may differ according to whether individuals are “near” each other, by some measure, or not. One may decompose this social welfare function into different components, one of which reflects the social welfare cost of local tax burden disparities at each income level. We derive the scalar index of horizontal equity by asking what uniform fraction of existing income would deliver the same level of social welfare if all such local disparities were eliminated. This index has a maximum possible value of 1.0, and values closer to 1.0 indicate greater horizontal equity. For example, a value of 0.999 indicates that we would be willing to give up 0.1 percent of total income to eliminate horizontal inequality.

Measuring horizontal equity requires the specification of the degree of inequality aversion for comparisons among members of a particular reference group and a definition of the reference group itself, in this case with respect to generational proximity. As we claim no particular insight as to which parameters are best here, we simply adopt those used for the base case in Auerbach and Hassett (2002), a CES degree of inequality aversion equal to 2 and a neighborhood based on a normal distribution with standard deviation equal to .10 times income. We estimate horizontal equity based on our NPV calculation, measuring income as the present value of earnings along the trajectory and taxes as the net present value of taxes minus benefits along the trajectory.

Results

For reasons discussed earlier, we will present results both for generations weighted by their size at birth, in Table 1, and unweighted by size, in Table 2. We believe that the population-weighted measures are most relevant, since results are otherwise strongly influenced by whether the pension structure favors larger or smaller generations. Table 1 presents the population-weighted means and variances, while the unweighted versions are given in Table 2. The social welfare measures based on all generations, presented in Table 3, are limited to the population-weighted case. We will start by considering the results for the *IRR* in Table 1.

IRR

The first row of Table 1 presents the mean *IRRs* for each system, weighted by population. These mean values are in the range .010 to .013, with most of the variation due to the specific nonlinear interactions of the plans, the disturbances, and this particular measure. In a deterministic steady state all would yield approximately the same *IRR*, .011, i.e., the rate of wage and productivity growth, differing only because of the small initial asset position that would increase the *IRR* very slightly. However, the *IRR* for the Swedish system is penalized by the asymmetric brake and has the lowest mean. The next row of the table displays the corresponding median values for the *IRR*. The median values are slightly higher for all systems, suggesting left-skewed distributions, but the relative picture is similar, with the Swedish system still lower than the others.

The next two lines of Table 1 present two variance measures for the *IRR*, first across trajectories (averaged across cohorts), and then across cohorts (averaged across trajectories). As discussed above, the first variance measure is more relevant when considering the impact of social security on risk-averse households, while the second measure conveys some information

about how different generations fare along any particular trajectory, which might also be useful for social welfare evaluations where horizontal equity is a concern. The two measures of variance provide quite different rankings of the different systems. In particular, the Swedish system has, on average, an extremely low variance along a given trajectory, but a relatively high variance across trajectories. As will be seen shortly, this variation between results along and across trajectories is not repeated for net present value measures, which relate more closely to concepts of individual and social welfare.

NPV

The mean *NPVs* are in the range $-.057$ to $-.063$ as a fraction of lifetime earnings. These values are negative, as one would expect for a social security system in a dynamically efficient economy. As also expected, the actual Swedish system has a lower *NPV* than the three variants that include a symmetric brake, as these alternative systems do not build up assets, on average.

What is somewhat surprising is that the German system and the three variants of the US system, although performing much better than the actual Swedish system, perform worse than the three variants of the Swedish system that do not accumulate substantial assets on average. There is no mathematical reason why this can't happen for these complex calculations, although in weighting the results by initial cohort population and by keeping the average payroll tax rate the same under each system we have removed some of the most obvious factors that could contribute to systematic differences in expected present values.⁸

Unlike in the case of the *IRR*, the two variance measures for *NPV* tell similar stories in terms of system variance. The actual Swedish system has the lowest variance in *NPV* across

⁸ Without population-weighting, a system that favored small generations could make the "average" outcome look better. Since social security yields a below-market rate of return for each dollar of payroll taxes, a larger social security system delivers a more negative present value.

both trajectories and generations, achieving this stability at the cost of having the lowest mean *NPV* for the population-weighted case, because its asymmetric brake lets assets accumulate on some sample paths and thereby reduces the need for taxes or benefits to respond to shocks. The US Tax Adjust system has the highest variance in each case, followed by the German system and the US 50-50 Adjust system, with the remaining four systems (the three Swedish variants and the US Benefit Adjust system) performing better. Note that the systems with the highest variances are those that adjust taxes, rather than benefits. Presumably, this reflects the fact that adjustments occurring earlier in life are less heavily discounted than those occurring later.

Based on these descriptive statistics, particularly the *NPV*, the Swedish system stands out for its low uncertainty but also for slightly lower mean performance. The expected utility measures to which we turn shortly take both into account in constructing a single measure. But first, we consider the effects of cohort size.

The Effect of Population Weighting on NPV and IRR Outcomes

Comparing Tables 1 and 2 we can calculate that with the US Tax Adjust system, population weighting raises average values of both the *NPV* and the *IRR* by about 10-15 percent relative to their values with no weighting, consistent with the earlier discussion about population weighting. With the US Benefit Adjust system, population weighting actually lowers the *NPV* and the *IRR* by about 1-2 percent, reflecting the fact that tax adjustments are more favorable to large cohorts than are benefit adjustments. The US 50-50 Adjust system lies in between, with smaller increases in the average *NPV* and *IRR* than the US Tax Adjust plan. The German system, which resembles the 50-50 Adjust plan in that it involves annual adjustments in both taxes and benefits, follows a similar pattern, although its somewhat larger differences between Tables 1 and 2 suggest that the German system lies somewhere in between the US 50-50 Adjust system

and the US Tax Adjust system in terms of the relative burden of adjustment placed on taxes and benefits. For the Swedish plans, the corresponding changes are generally small and not systematic, as these plans allocate different shares of shock absorption to workers and retirees, depending on each system's detailed specification.

The previous discussion referred to average values, but one can also argue that any fluctuation in cost of benefits will require a smaller adjustment in tax rates under the Tax Adjust system if a taxpaying generation is large, so we might expect the variance of *NPV* or *IRR* to be smaller under population weighting for the Tax Adjust system, and indeed it is. For the Benefit Adjust scheme, the pattern is less obvious, consistent with the results for average values suggesting that large and small generations are not treated very differently under this scheme. For the German system, as for the US 50-50 Adjust system, the impact of population weighting on variances falls in between that of the US Tax Adjust and US Benefit Adjust systems.

Expected Utility

In looking at our *IRR* and *NPV* measures, we have considered both mean values and variances as a way of describing the trade-offs of the different systems. However, these summary measures cannot adequately characterize impacts on individual welfare, because one cannot weigh the trade-off between mean and variance without an explicit utility function, and even then one needs more than these two moments to assess the impact on utility. For example, a high variance in the *NPV* could be helpful if the upper tail of its distribution coincides with states of nature in which wage growth is below its mean and thereby insures, rather than exacerbates, lifetime income risk. We have, therefore, developed a methodology for approximating the incremental impact on expected utility of any particular pension system, as described earlier and developed in more detail in the Appendix. As already discussed, this

method yields the *NPV* measure for the special case of risk-neutrality, but our motivation for deriving it is to consider the case of risk aversion.

Looking first at expected utility under population weighting, in Table 1, we can see a number of significant changes in the relative standing of the different plans. First, the actual Swedish system gains relative to the US and German plans, reflecting the smoothing of outcomes that its asset accumulation permits. Second, among the US plans, the Benefit Adjust plan fares best and the Tax Adjust Plan worst, a ranking that reflects the greater variance of the latter.

Our variants of the Swedish system include two NDC(*g*) versions with different severities of the symmetric brake: $A=.5$ and 1.0 . One might expect that the weaker ($A=.5$) brake would involve more risk sharing, since response to imbalances is slower. This is the case in Table 1 for $\gamma=3$ and for $\gamma=5$, although the differences are quite small. In fact, comparison of other outcome measures in Tables 1 and 2 shows that performances of these two systems are generally very similar except that when $A=.5$ the variance of the *IRR* becomes large.

The next two rows of Table 1 present some sensitivity analysis for the case of risk aversion. In the first of these rows, we assume that the marginal utility of consumption for retirees (see expression (A10') in the Appendix) is based on the wage rate at the cohort's age of retirement rather than at its entry into the labor force.⁹ While this variation in assumptions lowers expected utility, it has little impact on the relative attractiveness of the different systems. The final row of the table illustrates what happens when we assume that risk aversion changes over the course of life, in particular assuming that the coefficient of risk aversion is higher ($\gamma =$

⁹ In the notation used in expression (A10'), the wage w_s rather than w_t is used.

5) in retirement than during working life ($\gamma = 3$).¹⁰ Here, there are interesting patterns in the relative performance of the different systems. The smallest reductions in expected utility relative to the case for which $\gamma = 3$ throughout life are for the three NDC systems, while the biggest impact is for the German system. Throughout the sensitivity analysis, though, the NDC(g) systems with $A = .5$ and $A = 1$ provide the highest level of expected utility, and the US Tax Adjust system or the German system the second-lowest.

These conclusions change, however, when we don't weight by the size of generation, as seen in Table 2. In this case the NDC($n+g$) with symmetric brake ($A = .5$) does better for all four cases. We might expect this program to do well, because its rate of return reflects both productivity growth and labor force growth. It and the other NDC programs with symmetric brakes dominate all other pension forms. Indeed, the basic Swedish system also dominates some of the other systems, in particular the US Tax Adjust and German plan. Although we think the results in Table 1 are more relevant, the differences between the two tables are informative, confirming that systems that focus exclusively on benefit reductions (i.e., the Swedish system and its variants) look relatively better when more weight is given to smaller generations, because benefit reductions tend to hit larger cohorts whereas tax increases tend to hit smaller cohorts. This effect can also be seen by looking at the relative impact for the three US systems in moving from Table 1 to Table 2, with the biggest drop occurring for the US Tax Adjust system, followed by the US 50-50 Adjust system and the US Benefit Adjust system.

¹⁰ There is some evidence for such a pattern of increasing risk aversion, although the evidence is not overwhelming. See Poterba (2001).

Horizontal Equity

For reasons discussed earlier, we also believe it may be useful to evaluate the programs from the point of view of a social welfare function that weights variations in outcome more heavily when they affect nearby generations differently, since nearby generations seem likely to be the reference group for assessing the fairness of outcomes. This is the purpose of the Horizontal Equity measure. Referring first to the results for the population weighted measures in Table 1, we see in the last line of the table the Horizontal Equity results for the *NPV*.¹¹ In the table, the gap between this measure and 1.0 is the fraction of lifetime income that society would be willing to pay in order to remove horizontal inequity. The higher the value in the table, the greater is Horizontal Equity. The results are clear: the actual Swedish system dominates the other systems, its treatment of contemporaneous generations smoother than any other. The NDC systems and US Benefit Adjust systems do the next best. For the case in which generations are not weighted by population (Table 2) the Swedish systems based on the wage rate (g) appear slightly preferable to that based on total wages ($n+g$), but the difference is small there and even smaller when one weights by population (Table 1).

Social Welfare Calculations

As discussed earlier, we have attempted to scale the different public pension systems being considered here in order to make it easier to compare their long-run impact. In particular, we have kept the average size (in terms of tax payments) the same, to avoid confusing the systems' relative performance with the fact that PAYGO pensions yield below-market rates of return. However, as came up especially when contrasting the performance of the US Benefit Adjust and US Tax Adjust plans, systems that impose more of the adjustment risk on benefits

¹¹ The measure as originally developed is not easily applied for our expected utility measure with risk aversion.

also end up imposing more risk on initial transition generations. Therefore, leaving such initial generations out of the calculations may bias our conclusions in favor of systems that adjust benefits, even if all systems are scaled to have the same average size.

In order to deal with this potential bias, we expand our NPV and EU calculations to include all transition generations alive during our simulation period, including those already alive at the beginning of the transition and those still alive at the end. Following the methodology laid out in the Appendix, we calculate partial values of NPV and EU for these cohorts and then aggregate these with the corresponding values for complete generations, discounting everything back to the initial year of our simulations to get a single social welfare measure for each system. By construction, our social welfare measure will equal zero under zero risk aversion for any system that maintains a balance of zero assets throughout the simulation, because it then reduces to the discounted sum of benefits less taxes in all years.

The results of these calculations are provided in Table 3. The first row of the table gives the results for the case of risk neutrality. Except for the actual Swedish system, the results are all slightly positive, because the systems all start with the same small positive asset balance. For the US and German plans, the values are identical, because these plans maintain a constant asset-payroll ratio throughout each simulation trajectory. For the NDC variants of the Swedish system, the values are positive but slightly higher, suggesting that assets fall slightly from their initial value, on average. For the actual Swedish system, the value is negative because this system accumulates assets on average. To neutralize these small differences among the systems related to asset drift, we adjust the numbers in the table by these differences using the NDC(g) plan as the benchmark for adjustment. The adjusted version of the table is presented in the lower

panel of the table. Note that we do not adjust the actual Swedish system, because this system accumulates assets by design.

Focusing on the adjusted values in Table 3, we see once again that the actual Swedish system fares relatively better as risk aversion is taken into account. Unlike in the other tables, however, this system continues to perform much worse even when risk aversion is high. This relatively poor performance when transition generations are included makes sense, because the Swedish system's initial phase is one in which the brake mechanism is frequently in place as a buffer of assets is accumulated. The initial cohorts therefore are more likely to suffer under this initial adjustment process and to get little benefit from the subsequent reduction in volatility that this buffer provides.

Looking now at the US systems, we see that the apparent advantage of the Benefit Adjust plan has disappeared. This plan now always performs worse than the US 50-50 Adjust plan, which now is the preferred plan among the three US plans. Taking the impact on initial generations into account causes the Benefit Adjust plan to lose its apparent advantage. Indeed, when risk aversion is higher in old age ($\gamma = 5$ vs. $\gamma = 3$), the Benefit Adjust plan fares worse than the Tax Adjust plan, its reliance on adjusting benefits poorly suited to a situation in which beneficiaries are more risk averse than workers. That the 50-50 Adjust plan performs better than either of the other two also makes sense, in that it spreads the impact of each adjustment over more generations than either of the other plans. The German plan, which also distributes its annual adjustment over both beneficiaries and workers, performs only slightly worse than the US 50-50 Adjust plan and always better than the other two US plans.

Indeed, the US 50-50 Adjust plan is no longer dominated by the Swedish NDC plans, its performance now lying in the same range as these plans. Although the NDC plans adjust only

benefits, they spread each year's adjustment over a larger number of cohorts, including not just current beneficiaries but also future ones, i.e., current workers. Among the Swedish plans, there is a distinct partial ranking. While the strength of the brake ($A = .5$ vs. $A = 1$) still has virtually no impact, the inclusion of population fluctuations in the rate-of-return calculation ($n+g$ vs. g) has a distinctly negative impact on social welfare. Recall our previous finding that an NDC system based on the growth of wages, rather than of the wage rate, is inherently more stable and relies less on the brake mechanism for stability. Our findings here suggest that the added volatility in benefit accruals under the NDC($n+g$) system when the brake is not in place outweighs the volatility imposed by more frequent application of the brake under the NDC(g) system.

Discussion and Conclusions

The NDC systems aim to pay a rate of return to contributors that is warranted by the macroeconomic/demographic environment. However, Sweden, in setting up its system, chose to make that rate of return equal the rate of wage growth, g , rather than $n+g$ which is the rate payable in steady state. Because they also included a brake mechanism in their system design, if labor force growth should drop below 0 then the brake would eventually automatically reduce the rate of return below g . Our analysis shows that this program design insulates participating generations from variations in the economic/demographic environment. The asymmetric brake, which reduces the rate of return in some circumstances but never raises it, apparently plays a key role. This arrangement permits the system to accumulate undistributed assets and therefore makes it yield a lower mean NPV and IRR compared to NDC(g) systems with a symmetric brake. But, by accumulating more assets, it avoids having to apply the brake and thereby leaves the rate of return more stable along a trajectory. This makes the Swedish system look relatively

better when risk aversion is explicitly included in the calculation of expected utility, but the net benefit appears smaller once the welfare of initial transition generations is taken into account, for these are the generations that bear the brunt of the Swedish system's buffer stock accumulation.

Among other plans, the apparent advantage of the US Benefit Adjust plan over other stable variants of the US system disappears when transition generations are taken into account, and one finds that the US 50-50 Adjust system performs up to the standards of the NDC systems, all systems that distribute annual shocks among both workers and retirees. The German system resembles the US 50-50 Adjust system in many respects but its performance suggests that it places a higher relative burden of risk bearing on workers and spreads risk somewhat less efficiently. Among the NDC plans, there is relatively little difference apparent until transition generations are taken into account, at which point the NDC(g) systems look somewhat better than the inherently more stable NDC($n+g$) systems. This suggests that shifting more of fiscal adjustment to the brake mechanisms may also improve risk spreading.

Our results suggest, then, that spreading risk widely among generations improves welfare, and that the policy of reducing risk through asset accumulation, as the Swedish system does, offers a less attractive approach unless one places very high weight on horizontal equity, i.e., on maintaining a very smooth pattern of net benefits from one cohort to the next.

In future work, we hope to look more closely at the differences in performance we have uncovered here by looking at how the different systems distribute different types of shocks among cohorts. This will serve not only to confirm (or correct) some of the intuition provided here, but also to help us understand the extent to which different approaches might be used according to the source, strength and stochastic properties of shocks.

References

- Alho, Juha M., Svend E. Hougaard Jensen and Jukka Lassila eds. (in press) *Uncertain Demographics and Fiscal Sustainability* (Cambridge University Press).
- Alho J.M., J. Lassila and T. Valkonen (2005) Demographic Uncertainty and Evaluation of Sustainability of Pension Systems, pp. 95-115 in Holtzmann R. and Palmer E. (Eds.) *Non-Financial Defined Contribution (NDC) Pension Schemes: Concept, Issues, Implementation, Prospects*. World Bank, Washington D.C.
- Auerbach, Alan J., and Kevin Hassett (2002) "A New Measure of Horizontal Equity," *American Economic Review* (September) pp. 1116-1125.
- Auerbach, Alan J. and Ronald D. Lee (2008) "Notional Defined Contribution Pension Systems in a Stochastic Context: Design and Stability," in J. Brown, J. Liebman, and D. Wise, eds., *Social Security Policy in a Changing Environment*. University of Chicago Press: Chicago, forthcoming.
- Börsch-Supan, Axel, and Christina B. Wilke, 2003, "The German Public Pension System: How it Was, How it Will Be"
- Börsch-Supan, Axel, Annette Reil-Held and Christina Benita Wilke, 2003, "How to make a Defined Benefit System Sustainable: The 'Sustainability Factor' in the German Benefit Indexation Formula"
- Gruber, Jonathan, and David A. Wise (1999) *Social Security Around the World*. University of Chicago Press: Chicago.
- Holtzmann R. and Palmer E. (Eds.) (2005) *Non-Financial Defined Contribution (NDC) Pension Schemes: Concept, Issues, Implementation, Prospects*. World Bank, Washington D.C.
- Lee, Ronald (1993) "Modeling and Forecasting the Time Series of US Fertility: Age Patterns, Range, and Ultimate Level," *International Journal of Forecasting* v.9 pp.187-202.
- Lee, Ronald (2007) National Transfer Account Project, data for United States, 2003. <http://www.ntaccounts.org>.
- Lee, Ronald and Lawrence Carter (1992) "Modeling and Forecasting U.S. Mortality," *Journal of the American Statistical Association* v.87 n.419 (September 1992), pp.659-671, and "Rejoinder," same issue, pp.674-675.
- Lee, Ronald and Shripad Tuljapurkar (1994), "Stochastic Population Projections for the United States: Beyond High, Medium and Low," *Journal of the American Statistical Association*, v.89 n.428 (December) 1175-1189.

- Lee, Ronald and Shripad Tuljapurkar (1998) “Stochastic Forecasts for Social Security,” in David Wise, ed., *Frontiers in the Economics of Aging*. Chicago: University of Chicago Press, pp. 393-420.
- Lee, Ronald D., Michael W. Anderson, and Shripad Tuljapurkar (2003) “Stochastic Forecasts of the Social Security Trust Fund” Report for the Social Security Administration (January) posted on the web site of the Office of the Actuary.
- Lee, Ronald (2005) “Comments on ‘The Rate of Return of Pay As You Go Pension Systems’” in Holtzmann R. and Palmer E. (Eds.) (2005) *Non-Financial Defined Contribution (NDC) Pension Schemes: Concept, Issues, Implementation, Prospects*. World Bank, Washington D.C.
- Lee, Ronald, Timothy Miller and Michael Anderson (2004) “Stochastic Infinite Horizon Forecasts for Social Security and Related Studies,” National Bureau of Economic Research Working Paper No. 10917 (<http://papers.nber.org/papers/w10917.pdf>).
- Legros, Florence (2003) “Notional Defined Contribution: A Comparison of the French and German Point Systems”, paper presented at the World Bank and RFV Conference on NDC Pension Schemes, Sandhamn, Sweden (Sept. 28-30, 2003).
- Ludwig, Alexander and Michael Reiter, 2008, “Sharing Demographic Risk – Who is Afraid of the Baby Bust?” Mannheim Research Institute for the Economics of Aging working paper.
- Poterba, James M., 2001, “Demographic Structure and Asset Returns,” *Review of Economics and Statistics* 83(4), November, pp. 565-584.

Appendix. Valuing Flows from Social Security

In theory, the preferred way to evaluate social security is to specify the household's other sources of income and then solve for its optimal consumption and portfolio choice behavior as a function of state variables at each date. As part of this dynamic programming solution, we would also obtain the household's value function as of date t . Solving for the value function in the presence and absence of social security would give us the value the household would place on the social security system. This approach is not feasible, however, because of the very large number of state variables involved and the complicated methods of calculating social security benefits, particularly under the Swedish system and its balancing mechanism. Thus some simplification is necessary. The approach we take here is to treat differences among systems as small deviations and use second-order Taylor approximations in combination with assumptions about the marginal utility of consumption at the point of approximation.¹²

Before specifying our proposed methodology, it is useful to distinguish two ways in which risk aversion will affect the valuation of taxes and benefits:

1. The household will value flows differently in different states of nature (assuming that there is no perfect insurance against variations in productivity, etc.); and
2. The household will be averse to fluctuations in benefits and/or taxes, even absent other sources of income fluctuations.

¹² An alternative approach, taken by Ludwig and Reiter (2008) to analyze the German system, is to use a linear-quadratic approximation of the government's problem to derive linear decision rules and to evaluate the second-order costs of policy fluctuations from the approximately optimal rules.

Proposed Methodology

In the presence of uncertainty, the Euler equation for household optimization would imply that the expected marginal utility for $s > t$ would relate to marginal utility in the year of birth, t , by:

$$(A1) \quad U'_t = E_t \left\{ \left[\prod_{u=t}^s (1 + r_u) \right] U'_s \right\}.$$

We don't know these state-contingent marginal utilities without solving the full optimization problem, so we make a simplifying assumption, that marginal utility at date s in state i is proportional to some function of the vector of state variables at date s , (including the level of productivity, interest rates, population, etc.) $x_{s,i}$:

$$(A2) \quad U'_{s,i} \sim h_s(x_{s,i}) = a_s h_s(x_{s,i}),$$

where a_s is a constant at date s . The idea here is that incremental additions to or subtractions from resources will vary across states at a given date, these marginal valuations being higher in bad states (e.g., states with low levels of productivity) than in good ones. Note that the function is subscripted by date, indicating that it may differ over age. For example, an individual's marginal utility may be more sensitive to the economy's level of productivity during working years, when most resources come from wages, rather than during retirement years. We will discuss the specification of $h_s(\cdot)$ further below.

Substituting (A2) into (A1), we get:

$$(A3) \quad U'_t = E_t \left\{ \left[\prod_{u=t}^s (1 + r_u) \right] a_s h_s(x_s) \right\}.$$

Without loss of generality, we can normalize the utility function so that $a_t = 1$ and hence

$U'_t = h_t(x_t)$. This normalization then gives us the solutions for a_s at each date s :

$$(A4) \quad a_s = \frac{1}{E_t \left\{ \left[\prod_{u=t}^s (1 + r_u) \right] \frac{h_s(x_s)}{h_t(x_t)} \right\}}.$$

From (A2) and (A4), we have the solutions for marginal utility,

$$(A5) \quad U'_{s,i} = \frac{h_{s,i}(x_{s,i})}{E_t \left\{ \left[\prod_{u=t}^s (1 + r_u) \right] \frac{h_s(x_s)}{h_t(x_t)} \right\}}.$$

Calculating the Value of Social Security Taxes and Benefits

The value of taxes and benefits for the generation born in year t equals the change in utility associated with the flows of taxes and benefits, or

$$(A6) \quad V_t = \sum_i \pi_{t,i} \sum_{s=t}^{t+T} P_{s,i} \left[U \left(c_{s,i}^* + \frac{F_{s,i}}{P_{s,i}} \right) - U(c_{s,i}^*) \right],$$

where T is the maximum number of years of life, F equals a cohort's social security flows (either minus taxes or plus benefits) per original member, $c_{s,i}^*$ is some benchmark level of per capita consumption and $P_{s,i}$ is the surviving fraction of the population in state i at date s . The weights $\pi_{t,i}$ equal the cohort's initial population along trajectory i as a fraction of the sum of the cohort's initial populations along all trajectories. (We also provide calculations for the case in which we

ignore differences in initial populations within and across cohorts and set $\pi_{t,i} \equiv 1/N$, where N is the number of trajectories.)

Because we will be evaluating this difference in utilities using a Taylor approximation, it is better to consider small variations, so we will look not at V_t as specified in expression (A6), but at the difference between V_t and the value of some benchmark system of constant taxes and benefits,

$$(A7) \quad \Delta V_t = \sum_i \pi_{t,i} \sum_{s=t}^{t+T} P_{s,i} \left[U \left(c_{s,i}^* + \frac{F_{s,i}}{P_{s,i}} \right) - U \left(c_{s,i}^* + \frac{\bar{F}_s}{P_{s,i}} \right) \right].$$

Letting $f_{s,i}$ equal per capita flows, and taking second-order Taylor approximations around the benchmark social security system of the utility variations in (A7), we get:

$$(A8) \quad \Delta V_t = \sum_i \pi_{t,i} \sum_{s=t}^{t+T} P_{s,i} \left[U'_{s,i} \cdot (f_{s,i} - \bar{f}_s) + \frac{1}{2} U''_{s,i} \cdot (f_{s,i} - \bar{f}_s)^2 \right].$$

Assuming that households have CES utility with risk-aversion parameter γ , we know that

$$U'' = -\gamma \frac{U'}{c}, \text{ where } c \text{ is consumption including the base level of social security flows, } c^* + \bar{f},$$

around which the Taylor approximation is being taken. Thus, (A8) can be rewritten:

$$(A9) \quad \Delta V_t = \sum_i \pi_{t,i} \sum_{s=t}^{t+T} P_{s,i} U'_{s,i} \cdot \left[(f_{s,i} - \bar{f}_s) - \frac{1}{2} \gamma \frac{(f_{s,i} - \bar{f}_s)^2}{c_{s,i}} \right].$$

Without loss of generality, we can drop the term \bar{f}_s when it first appears in (A9), because this is a constant term that does not vary across social security systems. Dropping this term, using the facts that $f = F/P$ and $c = C/P$, and substituting (A5) into (A9) yields:

$$(A10) \quad \Delta V_t = \sum_i \pi_{t,i} \sum_{s=t}^{t+T} \left\{ \frac{h_{s,i}(x_{s,i})}{E_t \left[\prod_{u=t}^s (1+r_u) \frac{h_s(x_s)}{h_t(x_t)} \right]} \right\} \cdot \left[F_{s,i} - \frac{1}{2} \gamma \frac{(F_{s,i} - \bar{F}_s)^2}{C_{s,i}} \right],$$

where C is a generation's consumption per initial individual at the benchmark level.

Expression (A10) will be the basis for our valuation of flows. As discussed in the introduction, risk aversion affects valuation in two ways, through the variation in the value taken by the function h (the term in curly brackets in A10) and through the impact of fluctuations on the flows themselves (the next term).

As a benchmark, note that for risk-neutrality, $h(x) \equiv 1$ and $\gamma = 0$, so (A10) reduces to:

$$(A11) \quad \Delta V_t = \sum_i \pi_{t,i} \sum_{s=t}^{t+T} \frac{1}{E_t \left[\prod_{u=t}^s (1+r_u) \right]} F_{s,i}.$$

That is, under risk-neutrality, we should divide the average flow at each date by the average discount factor.¹³ After summing over the trajectories for each generation in (A11), we will divide by that generation's present expected discounted value of earnings,

$$(A12) \quad PDVE_t = \sum_i \pi_{t,i} \sum_{s=t}^{t+R} \frac{1}{E_t \left[\prod_{u=t}^s (1+r_u) \right]} Y_{s,i},$$

¹³ Note: we are implicitly assuming access to annuity markets; had we not, then there would be an extra P_s multiplied by the discount factors, so that (A11) would have become:

$$(A11') \quad \Delta V_t = \sum_i \pi_{t,i} \sum_{s=t}^{t+T} \frac{1}{E_t \left[\prod_{u=t}^s (1+r_u) P_s \right]} F_{s,i}.$$

where R is the retirement age and $Y_{s,i}$ is the generation's earnings in year s along path i . This normalization serves two purposes. First, it removes the productivity growth trend to avoid giving more weight to later generations when we calculate an average over generations. Second, it scales ΔV so that it is expressed as a fraction of expected lifetime earnings. We average these generation-specific ratios by average initial generation size to form an average estimate of expected utility.

Parameterization

To implement expression (A10), we need to make three sets of parameter assumptions.

Risk-aversion parameter, γ

We consider four cases, $\gamma = 0$ (neutrality) 3, and 5, and $\gamma = 3$ when working and 5 when retired.

State-contingent valuation

There are a variety of possibilities here. One is to assume that h is related to the contemporaneous wage, $h_{s,i} \sim w_{s,i}^{-\gamma}$, which would make sense if consumption were proportional to labor income. Another is to assume that h is related to the initial wage along the cohort's trajectory, $h_{s,i} \sim w_{t,i}^{-\gamma}$, which would take into account the fact that consumption is financed to some extent by past saving and social security benefits. The approach that we adopt as our base case is to assume that $h_{s,i} \sim w_{s,i}^{-\gamma}$, during working years and $h_{s,i} \sim w_{t,i}^{-\gamma}$ during retirement years. As an alternative assumption, we assume that $h_{s,i}$ when retired is based on the wage in the last

year of work as might be more appropriate in an economy in which the well-being of the elderly depends on contemporaneous wages.

There remains the question of how to scale the marginal utilities of different generations. One approach would be to assume a constant utility function over time, in which case successive generations would, on average, have lower and lower levels of marginal utility as a consequence of trend productivity growth. This approach would tend to make social security systems that transfer resources from future generations to current ones look better than those that do not involve such transfers. Although such transfers would be relatively subtle for the systems we are considering here, we nevertheless wish to avoid confusing intergenerational transfers with risk sharing. Thus, we scale each generation's marginal utilities by that generation's average initial

wage, that is, $h_{s,i} = \left(\frac{w_{s,i}}{\bar{w}_t}\right)^{-\gamma}$ during working years and $h_{s,i} = \left(\frac{w_{t,i}}{\bar{w}_t}\right)^{-\gamma}$ during retirement years.

Benchmark level of consumption, C

Here, we use just two such numbers (relative to trend) to keep fixed across scenarios, one for workers (C^L) and one for retirees (C^R), rather than age-specific values. The calculation of C^L and C^R will be discussed shortly.

In summary, with our parameterization, (A10) becomes:

$$(A10') \quad \Delta V_t = \sum_i \pi_{t,i} \left\{ - \sum_{s=t}^{t+R} \frac{\left(\frac{w_{s,i}}{\bar{w}_t}\right)^{-\gamma} \left[T_{s,i} + \frac{\gamma(T_{s,i} - \bar{T}_s)^2}{2C_s^L} \right]}{E_t \left\{ \left[\prod_{u=t}^s (1+r_u) \right] \left(\frac{w_s}{w_t}\right)^{-\gamma} \right\}} + \sum_{s=t+R+1}^{t+T} \frac{\left(\frac{w_{t,i}}{\bar{w}_t}\right)^{-\gamma} \left[B_{s,i} - \frac{\gamma(B_{s,i} - \bar{B}_s)^2}{2C_s^R} \right]}{E_t \left\{ \left[\prod_{u=t}^s (1+r_u) \right] \right\}} \right\},$$

with everything except taxes, $T_{s,i}$, and benefits, $B_{s,i}$, the same across different social security scenarios.¹⁴ For the benchmark values of taxes and benefits, \bar{T}_s and \bar{B}_s , we use the average values for the US benefits-adjust system. Note that the average values are indexed by time, because they will follow the trend in productivity. To calculate C_s^L and C_s^R , we compute the ratios of taxes to consumption and adjusted benefits to consumption for the US population in 2003, based on populations aged 20-64 and over 65 (excluding those in nursing homes), respectively, using OASI payroll taxes and benefits and adjusting benefits (and consumption of beneficiaries) downward until they equal taxes in the aggregate.¹⁵ The resulting ratios are .103 for taxes relative to non-retiree consumption, and .235 for adjusted benefits relative to adjusted retiree consumption. We multiply the inverses of these ratios by \bar{T}_s and \bar{B}_s in a given year to get the values of worker and retiree consumption around which the expected utility approximation is computed.

Finally, as in the case of risk neutrality, we divide expression (A10') by the cohort's present expected discounted value of earnings, as given in (A12), in order to weight the results equally across generations and express them as a share of lifetime earnings.

Estimating Social Welfare

Our methodology compares social security systems that are normalized to be of the same size (in terms of taxes relative to earnings) because larger systems, which make larger transfers to initial transition generations, provide lower present value returns to the generations that follow. However, even this normalization may fail to neutralize all differences in the relative

¹⁴ Note that in case where γ is allowed to differ between work years and retirement, the discount factor in the second term on the right-hand side of (A10') will include another term that reflects both values of γ .

¹⁵ The data for this calculation are based on the 2003 Consumer Expenditure Survey and other sources of data as detailed in the US National Transfer Accounts (NTA) at <http://www.schemearts.com/proj/nta>.

treatment of transition generations. In particular, systems that rely on immediate adjustments to benefits (in particular the US Benefit Adjust system) will impose greater uncertainty on initial transition generations, and in so doing may impose less uncertainty on subsequent generations. Therefore, such systems may appear more attractive than they actually are if we ignore the effects on the welfare of transition generations. Note that this is not simply an issue of what is being adjusted (taxes or benefits), but rather of which generations are affected. For example, all variants of the Swedish system also rely solely on benefit adjustments, but these adjustments are spread over more generations. That is, while the US Benefit Adjust system achieves current cash flow balance by reducing the benefits of current retirees, the Swedish systems also reduce the future benefits of current workers by reducing their current accumulations of notional pension wealth.

To deal with this issue, we construct an expanded welfare measure that takes account of transition generations, both at the beginning and the end of our simulation period. Our basic approach is to use a discounted sum of the values of ΔV as calculated above and to include as well the partial values for initial and terminal cohorts.

For full generations whose birth occurs after our initial simulation year 0 and whose final death is before year L (the last year of our computation), we simply take the values of ΔV but discount them back to year 0; expression (A10') becomes:

$$(A13) \quad \Delta V_t^F = \sum_i \pi_{t,i} \sum_{s=0}^{t+T} \left\{ \frac{h_{s,i}(x_{s,i})}{E_0 \left\{ \left[\prod_{u=0}^s (1+r_u) \right] \frac{h_s(x_s)}{h_t(x_t)} \right\}} \right\} \cdot \left[F_{s,i} - \frac{1}{2} \gamma \frac{(F_{s,i} - \bar{F}_s)^2}{C_{s,i}} \right]$$

For generations born in year $t > L-T$, which will still be alive at the end of our period of measurement, we compute partial sums, starting in the year of birth and going through year L , and discount from the year of birth back to year 0:

$$(A14) \quad \Delta V_t^P = \sum_i \pi_{t,i} \sum_{s=t}^L \left\{ \frac{h_{s,i}(x_{s,i})}{E_0 \left\{ \left[\prod_{u=0}^s (1+r_u) \right] \frac{h_s(x_s)}{h_t(x_t)} \right\}} \right\} \cdot \left[F_{s,i} - \frac{1}{2} \gamma \frac{(F_{s,i} - \bar{F}_s)^2}{C_{s,i}} \right]$$

For existing generations, born in year $t < 0$, who are already alive as of year 0, we define a partial value of ΔV , say ΔV^E , as

$$(A15) \quad \Delta V_t^E = \sum_i \pi_{0,i} \sum_{s=0}^{t+T} \left\{ \frac{h_{s,i}(x_{s,i})}{E_0 \left\{ \left[\prod_{u=0}^s (1+r_u) \right] \frac{h_s(x_s)}{h_0(x_0)} \right\}} \right\} \cdot \left[F_{s,i} - \frac{1}{2} \gamma \frac{(F_{s,i} - \bar{F}_s)^2}{C_{s,i}} \right]$$

That is, we treat these cohorts as if they are born in year 0 with the surviving population as of year 0 and with $T+t < T$ years of life remaining.¹⁶

We sum over the three groups to get the sum of the ΔV s, weighting each generation by its average initial size, and then divide by the sum of the present values of earnings for the different cohorts, computed for the three groups in parallel fashion.

¹⁶ Note that the terms $h_{x_t}(x_{x,i})$ are normalized relative to year 0 wages in this case, to be consistent with the calculation starting in year 0.

Table 1
Summary Statistics for Eight Pension Plans Based on 1000 Stochastic Paths of 500 Years
 (Weighted by Generation Size)

	US - Tax Adjust (Scaled)	US - Benefit Adjust	US - 50-50 Adjust (scaled)	NDC Sweden	NDC(g) Symmetric Brake A=5	NDC(g) Symmetric Brake A=1	NDC(n+g) Symmetric Brake A=5	German
<u>IRR</u>								
Mean	0.0129	0.0114	0.0122	0.0104	0.0125	0.0127	0.0131	0.0124
Median	0.0132	0.0116	0.0123	0.0108	0.0135	0.0133	0.0132	0.0126
Variance across trajectories generations	0.000058 0.000050	0.000067 0.000057	0.000050 0.000042	0.000353 0.000023	0.000707 0.000387	0.000438 0.000091	0.000063 0.000052	0.000052 0.000043
<u>NPV</u>								
Mean	-0.0593	-0.0605	-0.0599	-0.0627	-0.0572	-0.0576	-0.0577	-0.0602
Variance across trajectories generations	0.00165 0.00101	0.00106 0.00046	0.00122 0.00061	0.00098 0.00034	0.00103 0.00049	0.00100 0.00046	0.00103 0.00049	0.00143 0.00074
<u>EU</u>								
$\gamma=3$	-0.0809	-0.0782	-0.0792	-0.0801	-0.0751	-0.0753	-0.0758	-0.0792
$\gamma=5$	-0.1320	-0.1270	-0.1291	-0.1300	-0.1232	-0.1233	-0.1244	-0.1295
$\gamma=3, \text{ret. } w$	-0.0875	-0.0856	-0.0862	-0.0866	-0.0824	-0.0825	-0.0831	-0.0857
$\gamma=3, 5$	-0.1039	-0.0993	-0.1010	-0.1057	-0.0908	-0.0913	-0.0935	-0.1045
<u>HE</u>								
NPV	0.999621	0.999893	0.999848	0.999962	0.999885	0.999895	0.999882	0.999819

Table 2
Summary Statistics for Eight Pension Plans Based on 1000 Stochastic Paths of 500 Years
 (Not Weighted by Generation Size)

	US - Tax Adjust (Scaled)	US - Benefit Adjust	US - 50-50 Adjust (scaled)	NDC Sweden	NDC(g) Symmetric Brake A=5	NDC(g) Symmetric Brake A=1	NDC(n+g) Symmetric Brake A=5	German
<u>IRR</u>								
Mean	0.0112	0.0115	0.0114	0.0100	0.0106	0.0114	0.0133	0.0113
Median	0.0114	0.0118	0.0115	0.0109	0.0129	0.0129	0.0134	0.0115
Variance across trajectories generations	0.000060 0.000053	0.000064 0.000057	0.000050 0.000043	0.001048 0.000025	0.001783 0.000601	0.001225 0.000168	0.000061 0.000053	0.000051 0.000044
<u>NPV</u>								
Mean	-0.0669	-0.0601	-0.0637	-0.0620	-0.0582	-0.0583	-0.0570	-0.0673
Variance across trajectories generations	0.00206 0.00130	0.00102 0.00048	0.00137 0.00074	0.00095 0.00039	0.00102 0.00051	0.00098 0.00048	0.00101 0.00051	0.00178 0.00095
<u>EU</u>								
$\gamma = 3$	-0.0892	-0.0789	-0.0838	-0.0807	-0.0776	-0.0773	-0.0763	-0.0864
$\gamma = 5$	-0.1488	-0.1379	-0.1430	-0.1404	-0.1366	-0.1359	-0.1356	-0.1444
$\gamma = 3, \text{ret. } w$	-0.0957	-0.0859	-0.0906	-0.0871	-0.0844	-0.0841	-0.0832	-0.0928
$\gamma = 3, 5$	-0.1288	-0.1158	-0.1219	-0.1220	-0.1122	-0.1108	-0.1102	-0.1256
<u>HE</u>								
NPV	0.999563	0.999897	0.999834	0.999952	0.999893	0.999902	0.999880	0.999788

Table 3
Social Welfare Calculations for Eight Pension Plans
 (Weighted by Generation Size)

Unadjusted

	US - Tax Adjust (Scaled)	US - Benefit Adjust	US - 50-50 Adjust (scaled)	NDC Sweden	NDC(g) Symmetric Brake A=5	NDC(g) Symmetric Brake A=1	NDC(n+g) Symmetric Brake A=5	German
$\gamma=0$	0.00140	0.00140	0.00140	-0.00878	0.00189	0.00186	0.00181	0.00140
$\gamma=3$	-0.00406	-0.00308	-0.00273	-0.01175	-0.00226	-0.00226	-0.00252	-0.00287
$\gamma=5$	-0.02109	-0.01923	-0.01890	-0.02645	-0.01792	-0.01788	-0.01840	-0.01921
$\gamma=3, \text{ret. } w$	-0.01014	-0.00963	-0.00911	-0.01787	-0.00841	-0.00844	-0.00878	-0.00882
$\gamma=3, 5$	-0.00120	-0.00156	-0.00007	-0.00990	-0.00026	-0.00022	-0.00074	-0.00110

Adjusted for Initial Differences Under Risk-Neutrality

	US - Tax Adjust (Scaled)	US - Benefit Adjust	US - 50-50 Adjust (scaled)	NDC Sweden	NDC(g) Symmetric Brake A=5	NDC(g) Symmetric Brake A=1	NDC(n+g) Symmetric Brake A=5	German
$\gamma=0$	0.00186	0.00186	0.00186	-0.00878	0.00186	0.00186	0.00186	0.00186
$\gamma=3$	-0.00360	-0.00262	-0.00227	-0.01175	-0.00229	-0.00226	-0.00248	-0.00241
$\gamma=5$	-0.02063	-0.01877	-0.01844	-0.02645	-0.01796	-0.01788	-0.01835	-0.01875
$\gamma=3, \text{ret. } w$	-0.00968	-0.00917	-0.00865	-0.01787	-0.00844	-0.00844	-0.00873	-0.00836
$\gamma=3, 5$	-0.00074	-0.00110	0.00039	-0.00990	-0.00029	-0.00022	-0.00070	-0.00064

Figure 1. Ratio of Retirees to Workers, 15 Sample Paths

