Using a Life-Cycle Model to Predict Induced Entry Effects of a $1 for $2 Benefit Offset in the SSDI Program

Hugo Benítez-Silva, *SUNY Stony Brook*
Moshe Buchinsky, *UCLA and NBER*
John Rust, *University of Maryland and NBER*

First version: November, 2003
This version: June, 2004

1 Introduction

There have been various policy changes that have been considered by the Social Security Administration (SSA). One such policy change in the Social Security Disability Insurance (SSDI) program is what is termed internally as the "$1 for $2 benefit offset". This policy has not yet been enacted, but it has been actively supported by disability advocates as a way to provide greater incentives to return to work for SSDI beneficiaries who have fully, or partially, recovered from their disabilities.

Under the current law SSDI recipients who return to work during a 9 month "trial work period" (TWP) face no loss of benefits. However, continued work beyond the TWP, and an additional three-month grace period (TMGP), that results in earnings that are higher than the "substantial gainful activity" (SGA) ceiling will result in termination from the roles.[1] Under the $1 for $2 offset, an SSDI recipient who earns more than SGA after the TWP and TMGP would not be terminated from the roles. Rather, their DI benefits would be reduced by $1 for every $2 earned above the SGA disregard. In simple terms, the $1 for $2 amounts to replacing the current 100% forfeiture of benefits due to earnings in excess of SGA, by a 50% surtax.

Clearly, the $1 for $2 offset would increase the overall level of generosity of the SSDI program, because under the suggested plan no SSDI recipient would be made worse off, while some

[1] The SGA is currently $800 per month, and for statutorily blind individuals the SGA is currently $1,300.

recipients would be strictly better off. This is particularly true for those who have fully, or partially, recovered and who would like to return to work to supplement their SSDI benefits.

While it seems likely that the $1 for $2 offset would increase the amount of labor supply by some SSDI recipients, it is less clear whether or not it would result in significant cost savings from permanent *induced exit* from SSDI. It is possible that the $1 for $2 offset might make it more comfortable for SSDI recipients to remain on the roles and supplement their DI benefits with earnings from a part-time job. Hence, it is plausible that the $1 for $2 offset might not generate an increase, and might even lead to a decrease, in the number of beneficiaries who exit SSDI after the TWP.

Perhaps of biggest concern to policy makers is the possibility that the increased generosity of SSDI under the $1 for $2 offset would result in significant *induced entry* by individuals who are not yet on SSDI, but who could be induced, on the margin, to apply because of the change in policy. Previous studies by the SSA and the Congressional Budget Office (CBO) provide very different forecasts of the magnitude of the induced entry effect. The CBO estimated that the $1 for $2 offset would result in an increase of 75,000 SSDI beneficiaries over a ten-year period, whereas the SSA estimated that it would increase SSDI roles by 400,000 over the same time period. There are correspondingly large differences in the two forecasts of the net increase in the costs of the DI program resulting from the $1 for $2 offset. The CBO predicts that over five years, the net costs will increase by $410 million, while the SSA predicts that costs will increase by $5.1 billion.

In this paper we use a prototype of an empirical life-cycle model that we are currently developing. The model is quite rich and can also be used to provide detailed predictions of the behavioral responses to a wide range of hypothetical changes in Social Security policies. In this paper we use a calibrated version of the life cycle model to predict and examine the behavioral responses to the proposed $1 for $2 benefit offset plan.

The $1 for $2 offset has already been implemented for the Supplemental Security Income (SSI) program, but with a lower disregard of $65 per month. However, since there are strict asset/income tests imposed on the SSI applicants, the induced entry effect is likely to be much smaller than it would be for the SSDI applicants. A study by Muller, Scott and Bye (1996) concluded that: “While the research could not definitively assess the direct impacts of the benefit offset, the analysis found no major shifts in either labor force participation or earnings levels associated with the time periods corresponding to the implementation of the offset" (Muller 2000, p.??). In contrast, a recent study by Neumark and Powers (2003) found significant labor supply disincentives, due to state supplements to the SSI benefits. These benefits appear large enough to completely swamp the effect of the $1 for $2 work incentives.

The idea of extending the $1 for $2 offset to the SSDI program is extremely popular with disability advocates. They claim that the threat of loss of benefits due to earnings in excess of the SGA after the TWP is the primary reason that SSDI beneficiaries who are able to return to work, not to do so. In 1999 President Clinton signed a Federal law mandating that the SSA undertake a "demonstration project" (i.e., a controlled randomized experiment), to estimate the magnitude of labor supply response and the level of induced entry that would likely occur under the $1 for $2 offset.[2]

It turns out that it is exceedingly difficult to measure induced entry via controlled experiments. We will discuss some of the possible reasons below. However, a panel of consultants chosen by SSA to evaluate the feasibility of measuring induced entry by way of randomized experiments concluded that: "The classical experimental designs considered have very serious defects and should not be used to study induced entry." (Tuma, 2001, p. v). This study also concluded that: "Valuable information on the impact of induced entry may be acquired at relatively modest cost through the use of dynamic modeling of individual behavior and through responses to a hypothetical survey, such as the NSHA survey currently planned." (Tuma, 2001, p. v.). In this paper we provide a detailed examination of the induce entry effect, and other effects, using the former approach, that is, we provide predictions of the effects of the $1 for $2 offset based on a dynamic model of individual behavior.

The dynamic model we employ is a specially formulated version of the classical life-cycle model that incorporates a realistic treatment of the Social Security rules, particularly with regard

[2] U.S. P.L. 106-70, section 302 specifies: "The Commissioner of Social Security shall conduct demonstration projects for the purpose of evaluating, through the collection of data, a program for title II beneficiaries (as defined in section 1148(k)(3) of the Social Security Act) under which benefits payable under section 223 of such Act, or under section 202 of such Act based on the beneficiary's disability are reduced by $1 for each $2 of the beneficiary's earnings that is above a level to be determined by the Commissioner. Such projects shall be conducted at a number of localities which the Commissioner shall determine is sufficient to adequately evaluate the appropriateness of national implementation of such a program. Such projects shall identify reductions in Federal expenditures that may result from the permanent implementation of such a program. . . . The demonstration projects developed under subsection (a) shall be of sufficient duration, shall be of sufficient scope, and shall be carried out on a wide enough scale to permit a thorough evaluation of the project to determine—(A) the effects, if any, of induced entry into the project and reduce exit from the project; . . . (C) the savings that accrue to the Federal Old-Age and Survivors Insurance Trust Fund, the Federal Disability Insurance Trust Fund, and other Federal programs under the project being tested." In 2001, the SSA announced a separate set of demonstration projects in the *Federal Register* designed to evaluate the effects of altering various SSI program rules to improve incentives to return to work.

to the SSDI program. We have calibrated the model so that the behaviors observed in a simulated population of life-cycle optimizers resembles the behaviors we observe for a sample of real individuals born between 1931 and 1941, from the Health and Retirement Survey (HRS). Even though, due to computational limitations, the life-cycle model used here makes some simplifying assumptions, it is sufficiently realistic to be able to provide a convincing illustration of the potential value of life-cycle models for policy forecasting and evaluation, in general, and specifically for evaluating the effects of the $1 for $2 offset alternative.

The life-cycle model predicts that the $1 for $2 offset provides a very effective labor supply incentive. Under the *status quo* simulation of the model, we find that approximately 9.5% of the SSDI recipients eventually return to work (via the TWP). In sharp contrast, under the $1 for $2 offset 48.9% of the SSDI recipients eventually return to work at some point during their spells on SSDI. However, nearly all of the DI beneficiaries who return to work do so only on a part-time basis and for a relatively short duration. The average number of years worked while receiving SSDI benefits is about 2.9 years. The mean earnings of those who return to work is $9,096, significantly higher than the SGA for this cohort ($6,000 annually for a $500 per month SGA). Our model of health dynamics predicts that 75% of the individuals who enter the DI roles will eventually experience some partial recovery during their spells, while 50% will fully recover. Thus, under the $1 for $2 offset, nearly all of the fully recovered beneficiaries have sufficient incentive to return to work, whereas only 18% of them have sufficient incentive to do so under the *status quo*.

Another reason that the predicted labor supply response is relatively large under the $1 for $2 offset is that we explicitly assume that SSA is able to make a credible commitment not to increase the audit rates—known internally at SSA as *continuing disability reviews* (CDR's)—for DI recipients who return to work. Under the *status quo* simulations individuals of believe that engaging in the TWP would put them at substantially greater risk of being terminated due to a CDR. This is why only 10% of DI recipients take advantage of the TWP. If we assume that individuals continue to have these beliefs under the $1 for $2 offset, then the fraction of DI recipients who ultimately return to work falls from 48.9% to 36.8%.

Independent evidence that a $1 for $2 benefit offset provides a strong work incentive is provided in a study by Muller (1992) that followed 59,000 SSDI beneficiaries who were first entitled to benefits between age 55 and 64. A total of 6,518 of them, or 11% of the sample, returned to work at some point after their initial entitlement. Of these, 71% returned to work after age 62 and 47% returned to work after age 65. This is of significant importance, because SSDI beneficiaries can convert from DI to OA benefits at age 62, and the OA program has a $1 for $2 offset above a substantially higher disregard level.[3] At age 65 the earnings test falls to a $1 for $3 offset above the higher disregard level. Thus, it is not surprising to find that the majority of SSDI beneficiaries who return to work do so at ages where the earnings test drops significantly. The life-cycle model's predictions of the timing of return to work by DI beneficiaries are quite consistent with these findings. In our *status quo* simulations 9.5% of DI beneficiaries eventually return to work, all of whom first return to work after age 62, while 44% do some work even after age 65. The response is much larger and at earlier ages under $1 for $2 offset rule, because individuals who have recovered can take advantage of the tax incentives as soon as they are able to work, rather than having to wait until age 62.

The life-cycle model also predicts that the $1 for $2 offset will not have a very significant induced entry effect. In the simulations conducted here, the number of SSDI applications increases by only 2.2%, while SSDI roles increase by 3.2%. The percentage increase in the roles is larger than the percentage increase in applications because the $1 for $2 offset provides a greater incentive for SSDI recipients who have fully or partially recovered to stay on the program and supplement their benefits via part-time work. However, the mean duration of a beneficiary on the program increases only slightly, from 12.7 to 13.0 years.

Thus, the induced entry effect is primarily responsible for the 5.9% increase in the total number of person-years spent on SSDI. Due to the offsetting effects of benefit reductions of SSDI beneficiaries who return to work, the present value of benefit payments (discounted to age 21 at a 2% interest rate) increases by only 1.7%, from $115,000 per beneficiary to $117,000 per beneficiary. However, since there are more DI beneficiaries due to induced entry, the total discounted value of SSDI benefit payments is predicted to increase by 4.9%. The present value of Social Security contributions increases by 4.2% under the $1 for $2 offset, but even after subtracting the discounted value of these increased contributions from the present value of benefits, the net discounted cost of the SSDI program still increases by 5%.

The life-cycle model's prediction of the induced entry effect lies between the projections of

[3] For example, in 1989 the SGA was $300 per month, but the OA earnings test disregard for early retirees was $540 per month, and $740 per month for normal retirees.

the SSA and CBO. A 1994 study by James McGlaughlin, of the SSA Office of Actuary, predicted that the $1 for $2 program would result in 40,000 additional *induced filers* being awarded DI benefits each year, over a 10 year period. The number of new DI benefits per year awarded to disabled workers (we exclude their dependents since they would not be affected by the $1 for $2 offset) has ranged from a low of 587,000 workers in 1997 to a high of 662,000 in 1992. Using the annual average between 1994 and 2001, McGlaughlin's predicted increase of 40,000 induced awardees amounts to an approximate 6.4% increase in the annual number of workers awarded DI benefits. This is nearly three times larger than the 2.2% increase predicted by our life-cycle model. Projecting the latest data on DI roles from the 2001 Annual Statistical Report on the Social Security Disability Insurance Program, the 5.3 million adult beneficiaries will grow to 5.9 million by the end of 2004, if the 4% average growth rate in DI roles over the period 1994 to 2001 persists. Thus, McGlaughlin's study implies an induced entry effect of approximately 6.8% (i.e., the 400,000 induced filers divided by the projected 5.9 adult worker DI beneficiaries at the end of 2004).[4]

The CBO predicts that there would be 75,000 induced filers over the 10 year period considered by McGlaughlin. Dividing by the projected DI population of 5.9 million workers in 2004, the CBO predicts that induced entry will increase SSDI by approximately 1.3%. The life-cycle model predicts that DI roles will increase by 3.2%, which is in between the SSA and CBO predictions, but somewhat closer to the CBO prediction.

The life-cycle model predicts that the $1 for $2 offset will increase the cost of the SSDI program (net of contributions). However, it also provides a clear benefit to a subset of SSDI recipients, allowing them to achieve higher income, consumption, and wealth accumulation during, and following, their spells on SSDI. In particular, recipients' annual consumption increases by an average of 2.2% over their full lifetimes, and by 6.9% between the ages of 45 and 65. Nevertheless, due to the relatively high marginal tax rate of 50% and the increasing disutility of effort at older ages, the *ex ante* increase in welfare for a younger person who has not yet experienced a disabling condition is small. This is the main reason as to why the model predicts that the induced

[4] It is not clear whether McGlaughlin's estimates take account of mortality and conversions of DI benefits to OA benefits at the normal retirement age. Our projections do account for these sources of attrition from the DI program, as well as continuing disability reviews and voluntary exits due to the TWP. Also, McGlaughlin study assumed that the monthly disregard at which the $1 for $2 offset would take effect was $85, substantially lower than the prevailing SGA level at that point in time of $500. The McGlaughlin report assumed that the amount of induced entry would not change much if the disregard was increased to the full SGA level. The life-cycle model predicts that there would be a slightly smaller increase due to induced entry under the lower $85 disregard. DI roles increase by 2.1% under the $85 disregard compared to 3.2% under the $500 disregard.

entry effect is small. The main welfare gains of the $1 for $2 occur *ex post* for people who have already entered SSDI and who have experienced a full or partial recovery.

The remainder of the paper is organized as follows. Section 2 briefly describes the $1 for $2 offset proposal and explains why it is hard to predict the effects of this policy change using alternative methods, such as randomized experiments with human subjects. Section 3 provides a brief summary of the life-cycle model and shows that the model is capable of approximating the behavior of real individuals by comparing the predictions of the model to actual behavior observed in aggregate and micro panel datasets. Section 4 provides detailed analysis of the predictions provided by the life-cycle model regarding the impact of the $1 for $2 offset proposal. Section 5 offers some concluding remarks and discussion.

2 Evaluating the Effects of the $1 for $2 Offset

The essence of the $1 for $2 offset policy is to encourage SSDI beneficiaries to return to work on a full- or part-time basis, by reducing the effective tax rate paid by the SSDI beneficiaries who decide to return to work. Under the current policy, a SSDI beneficiary who desires to return to work can take advantage of a 9 month *trial work period* (and typically a 3 month *grace period* thereafter). During that period the person can earn any amount without any offsetting reduction in benefits. However, if a person continues to work beyond the trial work/grace period, their SSDI benefits will be terminated if they earn more than a relatively small earnings threshold amount known as the *substantial gainful activity* ceiling (SGA, $800 per month in 2003). Effectively, under the current policy an SSDI beneficiary is subject to a 100% *benefit tax* on earnings in excess of the SGA. Although there is no reduction in benefits or threat of being removed from the SSDI program if one earns less than the SGA limit, the discontinuous change in treatment of earnings in excess of the SGA can be viewed as a punitive tax that creates strong work disincentives.

The $1 for $2 offset proposal would allow any SSDI beneficiary who returns to work and earns above the SGA level beyond the 12 month trial work/grace period to remain on the roles. However, their DI benefits would be reduced by $1 for every $2 earned in excess of the SGA threshold (hence the name “$1 for $2 offset”). Thus, the policy change can be viewed as a reduction in the current 100% tax rate on earnings in excess of the SGA, to a 50% tax rate. The intention of the $1 for $2 offset proposal was to provide a tax incentive that would encourage

SSDI beneficiaries to leave the roles and return to work. However, since the $1 for $2 offset policy makes the DI program more generous than the *status quo*, there is a concern that it could actually result in a net increase in the cost of the DI program as a result of *induced entry*, i.e., an increase in the propensity to apply for benefits to take advantage of the increased benefits.

The U.S. Congress was sufficiently concerned about the possibility of induced entry that it mandated that the SSA undertake studies to predict the net effect of the $1 for $2 offset proposal. The 1999 *Ticket to Work Act and Work Incentives Improvement Act* (TWWIIA) authorized the SSA to carry out *demonstration projects*, that is, a large scale controlled experiments with human subjects, in order to predict the impact of the $1 for $2 offset.

Demonstration projects are regarded by many as the most credible way for predicting the impacts of policy changes. However, they also suffer from a number of disadvantages. First, they are extremely costly and time-consuming. Second, it is not clear that a demonstration project would be able to accurately predict the induced-entry effect. Most analysts anticipate the increase in the propensity to apply for SSDI benefits resulting from the $1 for $2 offset to be small. In which case, recent analyses by the Office of Research, Evaluation, and Statistics (ORES) at the SSA suggest that a huge demonstration project—possibly in excess of one million people in the treatment and control groups—would be necessary to obtain a statistically reliable estimate of the induced entry effect.

Furthermore, even a large scale experiment may not accurately estimate the ultimate impact of the $1 for $2 offset if it were adopted nationally. This is because it is difficult to control for informational and *social interaction effects* in an experimental setting. An individual's decision about whether or not to apply for SSDI is affected by his/her understanding of the program rules and procedures, and this information might come from a variety of sources other than the SSA. To control for these effects a panel of advisors to the SSA have considered the feasibility of a *county design* whereby all individuals in entire counties are assigned to either the treatment or the control groups. In this case, any information about the $1 for $2 could be made uniformly available in the counties that are randomly selected to receive the treatment. Eventually, the information about the new program should percolate through the various channels in the treatment counties. Hence, the opportunities that individuals in these counties have to obtain information about the new program will closely approximate the opportunities they had to obtain information about the existing SSDI rules. But even then, there are likely to be unobserved state and county level factors that also affect application decisions that will not be controlled for in a county level design. Because of these problems, the panel concluded that there are serious uncertainties about whether an expensive large scale demonstration project would necessarily be able to provide accurate predictions of the induced entry effect.

A recent paper by Tuma (2001) discussed a number of serious problems with various strategies for implementing a demonstration project for measuring the induced entry effect.[5] Although the report noted that: “Some of the consultants are uneasy with the necessary sample sizes that ORES calculated”, it also concluded that: “There is no disagreement, however, that the sampled numbers of counties would need to be large, that their populations would be very large, and that the cost of a demonstration study of induced entry could be huge” (p. 26).

These huge costs would be applicable for a demonstration project with a *single* design, i.e., where the treatment group is given a $1 for $2 with a fixed disregard. However, in general, it would also be of interest to study the incentive effects of other effective tax rates on benefits, such as a $1 for $3 or a $1 for $4 benefit offsets, or other *optimal* tax rates. As noted above the SSA made separate predictions for two different values of the disregard and there was considerable internal debate about whether a low or high disregard should be implemented. However, if it is not cost-effective to evaluate the induced entry effects of the $1 for $2 offset, then it is certainly hopeless to use the demonstration project methodology to systematically evaluate alternative policies, in order to isolate the most promising one.

The limitations of randomized experiments have long been recognized by economists and sociologists. Moffitt’s (2003) survey of this literature concluded that: “While randomized field trials in the area of welfare reform have been professionally conducted and well-run, and have yielded much valuable and credible information, their usefulness has been limited by a number of weaknesses, some of which are inherent in the method and some of which result from constraints imposed by the political process. The conclusion is that randomized field trials have an important but limited role to play in future welfare reform evaluations, and that it is essential that they be supplemented by non-experimental research.” (p.??). A particular limitation of experimental methods noted in Moffitt’s survey is their inability to estimate entry effects. He concludes that randomized experiments: “should be reserved for estimating the exit effects and effects on initial

[5] Prepared for a committee within the SSA, charged with designing a demonstration project to evaluate the $1 for $2 offset policy.

participant populations," and those experiments: "should be supplemented by non-experimental analyses of entry effects where it appears possible that those effects are significant." (p. 38).

In view of these problems, Tumas report recommended that a number of alternative, more cost-effective, strategies for predicting the impact of the $1 for $2 be considered, including "2) a dynamic model of individual behavior using results of previous studies, 3) individuals' responses to a survey with hypothetical questions" (p. 24). In the next two sections we demonstrate the feasibility of approach 2) using a particular type of dynamic model, namely the *life-cycle model*. The next section describes the model and, via a comparison of the predictions of the model and the observed behavior of real individuals from a variety of micro and macro data sources. We show that the life-cycle is capable of providing a realistic and accurate empirical model that could be of substantial value for forecasting the impacts of wide range of policy, changes including the $1 for $2 offset proposal.

3 The Life-cycle Model

The life-cycle model is one of the cornerstones of economic theory and is originally credited to the work of Modigliani, Brumberg, Ando and others. Generally, there is no single life-cycle model, but rather a class of models that could be described as life-cycle models, where specific models differ in the details about labor supply, consumption, savings, uncertainty, and details about the private and social insurance institutions. There have been some economists, such as Bernheim, Skinner and Weinberg (2001), who argued that the life-cycle model cannot account for observed levels of *under-saving*, and consequently low wealth accumulation by a significant fraction of Americans. We argue that this conclusion is erroneous since it is based on an oversimplified formulation of the life-cycle model, which can be solved analytically. Current versions of life-cycle model a lot more realistic and account for many more aspects of the individuals decision process, such as labor supply decisions, incomplete markets, Social Security, pensions, etc.[6] Although these models are typically too complex to be solved analytically, the advent of fast computers and improved algorithms now allows us to solve increasingly realistic versions of the life-cycle model numerically. Via computer simulations of these models, it becomes clearer

[6] An example of the latter type of model is Rust and Phelan (1997).

that the life-cycle model is sufficiently rich to be able to provide insightful explanations for a wide variety of previously puzzling aspects of savings, labor supply, pension, and social security application decisions.

A prime example is the *age 65 retirement puzzle*. Previous oversimplified life-cycle models were unable to explain the peaks in retirements, particularly at ages 62 and 65. Obviously, these peaks must have some connection to the fact that early and normal Social Security retirement benefits are available at these ages and that Medicare benefits are available at age 65. However, previous reduced-form models, and life-cycle models that failed to accurately model the Social Security rules, were unable to explain the peaks in retirements at these ages. Consequently, some economists conjectured that the only way one can explain the concentration of retirements at age 65 is via a *sociological age 65 retirement effect*. In contrast, Rust and Phelans (1997) showed that the peaks in retirements can be explained as a rational response to retirement incentives, using a version of the life-cycle model that accounted for incomplete markets and incorporated a more realistic treatment of the incentives created by the SSA.

The peak at age 62 is largely due to a significant number of low income, liquidity constrained, individuals who would have liked to retire earlier than 62 but are unable to borrow against their future Social Security benefits. The peak at age 65 is largely due to health insurance constraints, that is, individuals who have health insurance policies through their employer but who would be unable to purchase fairly priced retiree health insurance if they were to quit working before 65. Rust-Phelan model showed that these individuals have a strong incentive to continue working until age 65 when they become eligible for Medicare benefits.

We developed a version of the life-cycle model that is specifically focused on providing a realistic treatment of the U.S. Social Security program. We follow the general approach of Rust and Phelan (1997) who developed an econometric algorithm that repeatedly re-solves the life-cycle model until the values of the unknown parameters estimate enable the predictions of the life-cycle model to best fit the actual behavior of individuals that are observed in micro panel datasets.[7]

The parameters of the life-cycle model include parameters that determine individuals' preferences for consumption and leisure, and parameters that characterize their beliefs about their

[7] Rust-Phelan used the Retirement History Survey, while we use the Health and Retirement Survey, which began in 1991 and is continuing to collect data from an initial panel of 12,000 individuals in two year intervals.

uncertain future health, mortality, and earnings. Other parameters can be imposed from the outside, if one is willing to assume that individuals are rational and fully informed. These include the parameters determining the eligibility and benefits under the Social Security program, such as: (1) the ages of early and normal retirement; (2) the bend points in the function relating the average indexed monthly earnings (AIME) to the primary insurance amount (PIA); and (3) the actuarial reduction factors for payment of Social Security benefits at the early retirement age, and so forth.

Nevertheless, the Rust-Phelan model has several important limitations. First, it was estimated for the 1903-1911 birth cohort using the Retirement History Survey that was collected in the 1970s, and is thus out of date. Second, our model relaxes an important restriction of Rust-Phelan model that consumption equals income, which was a reasonable approximation for the predominantly lower income, blue collar workers in their RHS sample. However, their model would be of questionable value for evaluating more radical Social Security reform proposals such as *individual accounts*, for which modeling and understanding individual savings and wealth accumulation decisions is likely to be critical. Third, Rust and Phelan ignored the SSDI program, which is one of the most volatile components of the Social Security programs and DI roles and costs have been rising at unsustainable rates. A more comprehensive model that includes all the key components of the Social Security program is necessary to obtain accurate predictions of the net fiscal impacts of various policy changes. For example, an attempt to save money by increasing the early retirement age from 62 to 64 or 65 may be partially offset by an increased in applications for DI benefits at ages lower than the early retirement age.

We now describe a life-cycle model that overcomes many of the limitations of the Rust-Phelan model. The unknown parameters of this successor model will be estimated using the most recent available panel data from the six waves of the Health and Retirement Survey (HRS) over the period 1992 to present. Second, the model includes the consumption/savings decision in addition to the labor/leisure decision. The model includes an integrated treatment of the SSDI and the Old Age and Survivor's Insurance programs of Social Security (OASDI).[8]

[8] Because of computational difficulties the present version of the life-cycle model does not yet include Medicare/Medicaid, private health insurance, or the Unemployment Insurance components of the Social Security program. These will be incorporated in a future version of the life-cycle model that we present below, along with a more realistic treatment of the family that includes income from spouses and other sources of unearned income, asset/business income and other sources of unearned income. We believe that these additions will allow the life-cycle model to do a better job of fitting observed behavior, particularly with respect to wealth accumulation and labor supply.

We describe the life-cycle model briefly and intuitively and illustrate some comparisons of the behavior predicted by a calibrated version of this model with actual behavior observed in the HRS and other micro data, as well as data from aggregate program statistics from the SSA and other agencies. We demonstrate the richness of the life-cycle model and the additional insights that this model provides. While the results reveal a number of areas where changes need to be made to enable the model to do a better job of fitting the data, it is already sufficiently developed to be used for analyzing important policy issues, particularly the $1 for $2 offset proposal.

3.1 The Model

The life-cycle model predicts an individual's behavior over their full life-cycle starting at age 20 until their death. Each period, assumed to be one year in length, the individual makes decisions about how much to consume and how much to save, whether or not to work, and if so, whether to work full- or part-time, and whether or not to file an application for disability benefits, and, if the person is over age 62, whether to apply for Old Age benefits. An individual conditions his/her decisions on current information, which includes their current age, wealth, and health status. The individual faces uncertainty about future health status, mortality, and earnings. The individual saves in order to accumulate a precautionary buffer stock in the event of protracted periods of low earnings and/or bad health, as well as in order to prepare for retirement.

The model has three health states which can be classified as: (1) excellent/good health; (2) fair/poor health; and (3) "disabled". We put quotation marks around the latter state since it does not coincide with the Social Security definition of disability, but rather denotes a state of poor health that, we assume, is sufficiently severe that it prevents the individual from working entirely. The transition probabilities among the various health states were estimated from data on self-reported health and disability status in the HRS survey. Health states (a) and (a) are relatively persistent, in the sense that once one is either good health or one has a disability, they tend to continue to remain in good health, or continue to have their disability, for a long time. Specifically, if a person is currently in good health, there is a 95% probability that he/she will be in good health in the following year. Similarly, if a person is disabled, there is a 87.5% probability that he/she will remain disabled in the following year. The poor health state represents a *transitional state*, that is, when someone is in poor health there is a 20% chance that they will be in good health in the following year and a 12% chance that they will become disabled in the following year. Initially we assume that these transition probabilities do not depend on age. However, we do assume that the probability of dying depends on age, as well as on the person's health status. Using the HRS data we have estimated age and health-dependent survival probabilities. Not surprisingly, the results show that survival probabilities decline with age and decline with health.

Figure 1 compares the aggregate survival probability of a simulated sample of 1,123 individuals to the survival probabilities from the U.S. Decennial Life Tables for 1997. We see that the survival curve for our simulated population is very close to the survival curve implied by the life tables. This clearly indicates that we have accurately estimated the age-health survival probabilities from the HRS survey. Moreover, it also implies that our assumption that health status transition probabilities do not depend on age may be a reasonably good approximation, at least in the current context.

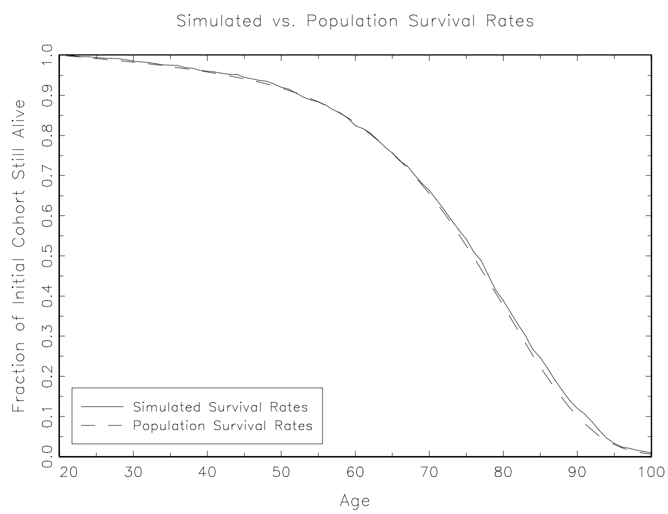

Figure 1: Simulated versus Actual Survival Curves

Individuals are assumed to maximize the expected discounted stream of future utility, where the per period utility function $u(c,l,h,t)$ depends on consumption c, leisure l, health status h, and age t. Obviously, we specify a utility function for which more consumption is better than less, and more leisure is better than less. The flip side of the utility of leisure is the disutility of work. We assume that the utility of leisure (disutility of work) is an increasing function of age and is higher for individuals who are in worse health than for individuals who are in good health. Thus, the main factor that distinguishes a person who falls into the health state *disabled* from a person who is in *fair/poor* health is that the former has a higher disutility of work. In addition, we assume that the worse a person's health is, the lower their overall level of utility is, holding age, leisure, and consumption constant. We also allow for a bequest motive by providing a utility of wealth bequeathed to heirs or to institutions (e.g. the college alma mater) after one dies.

Any person who is not already receiving SSDI benefits is eligible to apply for SSDI benefits provided they are younger than the *normal retirement age* (currently 65). If they are over the normal retirement age, then the only option they have is to apply for Social Security Old Age benefits. Prior to the normal retirement age any person has the option of applying for either SSDI or Old Age benefits, provided that if he/she is over the *early retirement age* (currently 62).).[9] While there is a 100% probability of being awarded OA benefits if one applies and is age-eligible, the probability of being awarded SSDI benefits is considerably less than 100%. Using data from the HRS and aggregate program data from the SSA web site, we estimated a probabilistic model of the SSDI award process (for details see Benítez-Silva, Buchinsky, Chan, Cheidvasser, and Rust (1999)).

The probabilities of being awarded benefits depend on the individual's health status and labor supply decision in the year of application. It is higher the worse one's health status is, but it is zero if they choose to work at a level in which earnings exceed the SGA level discussed in section 2. A person is allowed to appeal a denial, and also to repeatedly apply (and/or appeal) for SSDI benefits. This *repeated game* aspect of the SSDI program means that the *ultimate award rate* is about 70%, much higher than the 50% *initial award rate* that would be inferred by looking only at the initial decisions by the disability determination services (DDS). The reason that the ultimate award rate is higher than the initial award rate is because a significant fraction of applicants who are denied benefits by the DDS choose to appeal. The first level of appeal is to ask the DDS to perform a reconsideration, and if denied again, they can appeal the case to an Administrative Law Judge (ALJ). In principle an individual can subsequently appeal to the Social Security Appeals Board, and then to the Federal Court, but only a small fraction of awards are due to successful appeals to these last two stages. Our life-cycle model does not model each of these separate appeal stages. Instead, we simply assume that a denied applicant can reapply an unlimited number of times. At each reapplication, an applicant has the same chance of being awarded benefits as their initial application. As indicated above, in reality, the chance of being awarded benefits via an appeal to an ALJ is significantly higher than the chance of receiving an initial award via the DDS. On the other hand, the SSA is likely to keep track of

[9] The OA benefits are actuarially reduced based on the number of years prior to the normal retirement age at which the individuals first start receiving the OA benefits.

previous applications to the program, and a person who repeatedly applies for benefits may have lower chances of success. These two considerations have opposing effects, and we feel it is a reasonable approximation to model appeals as new applications.

A person that has been awarded SSDI benefits will not necessarily remain on the program until he/she dies. First, if an person reaches the normal retirement age he/she will be automatically transferred from the SSDI program to the OA program. Individuals can also decide to return to work on a full- or part-time basis, and consequently will be terminated from the SSDI roles after the 9 month trial work period, provided that their earnings exceed the SGA level. There is also a small probability that an individual will be involuntarily terminated as a result from random audits that are known internally in SSA as *continuing disability reviews* (CDRs). We allow the probability of being terminated due to a CDR to be a function of a person's health status, with persons in good health status having a substantially greater risk of termination than those who are in poor health, or who are disabled. Furthermore, as noted above, under the *status quo* version of the model we assume that DI recipients believe that engaging in a TWP will put them at permanently higher risk of termination due to a CDR. When calibrated the probability of termination due to a CDR 3 times higher after engaging in a TWP, the life-cycle model was able to provide a reasonable explanation as to why only about 10% of DI recipients ever take advantage of the TWP. Without these altered beliefs the life-cycle model over-predicted the fraction of DI recipients who would take advantage of the TWP option.

We solve the life-cycle model by backward induction, starting from the terminal age of 100 and working backward until age 21, when we assume individuals enter the labor force. Agents in our model make three decisions at the start of each period, denoted by $\{l_t, c_t, ssd_t\}$, where l_t denotes *leisure*, c_t denotes *consumption*, which is treated as a continuous decision variable, and ssd_t denotes the individuals *Social Security decision*. Here, l_t denotes the amount of waking time devoted to non-work activities, normalized to 1. Thus we define, $l_t = 1$ to denote not working at all, $l_t = .543$ corresponds to full-time work, while $l_t = .817$ corresponds to part-time work. These latter quantities correspond to the amount of waking time spent in leisure, assuming that a full-time job requires 2000 hours per year and a part-time job requires 800 hours per year.[10]

The quantity ssd_t assumes three possible values where $ssd_t = 1$ denotes the decision to apply

[10] This is how we get the leisure values $l = .543 = (12 * 365 - 2000)/(12 * 365)$ and $l = .817 = (12 * 365 - 800)/(12 * 365)$ corresponding to full and part-time-work respectively.

for Old Age benefits, $ssd_t = 2$ denotes the decision to apply for DI benefits, and $ssd_t = 0$ denotes the decision not to apply for benefits. Some of these choices may be infeasible under certain circumstances. For example, individuals who are below the early retirement age (denoted by ERA, currently set at 62) are not allowed to receive OA benefits. Hence, their choice set reduces to $ssd_t \in \{0,2\}$. Also, if a person is already receiving OA benefits they cannot re-apply for additional benefits, so they face no further choices unless their age t satisfies $\text{ERA} \leq t < \text{NRA}$ (where NRA denotes the normal retirement age, currently at 65), in which case they still have the option to apply for DI benefits, even while receiving OA benefits.

The *state* of an individual at any point in time can be summarized by four variables: (i) Current age t; (ii) net (tangible) wealth w_t; (ii) the individual's Social Security state ss_t; and (iv) the individual's average wage, aw_t. The ss_t variable can assume up to ten mutually exclusive values: $ss_t = 0$ (not entitled to benefits), $ss_t = 62$ (entitled to OA benefits at the early retirement age), and $ss_t = 63,\ldots,70$ represent the remaining 8 Social Security states corresponding to first becoming entitled for benefits at each of the ages $63,\ldots,70$, respectively. The reason these states are required is that under the SSA benefit formula, the individual's monthly old age benefit is based on their *primary insurance amount* (PIA) and a permanent actuarial adjustment factor that depends on the age at which the person was first entitled to OA benefits. If the age of first entitlement is before the NRA there is a permanent actuarial reduction. If it occurs after the NRA there is a permanent increase in benefits due to the delayed retirement credit (DRC).

In addition to age, wealth, health, Social security status, and current income, the average indexed wage is a key variable in the DP model, serving two roles: (1) it acts as a measure of *permanent income* that serves as a convenient *sufficient statistic* for capturing serial correlation and predicting the evolution of annual wage earnings; and (2) it is key to accurately modeling the rules governing payment of the Social Security benefits. An individual's highest 35 years of earnings are averaged and the resulting *Average Indexed Earnings* (AIE) is denoted as aw_t.[11] The PIA is the potential Social Security benefit rate for retiring at the normal retirement age (NRA). It is a piece-wise linear, concave function of aw_t, whose value is denoted by $pia(aw_t)$.

In principle one need to keep as state variables the entire past earnings history. To avoid this, we approximate the evolution of average wages in a Markovian fashion, i.e., period $t+1$

[11] If there is less than 35 years of earnings when the person first becomes eligible for SSDI, then the 5 lowest years of earnings are dropped and the remaining wages are averaged.

average wage, aw_{t+1}, is predicted using only age, t, current average wage, aw_t, and current period earnings, y_t. Specifically, we use the observed sequence of average wages as regressors to estimate the following log-normal regression model of an individual's annual earnings:

$$\log(y_t) = \alpha_1 + \alpha_2 \log(aw_t) + \alpha_3 t + \alpha_4 t^2 + y_t + \eta_t. \tag{1}$$

This equation describes the evolution of earnings for full-time employment. Part-time workers are assumed to earn a pro-rata share of the full-time earnings level (i.e., part-time earnings are $800/2000$ of the full-time wage level given in equation (1)).

The SSA's AIE is an indexed average of the 35 highest earnings years in a worker's earnings history. We have found that the AIME is well approximated by a simple moving average of indexed earnings (truncated at the Social Security maximum earnings limit), taken over the entire earnings history (i.e., we have not dropped out the lowest 5 years of earnings). This moving average wage, aw_t, can be written recursively as

$$aw_{t+1} = \frac{t}{t+1} aw_t + \frac{1}{t} y_t. \tag{2}$$

If we regress aw_t on the exact AIME (calculated from the person's earnings history using SSA's *ANYPIA* program), the R^2 of the regression is 98%, which confirms that aw_t is an accurate predictor of AIME. The advantage of using aw_t instead of AIME is that aw_t becomes a sufficient statistic for the person's earnings history. Thus we need only keep track of aw_t, and update it recursively using the latest earnings according to (2), rather than having to keep track of the entire earnings history in order to determine the 35 highest earnings years, which the AIME requires.

The DP model also accounts for actuarial reductions in old age benefits claimed prior to the NRA, and for the delayed retirement credit (DRC) for benefits claimed after the NRA. For the 1931-1941 cohort the NRA is 65 and the PIA is permanently reduced by an actuarial reduction factor of $\exp(-g_1 k)$, where k is the number of years prior to the NRA but after the ERA that the individual first starts receiving OA benefits. The actuarial reduction rate for the 1931 to 1941 cohort is $g_1 = .0713$, which results in a reduced benefit of 80% of the PIA for an individual who first starts receiving OA benefits at age 62. Note that a person who is accepted into the DI program prior to the NRA receives the full PIA regardless of his/her age. However, the SSA does apply an actuarial reduction to the DI benefits that are awarded after the ERA.

To increase the incentives to delay retirement, the 1983 Social Security reforms gradually increased the NRA from 65 to 67 and increased the DRC. This is a permanent increase in the PIA by a factor of $\exp\{g_2 l\}$, where l denotes the number of years after the NRA that the individual delays receiving OA benefits. The rate g_2 is being gradually increased over time. The relevant value for the 1931 to 1941 cohort is $g_2 = 0.05$. The maximum value of k is MRA$-$NRA, where MRA denotes a "maximum retirement age" (currently 70), beyond which further delays in retirement yield no further increases in PIA. As noted above, it is not optimal to delay applying for OA benefits beyond the MRA, because due to mortality further delays generally reduce the present value of OA benefits the person will collect over their remaining lifetime.

A final aspect of the Social Security rules that we model concerns taxation of benefits. We are solving the life-cycle model for a cohort of individuals who were born around 1930, and thus we have implemented a version of the Social Security benefit formula that was in effect during the mid 1990s, when these individuals started to reach the NRA (65 for this cohort). Individuals whose combined income (including Social Security benefits) exceeds a given threshold must pay Federal income taxes on a portion of their Social Security benefits. We incorporate these rules in our model as well as the 15.75% Social Security payroll tax.

In addition to these taxes, we account for the Social Security *earnings test.* If a person retires between the ERA and NRA, each dollar of earnings above a certain threshold (approximately $10,800 in the mid 1990s) results in a 50 cent reduction in Social Security benefits. Between the NRA and MRA the implicit earnings test tax rate falls to 33% for earnings above a higher threshold ($17,000 in the mid 1990s). For individuals who are above the MRA, there is no earnings test. The earnings test provision has been recently eliminated for individuals who are over 65. However, since the earnings test was relevant for most of the HRS birth cohort at the time of their retirement, we include it in our model. Our model also incorporates a detailed model of taxation of other income, including the progressive Federal income tax schedule (including the negative tax known as the EITC – Earned Income Tax Credit), and state and local income, sales and property taxes.

We assume that the individual's utility is given by

$$u_t(c, l, ssd, h, age) = \frac{c^\gamma - 1}{\gamma} + \phi(age, h, aw) \log(l) - 2h - K \tag{3}$$

if $ssd = 2$. Otherwise, the individuals utility is given by

$$u_t(c, l, ssd, h, age) = \frac{c^\gamma - 1}{\gamma} + \phi(age, h, aw) \log(l) - 2h, \tag{4}$$

where h denotes the health status and $\phi(age, h, aw)$ is a weight that can be interpreted as the *relative disutility of work*. We assume that ϕ is an increasing function of age and health status (that is, individuals in worse health have a higher disutility of work, as discussed above). We also assume that ϕ is a decreasing function of *aw*, reflecting the fact that individuals with higher wages typically have more interesting and physically less demanding jobs, and thus a lower disutility of work than a "blue collar" worker who typically earns lower wages.[12] The parameter K represents the "hassle" and "stigma" costs involved in applying for DI benefits. One can allow K to be a function of observed covariates (such as age and average wage), as well as unobserved heterogeneity, but we abstract from this specification in this paper. In the results shown below we have used a value of $K = .001$. We assume that there are no time or financial costs involved in applying for OA benefits, but we do explicitly account for the time and "hassle" costs involved in applying for DI. Accounting for these costs is essential, otherwise the life-cycle model substantially over-predicts the number of people who apply for DI benefits.

The parameter γ indexes the individual's level of risk aversion. As $\gamma \to 0$ the utility of consumption approaches $\log(c)$. We use $\gamma = -.37$, which corresponds to a moderate degree of risk aversion, i.e., implied behavior that is slightly more risk averse than that implied by logarithmic preferences.

Figure 2 plots the function ϕ that we used in the solution and simulations of the life-cycle model. The left panel shows that the disutility of work increases with age, and is uniformly higher the worse one's health is. If an individual is in good health, the disutility of work increases much more gradually with age compared to the poor health, or disabled health, states. The right hand panel of figure 2 shows how the disutility of work decreases with average wage. This is included to be a parsimonious way of allowing the life-cycle model to capture differences in the working conditions faced by high and low wage workers. We postulate that high wage workers, especially highly educated professionals, have better working conditions than most lower wage blue collar workers, whose jobs are more likely to involve less pleasant, more repetitive working conditions and a higher level of physical labor.

[12] In the subsequent econometric analysis we will allow the disutility to contain parameters reflecting unobserved heterogeneity for leisure, and let the data determine the distribution of the disutility of work conditional on the average wage and other observable variables.

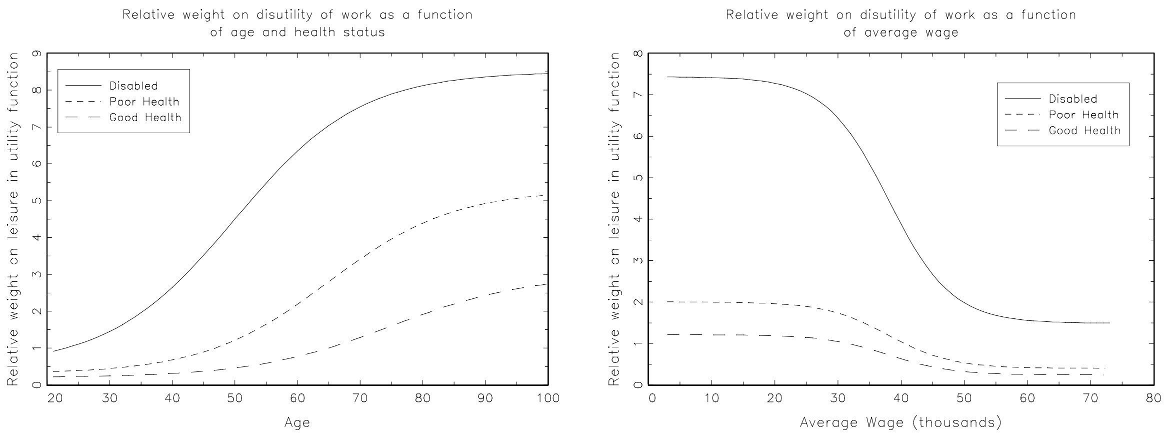

Figure 2: Relative Weight on Leisure as a Function of Age, Health and Average Wage

3.2 Simulations of the Life-cycle Model

Figures 3 through 10 illustrate the rich types of behavior that the DP model predicts. Each of the curves is an average of 1,123 independent and identically distributed *iid* simulations, with each simulation corresponding to a separate individual followed from age 21 until their death. The averages were computed at each age, for the sub-population of survivors who lived until at least that age.

Figure 3 compares actual and simulated health status by age. The simulated health status in the right panel of Figure 3 does a reasonable job of matching the actual pattern in the left panel of the figure. The fraction of people reporting good health status is declining with age a little more rapidly than in our model, and the fraction in poor health and disabled is increasing more rapidly with age than in our model. This suggests that our initial assumption of age-invariant health transition probabilities is one that we should relax, and this should enable us do a better job matching the "age-health profiles" in Figure 3.

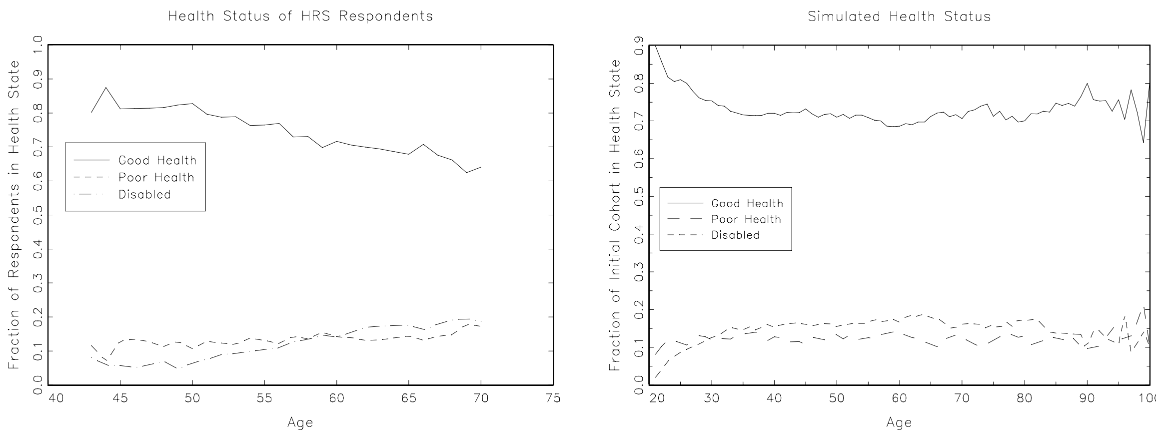

Figure 3: Actual vs. Simulated Health Status

The left hand panel of Figure 4 shows the employment status from the HRS data as a function of age. Note that there is a clear decline in labor force participation starting at about the age of 54. There is also significant increase in part-time work after the age of 60. The simulation results shown in the right hand panel of Figure 4 exhibits a similar, but more exaggerated pattern. The DP model under-predicts the amount of part-time work between ages 45 and 60, and, generally, over-predicts the amount of part-time work at later ages. The pronounced peaks in part-time work at ages 65 and 70 (when the earnings test tax falls from 50% to 33% and 0%, respectively) are absent in the HRS data. It appears that the life-cycle optimizers are far more responsive to these incentives than real people are, a point which should be kept in mind when evaluating the predicted impact of the $1 for $2 offset. Another discrepancy is that the life-cycle model over-predicts the fraction of full-time workers between ages 45 and 60, and under-predicts this fraction at later ages. None of the 1,123 individuals in the simulated data worked full-time after age 65, whereas approximately 20% of the HRS sample in this age ranged continued to work full-time.

We believe that many of these discrepancies can be reconciled in future versions of the life-cycle model. Specifically, adding more heterogeneity with regard to how rapidly the disutility of work increases with age, will make it possible to better predict the number of part-time workers at younger ages as well as the number of full-time workers at later ages. In the current version all individuals share the same utility function and the only source of heterogeneity is differences in health, average wages, and wealth. In fact, it is quite remarkable that we are able to capture the basic qualitative features in the data with a model with such limited heterogeneity. In reality, many of the individuals who continue working full-time into their 70s may be high paid professionals such as doctors, lawyers or academics, who have much better working conditions and more job flexibility, and who are more likely to love their work. As noted in the discussion of Figure 2, we have tried to capture some of this effect in a parsimonious way by allowing the disutility of work to be a declining function of average wage.

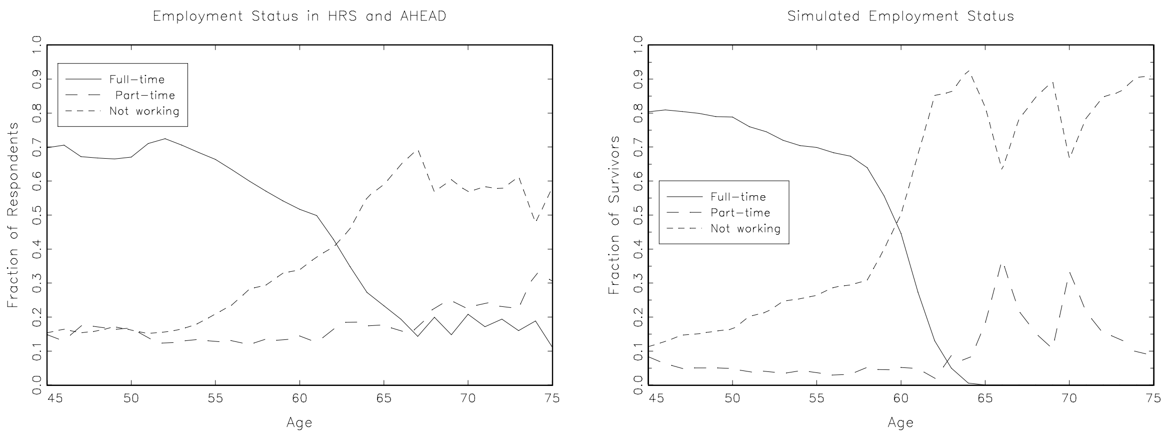

Figure 4: Actual vs. Simulated Labor Force Participation

Figures 5 and 6 compare the distribution of ages at which people first receive OA benefits. In the left panel of Figure 5 we present the actual distribution of retirement ages in 1998, from the 1999 Annual Statistical Supplement to the *Social Security Bulletin.* In the right panel we depict the results of our simulation of the life cycle model. The model captures the main features of the data, particularly the large peak in retirements at age 62. However, the current version of the model does not capture the relatively small fraction of individuals who claim benefits at age 63 and 64, and at ages after the normal retirement age. Again, the main reason for this apparent discrepancy is because of lack of heterogeneity in the life-cycle model, other than that which is produced by randomly evolving incomes, average wages, and health.

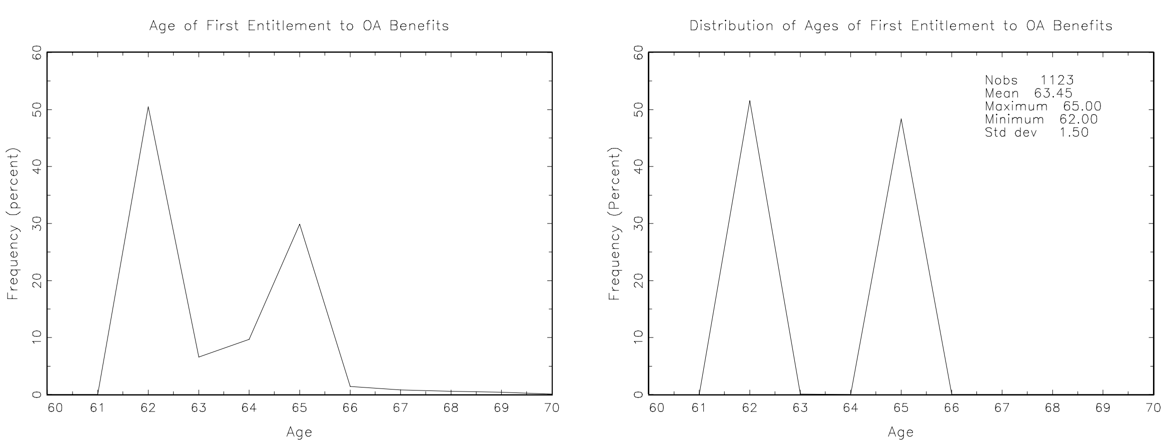

Figure 5: Actual vs. Simulated Distributions of Ages of First Receipt of OA Benefits

Figure 6 shows the evolution of the population between the 3 states of: (1) not receiving any

Social Security benefits; (2) receiving SSDI benefits; and (3) receiving OA benefits. We see that the model captures these trends fairly well, except that in the model, entry into SSDI is more concentrated in the 45 to 60 age range than it is in the HRS data.

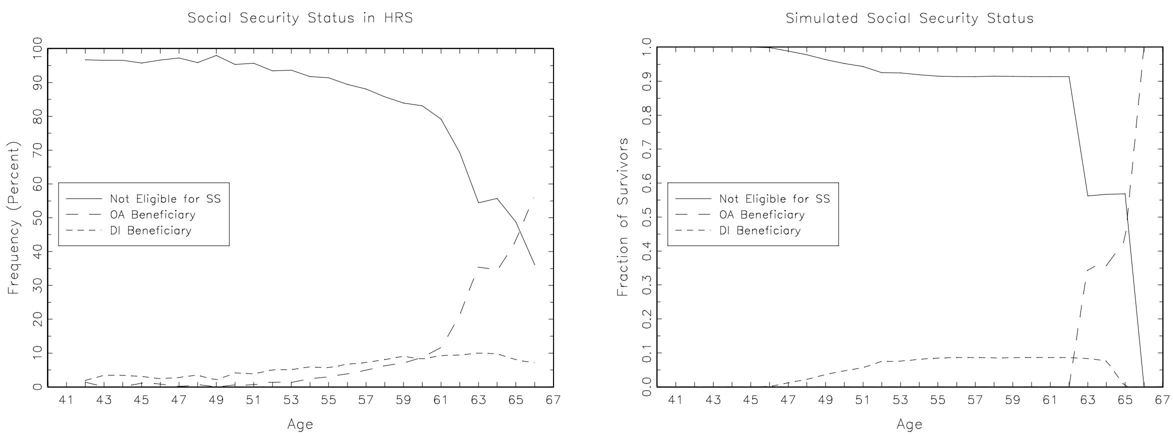

Figure 6: Actual vs. Simulated Social Security Status

Figure 7 shows actual and simulated trajectories for wages, and wealth. In the right panel of the figure we also provide the simulated trajectories of Social Security benefits and consumption over the life-cycle. First, we see that wages increased over the first part of the individuals' life-cycle and start dropping in their late 50's in both panels of the figure. During the first 30 years, individuals consumed only about 70% of their wage earnings, resulting in a rapid buildup of net worth that peaks at age 60 in our simulations, and slightly later in the actual data. The maximum level of wealth accumulation is about the same in the data and the simulations, but the life-cycle model predicts a more peaked trajectory for wealth. Wealth accumulates up faster than we observe in the HRS prior to age 60, and then decumulates at a faster rate than we observe after at 60.

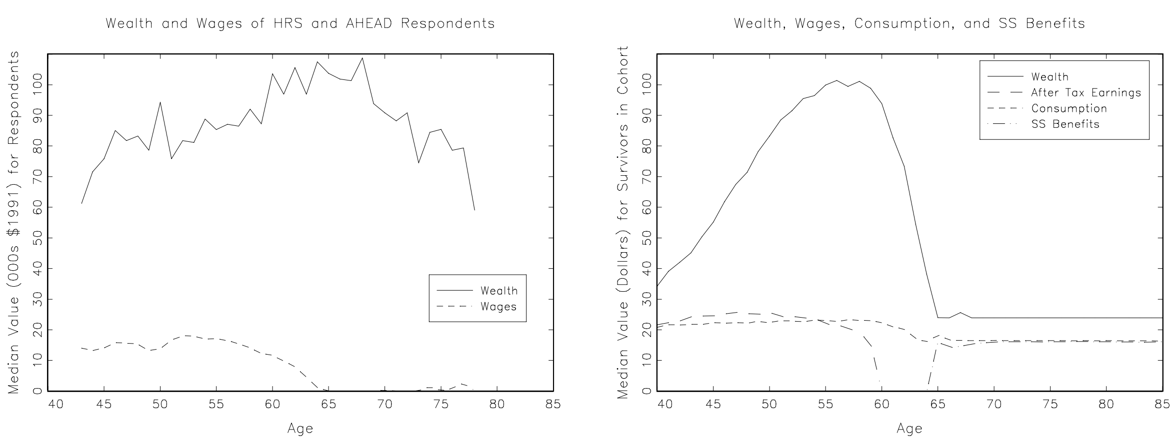

Figure 7: Wealth, Earnings, Social Security Benefits, and Consumption

The actual distribution of wealth is more skewed in the HRS data than in our simulations (not reported directly for brevity). In particular, the mean net worth in at age 60 for the HRS respondents is more than two times greater than median wealth at that age, whereas in our simulations mean wealth is only slightly higher than median wealth. It is conceivable that the reason that the life-cycle model does not generate a sufficient skewed wealth distribution is that we have not yet incorporated other sources of income, such as spousal income and inheritances. Additional heterogeneity in earnings processes could also help generate extra skewness in the wealth distribution. We believe that once we account for other risks such as the risk of involuntary unemployment and uninsured medical costs, the life-cycle model will predict substantially higher precautionary savings rates than we observe in the current model. Recall that in the current model the only risks an individual faces is loss of job due to health problems, mortality, and uncertainty about future wage rates.

Overall, it appears that the simulation results provide a reasonable first approximation to the data. Contrary to claims that individuals have inadequate savings at the eve of retirement that Bernheim and others have made, our simulations of the life-cycle model suggest that the individuals born between 1931 and 1941, that were followed by the HRS, have adequately prepared for retirement. If anything, these individuals have a higher level of wealth accumulation both before and after the retirement age than what is predicted to be optimal by our life-cycle model. Our conclusions are similar to those obtained by Engen, Gale and Uccello (1999), in a study that also analyzed simulations of a numerically solved life-cycle model.[13] Thus, we see little evidence that individuals in the HRS are under-saving for retirement, a conclusion supported by direct empirical analyses of wealth and pension accumulations of HRS respondents by Gustman and Steinmeier (1999). The life-cycle model shows how wealth accumulation plays an important role in consumption smoothing over the life cycle. In particular, the rapid decumulation of wealth after retirement allows individuals to maintain a relatively smooth pattern of consumption over their life cycle, even when major changes in labor supply occur.

Figure 8 depict the net wage (net of taxes) Social Security benefits and consumption over the life cycle. Clearly, the life-cycle model predicts that as individuals enter retirement, they make

[13] Engen *et. al.* life-cycle model did not allow a choice of labor supply and did not incorporate a realistic treatment of the Social Security. Allowing for Social Security and endogenous labor supply, *ceteris paribus*, leads to lower levels of wealth accumulation. However, our model also incorporates an additional risk that Engen *et. al.* did not consider—uncertain health status—that creates a motive for higher precautionary wealth accumulation. These two effects seem to counterbalance each other. Consequently, both our model and the Engen *et. al.* model lead to roughly similar levels of wealth accumulation over the life-cycle.

significant reductions in their consumption spending. The right hand panel of Figure 8 shows that mean consumption spending declines from a peak of $22,000 per year at age 60 to about $19,000 per year at age 65. There is a sharp peak in consumption at age 70 that reflects the significant number of individuals who return to work to take advantage of the fact that there is no earnings test after age 70. The large drop in consumption is caused primarily by the large drop in wealth, which is being consumed as a result of early retirement. As we can see in the left hand panel of Figure 8 individuals are starting to retire in their 50s, whereas Social Security benefits don't begin to kick in until individuals are eligible to receive them at ages 62 and 65.

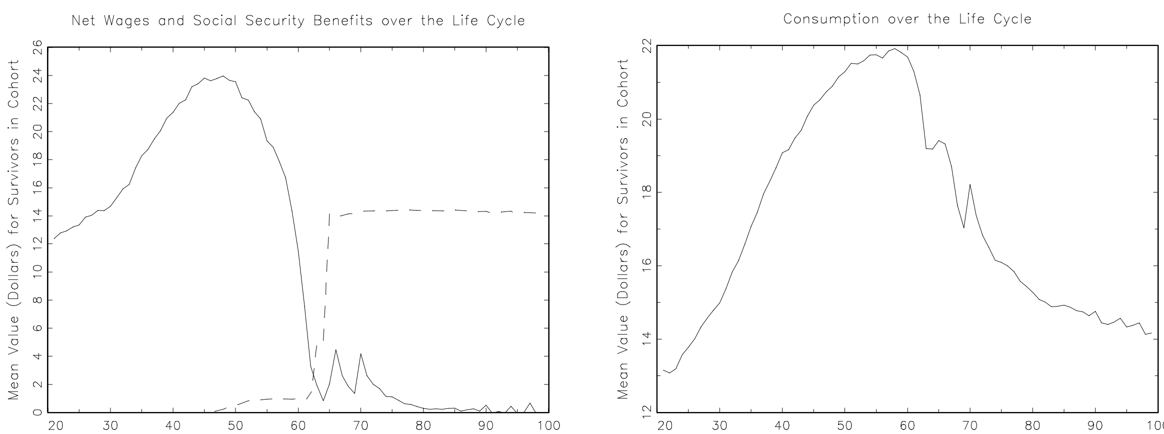

Figure 8: Detail on Net Wages, Social Security Benefits and Consumption

A number of previous authors such as Bernheim, Skinner and Weinberg (2001) have argued that large drops in consumption after retirement are inconsistent with the predictions of the life-cycle model. Figure 8 demonstrates that this claim is incorrect. We believe that it is premature to conclude that the life-cycle model is incapable of explaining observed behavior. On the contrary, sufficiently rich and realistic versions of the life-cycle model do provide extremely good approximations to observed behavior.

Figure 9 compares the distribution of ages of first receipt of DI benefits. The right hand side panel presents the simulation results, while in the left panel we provide the distribution from the HRS. Our simulations are qualitatively similar to the actual distribution, except that the life-cycle model under-predicts the mean age of first receipt of DI benefits. We have calibrated the life-cycle model to approximate the average age of first receipt of SSDI benefits in the aggregate, which is 49.3 for 1997 from Table 6.C in the 1999 Annual Statistical Supplement to the *Social Security Bulletin.* Our model is also consistent with the fraction of DI recipients who are ultimately awarded benefits, namely 70%, that we have found in previous work with the HRS data (see

Benítez-Silva, Buchinsky and Rust, 2003a and 2003b). Furthermore, we find that nearly 9.5% of SSDI recipients ultimately return to work, all via the TWP, a result that is consistent with Muller's (1992) analysis of the New Beneficiary Survey. In our simulation, approximately 10% of all SSDI recipients leave the roles as a result of continuing disability reviews (CDRs). This is also consistent with Muller's analysis of the New Beneficiary Survey. Finally, the mean duration on SSDI for our simulation sample, 12.5 years, is somewhat higher than the 10.9 year duration for the SSDI recipients in 1993 (Wheeler, 1996).

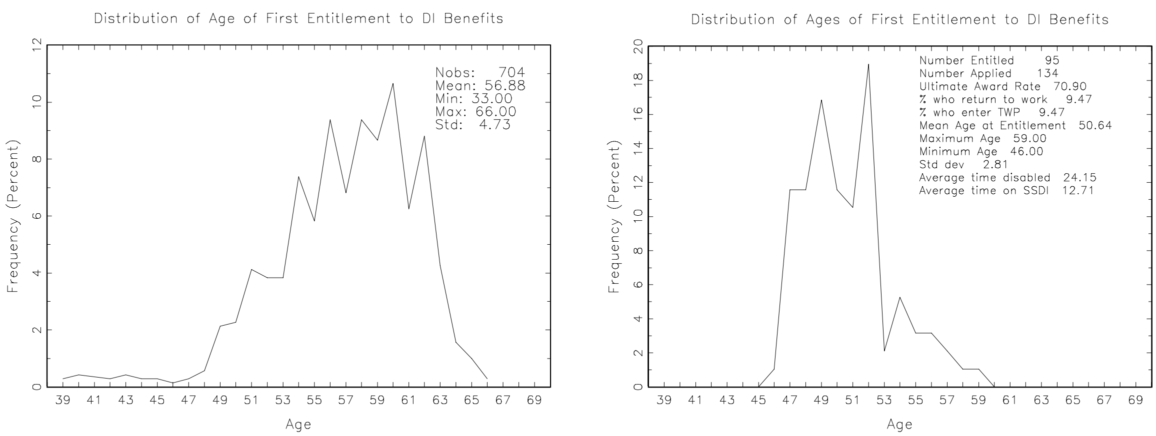

Figure 9: Distribution of Age of First Receipt of DI Benefits

There are two key aspects of individuals' preferences and beliefs that affect their decisions to apply for SSDI, and to return to work under the TWP. The first is the perceived stigma of being on SSDI. The second is an individual's beliefs about the chances that they will be a subject of a CDR if they reveal themselves as no longer disabled by taking advantage of the TWP. Without any stigma effect the life-cycle model greatly over-predicts the number of individuals who apply to SSDI. In our calibrations we have found that we can approximate the fraction of the population who applies for SSDI by including a fairly small stigma effect, $K = .001$, in the utility function in (3). The stigma effect also helps generate an incentive for individuals to leave the DI roles, since without it, we find that very few DI recipients have an incentive to permanently exit the program, even when they have completely recovered from their disability.

The life-cycle model also over-predicts the number of DI recipients who use the TWP opportunistically, that is, those who return to the DI roles immediately after the TWP. As noted in the introduction, our model of health dynamics predicts that about 50% of new awardees to the DI program will eventually experience a recovery sometime during their spell on the DI program. The life-cycle model predicts that nearly all of these individuals should take advantage of the return to work incentives provided by the fact that they can keep 100% of their DI benefits during the 9 month TWP and the subsequent 3 month grace period. However, as noted above, the work of Muller (1992, 2000) indicates that only about 11% of new DI awardees eventually take advantage of the TWP. One way the life-cycle model to capture this phenomenon is via a belief on the part of recipients that engaging in TWP will *reveal* to the SSA that they are no longer disabled, and hence will put them at much greater risk of leaving the roles via a CDR.

The results provided in Table 1 of Muller (1992) suggest that these beliefs may be well-justified. Among the 405 DI beneficiaries in the New Beneficiary Survey who returned to work at some point in their DI spells, 46 of them were removed from the roles due to: "Report of work led to continuing disability review and termination for medical recovery" (p. 5). This appears to be a significantly higher termination due to CDRs than for the DI population as a whole. We therefore assume that DI beneficiaries believe that their chances of being removed from the DI roles due to a CDR is permanently 3 times higher after engaging in a TWP, compared to not doing so. This enables us to match the fact that only about 10% of DI beneficiaries take advantage of the TWP at some point during their spells.

The main aspect of the simulation that does not seem to be consistent with the data is the fact that the actual distribution of ages of first receipt of SSDI tends to be more spread out over ages, but with an upward slope that peaks around age 60 and falls off rapidly thereafter. The distribution of ages from our simulations of the life-cycle model is more concentrated in the 40s and 50s age range and falls off to zero well before age 60.

We conclude with Figure 10, which shows two interesting implications of the life-cycle model. The left panel shows the distributions of bequests, while the right panel of Figure 10 plots the distribution of internal rates of return (IRR) on Social Security contributions implied by our model. The distribution of bequests is highly skewed, with a small number of relatively large bequests. However, none of the bequests in our simulation of 1,123 relatively lower income individuals exceeded the $600,000 exemption that would have subjected them to Federal estate taxes. Although there is no direct data on the size of bequests in the HRS, there has been some work using the first two waves of the AHEAD by Hurd and Smith (2001). Their work indicates that our simulation results are quite reasonable, and that the distribution of bequests is indeed skewed, but seem to somewhat understate the level of bequests. [14]

[14] We plan to update part of their work in order to be able to appropriately characterize this aspect of the model.

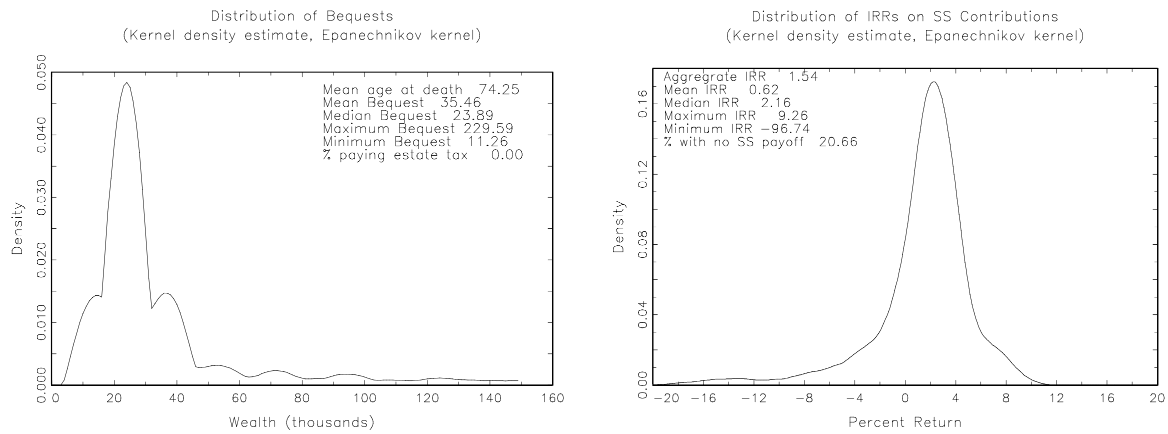

Figure 10: Implied Distributions of Bequests and IRRs on Social Security Contributions

As for the IRR, for each of the 1,123 people in our simulation, we computed the IRR on Social Security contributions. That is, we computed the interest rate which equates the discounted value of Social Security taxes (including the employer share) to the present value of Social Security benefits (including disability benefits). Consistent with other studies using actual Social Security benefits (e.g. Geanakoplos, Mitchell, and Zeldes 1999), we see that the average internal rate of return on Social Security is quite low, namely less than 2%. This low internal rate of return is the main cost of continuing to operate a fundamentally pay-as-you-go Social Security system in a rapidly aging society.

We believe the results presented here clearly show the richness and the insights that can be obtained from analyzing a sufficiently realistic version of the life-cycle model. While the comparisons of model predictions to the data in this section have revealed some discrepancies, the main features of the data have been captured quite accurately. The model analyzed here is under continual development and new features are added, but it is already sufficiently developed to provide new insights on important policy issues.

Regardless of the refinements that will be introduced to the model in the future, it to provide a perfect fit to the data. All models are approximations to reality, so the relevant issue is not whether a particular model is rejected by statistical goodness of fit tests (almost any model will be rejected by such tests), but instead: (1) whether the model is a sufficiently good approximation to be a useful and credible input into policy making; and (2) whether the model provides more accurate forecasts and predictions of behavior than competing statistical and behavioral policy forecasting

If necessary, we also plan to seek out other sources of data, including probate records, to check the predictions of our model.

models. We are not aware of any other model that is currently available that integrates SSDI and Social Security Old Age benefits and that can provide individual-level predictions of labor supply, savings, and DI and OA benefit application decisions, and which can also predict how behavior and welfare will change in response to changes in Social Security policies. Hopefully, in the near future there will be other competing models that can predict behavioral responses to policy changes.

4 The Impact of the $1 for $2 Offset

We now use the life-cycle model described in the previous section to predict the behavioral, welfare, and fiscal impacts of the imposition of the $1 for $2 benefit offset described in section 2. We want to emphasize that the predictions are based on the calibrated version of the life-cycle model that was presented in the previous section. Also, several components are still missing from this model including: (1) health care costs and Medicare/Medicaid and the availability of private health insurance; and (2) involuntary unemployment and the Unemployment Insurance (UI) program. Integrate of these components might alter the forecasts somewhat. For example, people who are uninsured may have a stronger incentive to enter the DI role in order to get access to Medicare benefits (after 2 years in the program). Hence, the induced entry effect might be larger once the *Medicare incentive* is incorporated.

The great advantage of the life-cycle model is it provides an opportunity to conduct a very special type of "controlled experiment" that would be impossible to conduct in the real world. The experiment works as follows. We simulate a population of 1,123 individuals starting at age 20 under the *status quo* Social Security policy with the de facto 100% tax on earnings above the SGA (up to the amount paid by the SSA). Then we re-solve and re-simulate the life-cycle model under the hypothesis that the $1 for $2 offset is in effect. In a computer simulation it is possible to save the "random seeds" that are used to generate stochastic trajectories for health, mortality, wages, and so forth. Some of these variables, particularly health and mortality, evolve according to *exogenous* stochastic processes, in the sense that the individual's decisions about the *endogenous variables* (labor supply, earnings, consumption and savings, and Social Security status) do not affect the evolution of the *exogenous variables,* which are health and mortality in this version of the life-cycle model.

Therefore we can save the random seeds for the simulation of the 1,123 individuals under the *status quo* and treat them as “experimental controls” to apply to another simulated population of 1,123 individuals who are given the $1 for $2 benefit offset “treatment”. This means that the trajectories for health and mortality in the control and treatment groups are *identical,* and thus the only differences in the outcomes for the two groups are changes in the endogenous variables, which reflect the behavioral responses to the $1 for $2 treatment. This is an especially powerful type of experimental control that is clearly not possible to conduct in any experiment with human subjects.

Figure 11 shows the labor supply effects of $1 for $2 offset. The left hand panel shows the impact on the fraction of the sub-sample of SSDI recipients who engage in full-time work and the right hand panel shows the fraction that engage in part-time work. We see that the $1 for $2 offset has very little effect on full-time work: it results in a slight reduction in the number of individuals engaging in full-time work between age 51 and 60. For example, at age 55 2.1% of the sample works full-time under the *status quo,* whereas only 1.0% work full-time under the $1 for $2 offset.

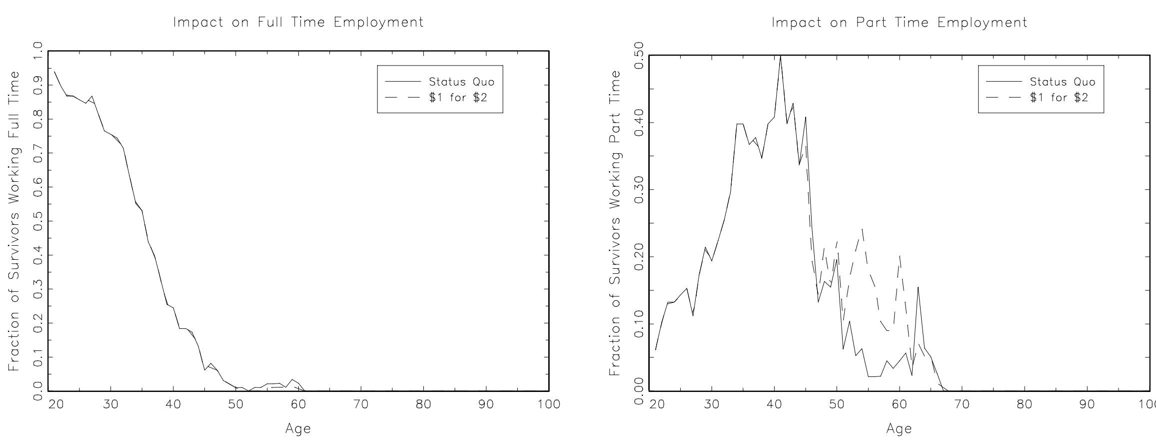

Figure 11: Impact on Full and Part Time Employment

The right hand panel of Figure 11 shows that there is a significant impact of the $1 for $2 offset on part-time work, concentrated between the ages of 50 and 61, when these individuals are receiving DI benefits, and are thus taking advantage of the $1 for $2 offset provision. There is virtually no change in part-time work earlier in the life-cycle, especially not before the age of 45, which is the age at which individuals start to enter the DI program. There is some part-time work at ages 45 and 46 under the $1 for $2, reflecting the fact that a number of individuals were induced to begin their application for DI earlier than they did under the *status quo.* From ages 47 to 50, the fraction of individuals engaging in part-time work is uniformly higher under the $1 for $2 offset. There are particularly large peaks in part-time work at ages 54 and 60, when 24% and 20% of the sample, respectively, are working part-time.

Under the *status quo*, we see a big peak of part-time work at age 63, where 14% of the sample are working part-time. The significant rise in return to work behavior after age 62 in the *status quo* simulations is reflect the opportunistic use of the TWP option. Why should such opportunistic use of the TWP increase after age 62 as opposed to earlier ages? This is explained by the fact that after age 62 DI recipients have a fall back option in the form of Social Security early retirement benefits, that helps shield them from the risk that a return to work could trigger a higher chance of termination due to a CDR.

Overall, the fraction of SSDI recipients who choose to return to work at some point during their DI spell increases from 9.5% under the *status quo* to 48.9% under the $1 for $2 offset. Furthermore, those who return to work, work for more years: under the $1 for $2 offset they work an average of 2.9 years, whereas under the *status quo* they work only 1 year. This is because under the *status quo* DI recipients are largely exploiting the trial work provision and do not work beyond it since it would result in termination from the DI program. In contrast, under the $1 for $2 offset DI recipients are significantly more likely to enter a TWP and to continue to work a few years after the end of the TWP.

Figure 12 shows the life-cycle model's prediction of the other major effect of the $1 for $2, namely the *induced entry* effect. The left hand panel depicts the fraction applying for DI benefits, while in the right panel the fraction on SSDI roles is depicted. We see only a modest induced entry effect here. Overall, 134 individuals applied for SSDI and 95 of these were ultimately awarded SSDI benefits under the *status quo* simulation whereas 137 individuals applied and 98 of them were ultimately awarded SSDI benefits under the $1 for $2. Evidently, the net gain in expected discounted utility resulting from the option to work while receiving DI benefits under the $1 for $2 offset is not large enough to induce people to apply for benefits. This is because also because of the hassle costs and the significant risk that they will be denied benefits. Note that under the $1 for $2 offset the ultimate award rate is 71.5% ($= 98/137$), only slightly higher than 70.9% ($= 95/134$), the ultimate award rate under the *status quo*. This shows that there is very small *induced persistence* effect, that is, applicants who are initially rejected have a slightly stronger incentive to appeal an initial rejection under the $1 for $2 offset than under the *status* *quo*.

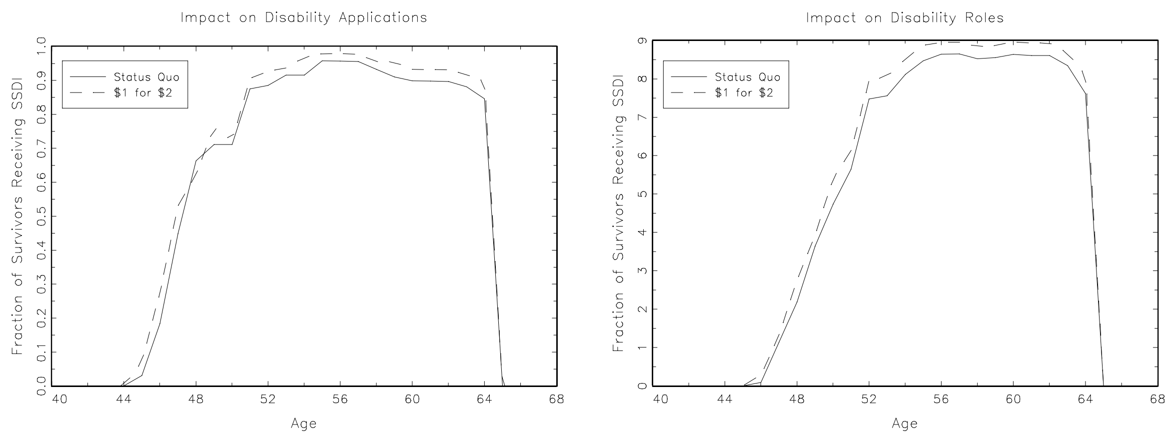

Figure 12: Impact on SSDI Applications and Roles

In the right hand panel of Figure 12 we see that SSDI roles increase by 3.1% in our simulations, which is somewhat larger than the 2.2% increase in applications for SSDI. The DI roles increase by a larger amount due to the fact that the mean duration on the DI program increases by .3 years, from 12.7 years under the *status quo* to 13.0 years under the $1 for $2 offset. It is tempting to blame this increase on *reduced exit* i.e., that DI recipients have less incentive to leave the roles under a $1 for $2 offset. However, the increase in durations is actually due to a .3 reduction in the mean age of first receipt of DI benefits. The reduction in age of application is closely related to the induced entry effect. At the margin some new disabled individuals who did not apply for DI benefits under the *status quo* are induced to apply under the $1 for $2 offset. In addition there are also individuals who applied in both cases, but who were induced to apply slightly *earlier* under the $1 for $2 offset. Consequently, the combined effects of induced entry and the increase in mean duration on the program result an increase of 5.9% in the number of *person-years* spent on SSDI roles.

Figure 13 shows the impact of the $1 for $2 offset on Social Security contributions and benefit payments for the sub-population of SSDI recipients. *A priori* the effect of the $1 for $2 on net benefit payments is unclear. Although recipients are on the roles for longer, they are receiving reduced benefits when they are working due to the $1 for $2 offset. Also, when they are working they are making Social Security contributions on their wage earnings. The left hand side panel of Figure 13 shows that despite the reduction in benefits due to increased part-time work under the $1 for $2 offset, DI benefit payments are slightly higher. In particular, between the ages of 50 and 60 DI benefits are 4.3% higher under the $1 for $2 offset compared to the *status quo*.

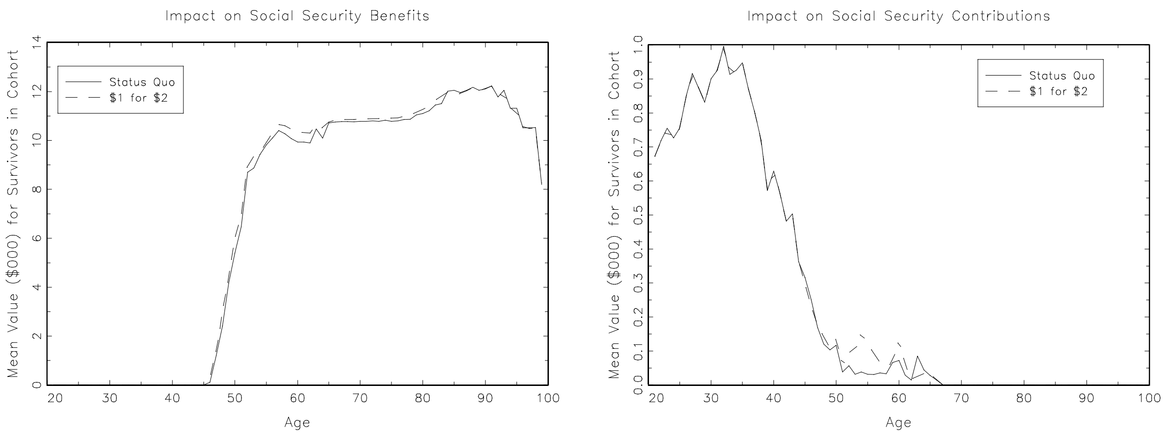

Figure 13: Impact on Social Security Contributions and Benefits

The right hand panel of Figure 13 shows the impact of the $1 for $2 on Social Security contributions. Not surprisingly, the higher level of part-time work between the ages of 50 and 60 results in higher Social Security contributions. However, contributions are lower at ages 62 and 63 since, as we noted above, there is a surge of opportunistic re-entry under the *status quo* at those ages. Under the $1 for $2 offset individuals tend to get higher earnings, which in turn, increases the average wages of DI beneficiaries. Consequently, this leads to a lasting increase in Social Security benefits at retirement. This is why even though the SSDI $1 for $2 offset is no longer relevant after age 65, since all DI beneficiaries have been converted to OA benefits, their Social Security benefits continue to exceed the benefits that they would have received under the *status quo*.

Figure 14 shows the impact of the $1 for $2 offset on the distributions of ages of first receipt of OA and DI benefits. The right hand panel of Figure 14 shows that under the $1 for $2 offset there is a slight downward shift in the ages of first receipt of DI, from 50.6 years under the *status quo* to 50.3 years under the $1 for $2 offset. The left hand panel shows that there is hardly any effect on the ages of first receipt of OA benefits. We do see a very small increase in claims for OA benefits at age 65 relative to age 62. This also contributes, via few actuarial reductions in benefits, to higher benefit payments after age 65 under the $1 for $2 offset.

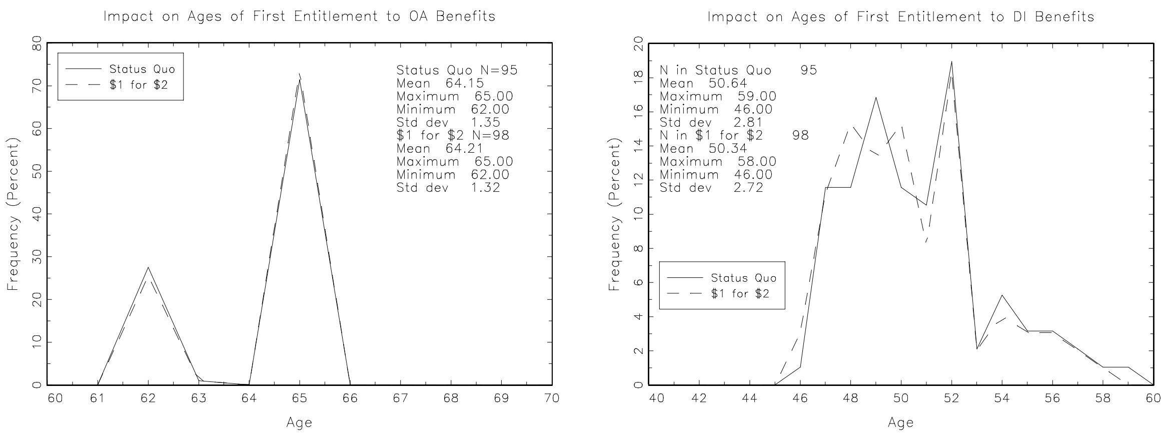

Figure 14: Impact on Age of Entitlement to OA and DI Benefits

In order to summarize the net impact of the $1 for $2 on the fiscal balance of the SSDI program, we compute the net present discounted value of SSDI benefits, where both contributions and benefits paid to the sub-population of SSDI recipients over their lifetimes is discounted at a real interest rate of 2%. Table 1 summarizes the fiscal implications of the $1 for $2 offset for the sub-population of SSDI recipients in the simulation cohort.

Table 1: Summary of the Budgetary Impacts of the $1 for $2 Offset

Item	Status Quo	$1 for $2 Offset	% Change
Number of DI applicants	134	137	+2.2%
Number of DI recipients	95	98	+3.2%
Person-years spent on DI	1273	1348	+5.9%
Number of DI recipients who returned to work	9	48	+533%
Mean Years of Part Time Work on DI	1	2.9	+290%
Mean Years of Full Time Work on DI	0	0	0%
Present value of Federal Tax Payments ($000)	$1938.92	$ 1872.75	-3.4%
Present value of Social Security contributions ($000)	$3,003.26	$3,129.84	+4.2%
Present value of Social Security benefits ($000)	$10,940.08	$11,474.07	+4.9%
Net Present value of Cost of SSDI beneficiaries ($000)	$ 5,997.90	$6,471.48	+7.9%
Present value of pre-tax wage earnings ($000)	$24,219.80	$25,240.65	+9.3%
Present value of consumption ($000)	$ 33,116.34	$ 34,684.56	+4.6%

We see that in discounted terms, the present value of benefits paid under the $1 for $2 offset increases by 4.9% from $10.9 million to $11.5 million. However, on a per beneficiary basis, the present value of benefits increases by only 1.6%, from $115,159 to $117,082. Thus, while each beneficiary is collecting slightly more benefits due to a longer duration in the program under the \$1 for \$2 offset, the majority of the increase in aggregate benefit payments is due to the induced entry effect.

Note that the present value of Federal tax payments is actually slightly *lower* under the \$1 for \$2 offset. This is due to the fact that the low part-time earnings of many of the DI recipients who return to work qualify them for tax rebates under the EITC program. If we calculate the net present value of the cost the DI program in our simulations (i.e., the present value of benefits less the present value of contributions and Federal taxes), we find a 7.9% increase in net costs under the \$1 for \$2 offset, from \$6.0 million to \$6.5 million. On a per beneficiary basis, discounted costs rise by 4.6%, from \$63,147 per beneficiary to \$66,041 per beneficiary.

Thus, these calculations suggest that implementation of the \$1 for \$2 offset policy would result in a modest increase in DI applications, awards, roles, and net expected discounted costs. However, the \$1 for \$2 offset provides some of the SSDI recipients with clear welfare improvement. The \$1 for \$2 policy is found to be strictly welfare enhancing, since it raises the payments (and thus consumption) to SSDI recipients under states of the world where they experience a recovery and wish to return to work. On the other hand, these increase payment do not reduce payments to the SSDI beneficiaries in other states of the world. Indeed, we see from Table 1 that the present value of pre-tax wage earnings is 9.3% higher, and the present value of consumption is 4.6% higher, under the \$1 for \$2 offset policy than under the *status quo*.

Figure 15 shows the impact of the \$1 for \$2 on consumption and wealth accumulation by age. The left hand panel shows the average consumption under the *status quo* and under the \$1 for \$2 offset. The mean consumption is higher under the \$1 for \$2 offset at every age. Similarly, the right hand panel of Figure 15 shows that net worth is also higher under the \$1 for \$2 offset at all ages. Nevertheless, there is virtually no increase in consumption and net worth prior to age 45, and the biggest increases occur between the ages of 55 and 70. It is evident that individuals who return to work under the \$1 for \$2 offset are using part of their earnings to generate an immediate increase in consumption, but they are also saving a significant fraction of these earnings to increase their buffer stock of savings for their retirement years. This additional buffer stock lasts well into their 80s.

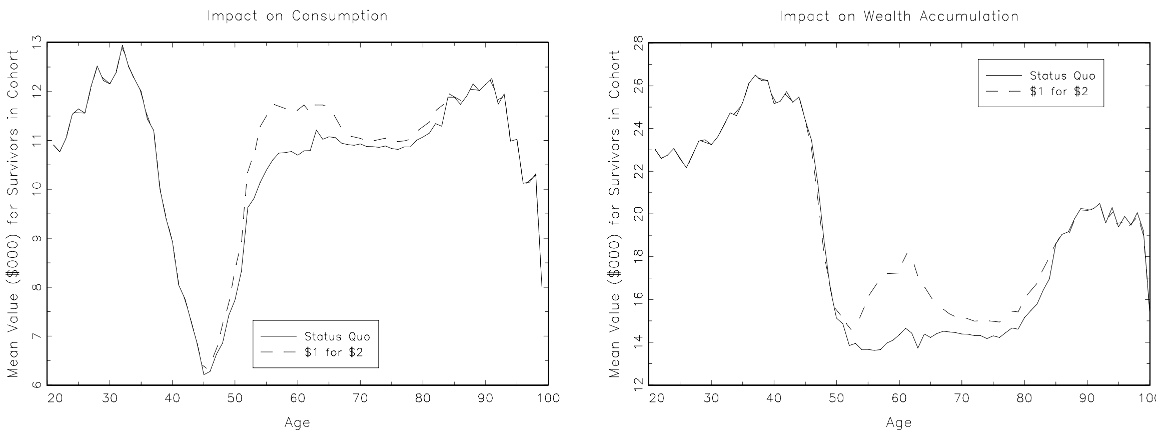

Figure 15: Impact on Consumption and Net Worth

Figure 16 plots the ratio of the mean consumption for DI beneficiaries under the $1 for $2 offset and mean consumption under the *status quo*. We see that there are particular ages at which there are fairly big increases in consumption, such as around age 45, where consumption is more than 18% higher under the $1 for $2 offset. Over the entire life-cycle, consumption increases by an average of 2.2% per year, and by 6.9% per year between ages 45 and 65.

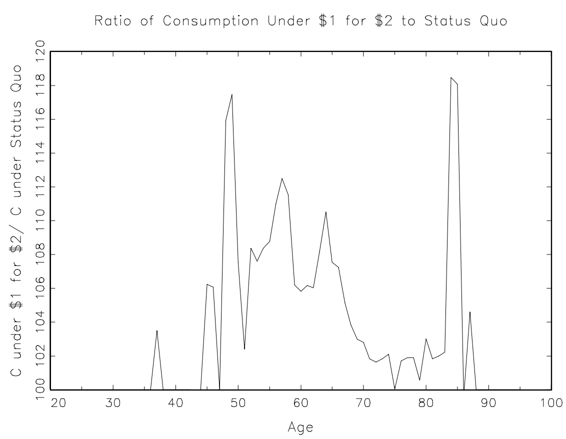

Figure 16: Ratio of Consumptions: $1 for $2 Offset vs. status quo

We conclude with Figure 17, which summarizes the welfare effects of the $1 for $2 offset policy. The $1 for $2 offset must be strictly welfare improving relative to the *status quo*, since no SSDI beneficiary can be made worse off as a result of the $1 for $2 offset, and can result in a strict increase in the well-being of those SSDI recipients who experience a recovery and wish to return to work. Under the $1 for $2 offset policy, these individuals have the option of continuing to receive reduced SSDI benefits even though they continue working past the trial work period. The simulations of the life cycle model show that this is an attractive option for nearly 50% of SSDI beneficiaries, who have experienced an improvement in their health status and returned to part-time work. These individuals are now receiving a combination of wage earnings and (reduced) SSDI benefits, rather than not working at all and relying exclusively on a relatively small SSDI benefits.

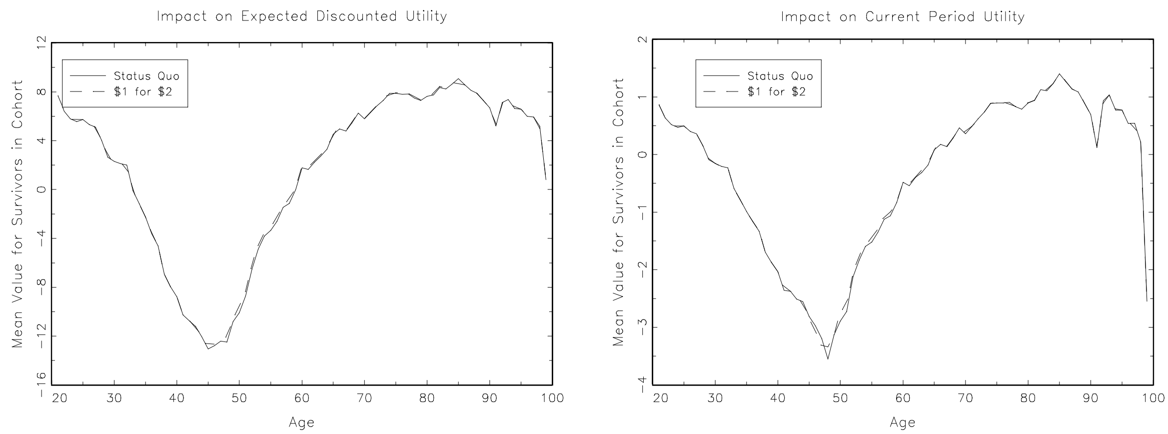

Figure 17: Impact on Expected Discounted and Current Period Utility

The left hand panel of Figure 17 plots the mean of the value function $V_t(aw, w, h, ss)$, i.e., the expected discounted utility from the current age onwards, as a function of age. The right hand panel plots the current period utility function $u(c, l, h, ss)$, as a function of age. We see that both V and u initially decline with age until about age 45 and then they increase until leveling off around age 70. (The large decreases in utility at the very end of the life-cycle is probably a statistical fluke due to small numbers of simulated observations at this age range.)

The decline in utility between ages 20 and 45 is a result of the fact that this population started off healthy when they were in their 20s, but their health declined in their 30s and 40s. Many actually became disabled in these age ranges, but were deterred from applying for DI benefits by the “hassle costs”. They stopped working and lived off of their savings, hoping to experience a recovery. However, as their disability persisted and their savings dwindled, most of these individuals decided to incur the hassle costs of applying for DI in their 40s. Once they got onto DI their welfare started to improve since their DI benefits enabled them to increase their consumption without relying entirely on consuming from their depleted savings. Also, about 50% of these individuals started to experience recoveries after several years on DI, and this lead to increasing welfare between ages 50 and 75.

In both panels of Figure 17 we observe that welfare is only slightly higher under the $1 for $2 offset than under the *status quo*. Indeed the increase in utility and welfare is barely visible in these graphs. There are several reasons for this finding. The first and most important point to note is that our welfare calculations account for the disutility of work effort. Thus, while consumption does rise significantly under the $1 for $2 offset, the rise in consumption came at a cost of forgone leisure by individuals who decided to return to work. While individuals were better off from having done so, in an *ex ante* sense, the disutility of work effort very nearly counterbalanced the increased utility from the higher consumption they could afford as a result of their work effort.

The second point to note is that although the $1 for $2 offset is strictly more generous than the *status quo*, it still amounts to a very high 50% surtax on earnings. Furthermore, this surtax applies to individuals with low average wage earnings anyway, who are not able to find very lucrative jobs, and who are older and face a higher disutility of work. Thus, the combination of the high tax rates, the higher disutility of work at older ages, and the relatively low earnings prospects for DI beneficiaries implies that the $1 for $2 offset is, on average, not a great deal for DI beneficiaries.

The final point to note is that the utilities in Figure 17 are an average of utilities of 50% of the population of DI beneficiaries who never experienced a recovery from disability and 50% who did. For those who never experienced any recovery, the $1 for $2 offset is not a highly valuable option. It is always possible that a disabled DI beneficiary might recover in the future and thus the $1 for $2 offset is an option that has some value. However, the value of this option is not very high for DI beneficiaries since the degree of persistence in the disabled state is quite high as noted in section 3.

Clearly, the $1 for $2 option is most valuable for DI beneficiaries who have recovered from their disabilities. However, when we work backward and compute the increase in welfare due to the $1 for $2 offset to a younger person who has not yet become disabled or considered applying to the DI program, the increase in *ex ante* welfare is negligible. The increase in welfare due to the $1 for $2 does become more significant once a person gets into their middle ages and becomes disabled. The welfare gain is highest for those who are already on DI and who have completely recovered from their disability. Obviously, the $1 for $2 offset has the highest value for this group.

The analysis of welfare changes is the key to understanding why the induced entry effect is not very large in our model. The *ex ante* gains facing a prospective SSDI applicant are not large when one factors in: (1) the chance that after incurring the hassle of submitting an application it would ultimately be rejected; (2) the chance that once on the program the person would actually experience a medical recovery that would make it possible to work; and (3) the high effective 50% surtax on the benefits for SSDI beneficiaries who do return to work. The combination of all these factors implies that the net gain from the $1 for $2 from an *ex ante* point of view is not large. Unless the “gradient” in the hassle costs for those who are at the margin of indifference between applying and not applying for SSDI benefits is very flat, the small increase in net expected discounted utility from the $1 for $2 offset will not result in a large amount of induced entry.

5 Conclusions and Discussion

This paper examines the impact of a new policy that has been considered recently by the SSA, namely the $1 for $2 benefit offset. We use a numerically solved life-cycle model and show that it can be an extremely useful tool for Social Security policy evaluation and forecasting, particularly for examining the proposed $1 for $2 benefit offset policy in the SSDI program. The SSA was initially mandated to evaluate the proposed policy change using a large scale demonstration project. The key issue that motivated the U.S. Congress to obligate the SSA to undertake a costly, time-consuming demonstration project, is the concern about *induced entry*. The key question that has been raised is: Would the $1 for $2 offset be considered a significant increase in the generosity of the SSDI program by potential DI applicants resulting in large increases in applications, awards, and program costs?

Unfortunately, a team of consultants advising the SSA on the design of this demonstration project concluded that it would be prohibitively costly and therefore infeasible to estimate the induced entry effect via a classical randomized experimental methodology. As a result, it appears that SSA has decided not to proceed with the demonstration project that was mandated under P.L. 106-170 sec. 302, or any other related data collection efforts, such as the National Survey on Health and Activity (NSHA). This is unfortunate, because both surveys and experiments are extremely valuable sources of data to improve our scientific understanding of how individuals react to, and are affected by, changes in social insurance policies.

The SSA’s reluctance to carry out an expensive, lengthy, large scale demonstration project is understandable given the inherent limitations of randomized experiments to evaluate the effects of policy changes. Even under the best of circumstances randomized experiments can only evaluate a very limited number of policy options. However, the U.S. Congress has mandated these experiments as virtually the only allowable methodology for evaluating the impacts of policy changes. In this paper we provide an alternative evaluation method for cases similar to the $1 for $2 offset, where it appears infeasible to evaluate all aspects of a policy change using a randomized experiment.

In the case of the $1 for $2 offset, the consultants advising the SSA recommended that alternative methods be pursued, including a dynamic model of individual behavior using results of previous studies and studies of “individuals’ responses to a survey with hypothetical questions” (Tuma, 2001, p. 24). Demonstration projects have also been mandated by the U.S. government to evaluate the impacts of other types of policy changes, particularly the impact of training programs and welfare reforms. However, leading experts have reached conclusions similar to those expressed in Tumas report, namely that “randomized field trials have an important but limited role to play in future welfare reform evaluations, and that it is essential that they be supplemented by non-experimental research.” (Moffitt, 2003, p. 38).

The empirical life-cycle model we have developed is a prime example for the non-experimental alternatives that have been advocated by Moffitt (2003) and Tuma (2001) as a necessary alternative to randomized experiments. Indeed, the approach used here may be one of a few alternatives at our disposal when it comes to the evaluation of program entry effects. The calibrated model we have used here to forecast the induced entry effects associated with the $1 for $2 offset deliver detailed and credible predictions about the implication of the proposed policy. In fact, we are not aware of any other alternative modeling approach that is capable of generating predictions that would be more detailed and accurate than the approach we used here.

Our forecast of the level of induced entry lies in between two independent forecasts made by the Congressional Budget Office and the SSA. Our analysis has provided a number of new insights into how entering the DI role and return to work incentives might be affected by the way the SSDI program is being administered. Specifically, the life-cycle model indicates that individuals are very sensitive to the SSA’s policy regarding CRDs. If DI recipients believe that taking advantage of the $1 for $2 offset will result in them being a target for higher rates of CDRs in the future, then a significantly lower number of DI recipients will return to work and, in turn, the induced entry effect becomes negligible.

It is worth noting that there are some who do not trust econometric models in general, and particularly the life-cycle model that we developed here. Some claim that reliability and credibility are questionable, while others simply see them as “black boxes”. Indeed, there are some papers in the literature, such as the paper by LaLonde (1986), which argue that econometric “selectivity bias” models were incapable of adequately predicting the behavioral responses observed in randomized experiments. This is probably the reason that Congress mandated the use of demonstration projects to evaluate the most significant policy changes. However, there are just as many papers in the literature, such as the one by Heckman, Hotz and Dabos (1987), that showed that Heckman-style econometric models were able of accurately forecast “treatment effects” of, say, job training programs. Of course, bad econometric modeling is not better than no modeling at all. And, clearly there is no substitute for human judgment and understanding by leading experts of the potential impacts of policy changes. However, when the government enacts new policies without significant prior analysis and testing, it effectively makes the entire population subject to an “uncontrolled policy experiment”, rather then what is referred to in the literature as “natural experiment”. There is a large literature in economics that evaluates the *ex post* impacts of these uncontrolled policy changes. But that leaves the economic science with little input about the *ex ante* design of efficient policies. This is not because economists have not given considerable attention, or have little to offer, but rather that the state of economic theory is far ahead of what is actually done in practice. Politics, both within academia and within government, is one of the many reasons why there is such a large gulf between the *science* of policy making and the much less informed and *ad hoc* way that policy decisions are made in practice.

While we make no claim that randomize experiments are not useful, we do argue that they are not sufficient, and in many cases are not feasible. When randomize experiments are not feasible, the most obvious alternative is the approach we have adopted here, not simply introducing new policies without any pretesting. This is particularly true when it comes to something as important as Social Security, since many people are crucially dependent on Social Security benefits. Frequent erratic policy changes can put them in a very precarious situation. While some policy changes are inevitable, it is better, in general, to have fewer well thought out policy changes than to have many semi-experimental or haphazardly chosen ones.

There are many examples of policies that have been implemented without any significant advance testing or analysis. One example is the “Ticket to Work Act” which was a separate part of P.L. 106-170 that mandated a demonstration project to evaluate the induced entry effects of the

\$1 for \$2 offset. Congress saw fit to implement the Ticket to Work with no advance experimental evaluation, whereas it was unwilling to enact the \$1 for \$2 offset until a thorough demonstration project had been completed. Under the Ticket to Work act, it is now possible for DI recipients to obtain free vocational counseling from thousands of different agencies located around the U.S. to help them return to work. An agency that provides the counseling is paid *prospectively,* that is, for every year a DI beneficiary stays off the roles due to the vocational training and rehabilitation services it had provided up front. The agency receives 40% of the DI benefit the person would have been paid if they had stayed on the roles. While the Ticket to Work Act seems like a highly innovative and creative idea, there are few regulations about the qualifications of agencies that provide the vocational counseling and rehabilitation. Clearly, there are vast opportunities for businesses in exploiting loopholes in the current policy. However, it is questionable as to whether the Ticket to Work Program will represent a cost-effective use of government funds, and whether it will help a significant number of DI beneficiaries return to work. Since the policy is already implemented, we have no alternative but to do an *ex post* evaluation of this uncontrolled policy experiment. It is quite possible that we will subsequently find that the Ticket to Work Program is not cost-effective, but it may be hard to reverse, since thousands of agencies around the country will have become beneficiaries of potentially lucrative government payouts.

In this paper we clearly demonstrate the usefulness of the life-cycle model in providing accurate predictions for the possible impact of the \$1 for \$2 benefit offset proposal that has been considered recently by the SSA. In this study we calibrate the life-cycle model to match observed data from the HRS and other aggregate statistics provided by various government agencies. The model performs very well, in the sense that its predictions are very close to the observed behavior along dimensions that are most important and most relevant when it comes to application for SSDI benefits. In particular, the model is able to accurately predict individuals observed data in the HRS regarding: (1) health status; (2) employment decisions; (3) age of first receipt of DI and OA benefits; (4) wealth; and (5) wages. Since these are the most important factors affecting the application decisions for DI benefits, and/or are the most obvious outcomes of such decision, it make makes one comfortable using it for the purpose of evaluating the proposed \$1 for \$2 benefit offset policy.

Careful examination of the various simulations provided in the paper indicates that the cost of the newly suggested policy is not huge. Moreover it provides some of the people with very strong incentive to return to work, even if in many cases it is only to part-time jobs. The induced entry effect that Congress and the SSA are very concerned about seems to be quite small. This is mainly because the proposed policy provides a welfare enhancement tool only for people who found themselves disabled at some point of their career. From an ex-ante point of view, the proposed plan does not give people enough incentives to apply for disability benefits. However, once a person finds himself on the DI roles, it give him/her ex-post incentives to get of the role. Consequently, the overall net increase in costs of the SSDI program are relatively small, and certainly relative to the overall spending of the SSDI.

While we argue that the life-cycle models are sufficiently realistic to provide credible predictions of a wide range of policy changes, we also argue that there is a very important complementarity between survey data collection, randomized experiments, and econometric modeling. We believe that the government should be investing in all three areas, and that these "R&D expenditures" will have a very high long term payoff in enabling the government to develop more cost-effective policies, particularly with respect to welfare and Social Security.

A very good illustration for this complementarity is provided in a recent study by Todd and Wolpin (2003). In their study, the authors used data from a large scale randomized experiment in Mexico (known as PROGRESA), that attempted to determine whether subsidies to parents who keep their children in school will significantly improve educational outcomes. They estimated a DP model, which has many similarities to the life-cycle model use here, that predicts a family's decision making about how much education their children will receive for a control group who did not receive any subsidies. They then re-solved the estimated DP model, accounting for the subsidies, and used it to predict the schooling attainments of the children in the treatment group families, i.e., those families who did receive the educational subsidies. The DP model provided surprisingly accurate fit, both for the control group that did not receive the subsidy and for the treatment group that did receive the subsidy. An important lesson to learn from this project is that it was possible for them to develop an accurate behavioral model, because the experiment was accompanied by the collection of a detailed panel data set on the families in both the treatment and control groups. Furthermore, the data collection extended well beyond the duration of the subsidies making it possible for researchers to assess the longer term impacts of the educational subsidies.

In view of the large and steadily growing share of GDP that is spent by the U.S. on transfer programs such as Social Security, it would seem that expenditures on “R&D” to develop better models and methods for validating this models would be money well spent.[15] Regardless of whether or not the $1 for $2 offset is an interesting policy reform on its own merits, we believe that there is a significant scientific payoff to running a similar demonstration project in the U.S. It might be argued that it is understandable that the SSA would not find it worthwhile to invest in a very costly large scale demonstration project to evaluate induced entry. However, it less clear why it does not see fit to invest in much less costly small scale demonstration projects that can provide a very valuable source of information that will assist it in developing better models (or in validating its own previous forecasts, e.g. the McGlaughlin forecast). The costs of data collection and model building are likely to be very small relative to the long term cost savings resulting from an improved understanding of how government policies affect individual incentives, decision making, and welfare, enabling us to design more efficient policies.

Although demonstration projects are an important means of confirming and validating behavioral models, it is not essential to wait for experimental data before undertaking the process of developing better models. Model building is the most inexpensive investment that can be made to improve the nation’s analytical capabilities. This paper has illustrated a case where it seems that a demonstration project mandated by law may in fact never actually be carried out. But it should at least be possible to provide detailed predictions of the impact of the policy via a credible behavioral model. We leave it to the reader to judge whether the behavioral model described here is credible or not. However, ultimately the real test is whether these models will be able to accurately forecast the behavioral responses to a wide range of policy changes, both those done under controlled settings (e.g. demonstration projects), and other policy changes that are “done in the wild” (i.e., what we referred to as “uncontrolled policy experiments”).

It is an open question whether any of these models will actually be used in practice as inputs to policy making. However, some government agencies, particularly the Congressional Budget Office, are already using life-cycle models as “behavioral inputs” to their long term forecasting model. It seems likely that inter-agency competition will ultimately lead to the adoption of better analytical tools, and hopefully greater investment in data collection and (controlled) experimentation.

[15] Social Security, Medicaid and Medicare constituted 7.6% of GDP in 2000 and nearly 50% of the Federal Budget, and is projected to grow to 8.3% of GDP by 2005.

6 References

Benítez-Silva, H., Buchinsky, M., Chan, H. Cheidvasser, S. and J. Rust (1999): "An Empirical Analysis of the Social Security Disability Application, Appeal and Award Process" *Labour Economics*, **6**, 147–178.

Benítez-Silva, H. Buchinsky, M. and J. Rust (2003a): "How Large is the Bias in Self-Reported Disability?" forthcoming, *Journal of Applied Econometrics*.

Benítez-Silva, H. Buchinsky, M. and J. Rust (2003b): "How Large are the Classification Errors in the Social Security Disability Award Process?" manuscript, University of Maryland.

Berlin, G., Bancroft, W., Card, D., Lin W., and P.K. Robins (1998): "Do work incentives have unintended consequences? Measuring 'entry effects' in the Self-sufficiency project," working paper, SRDC.

Bernheim, D., Skinner, J., and S. Weinberg (2001): "What Accounts for the Variation in Retirement Wealth Among U.S. Households?" *American Economic Review*, **91-4**, 832–857.

Card, D., Robins, P.K., and W. Lin (1997): "How Important are 'Entry Effects' in Financial Incentive Programs for Welfare Recipients?" working paper 97-01-E, Social Research and Demonstration Corporation.

Engen, E., Gale, W.G., and C.E. Uccello (1999): "The Adequacy of Household Saving," *Brookings Papers on Economic Activity*, **2**, 65–165.

Ford, R., Gyarmati, D., Foley, K., and D. Tattrie (2003): "Can Work Incentives Pay for Themselves? Final Report on the Self-Sufficiency Project for Welfare Applicants," manuscript, Social Research and Demonstration Corporation.

Geanakoplos, J., Mitchell, O.S., and S. Zeldes (1999): "Social Security Money's Worth," in *Prospects for Social Security Reform,* O.S. Mitchell, R.J. Myers, and H. Young (eds.), University of Pennsylvania Press.

Heckman, J.J., Hotz, V. J., and M. Dabos (1987): "Do We Need Experimental Data to Evaluate the Impact of Manpower Training on Earnings?" *Evaluation Review*, **11-4**, 395–427.

Hurd, M. and J.P. Smith (2001): “Anticipated and Actual Bequests,” in *Themes in the Economics of Aging*, David A. Wise (ed.), 357–391.

LaLonde, R. J. (1986): “Evaluating the Econometric Evaluations of Training Programs with Experimental Data” *American Economic Review* **76-4** 604–620.

Moffitt, R. (2003): “The Role of Randomized Field Trials in Social Science Research: A Perspective from Evaluations of Reforms of Social Welfare Programs” manuscript, Department of Economics, Johns Hopkins University.

Muller, L.S. (1992): “Disability Beneficiaries Who Work and Their Experience Under Program Work Incentives” *Social Security Bulletin*, **55-2**, 2–19.

Muller, L.S. (2000): “The Search for Evidence of a Labor Supply Response to a Benefit Offset” Office of Research, Evaluation and Statistics, Social Security Administration.

Muller, L.S., Scott, G., and B.V. Bye (1996): “Work and Earnings of SSI Disability Recipients” *Social Security Bulletin* **59-2** 22–42.

Neumark, D. and E.T. Powers (2003): “The Effects of Changes in State SSI Supplements on Pre-Retirement Labor Supply” NBER working paper 9851.

Rupp, K. and C.G. Scott (1996): “Trends in the Characteristics of DI and SSI Disability Awardees and Duration of Program Participation” *Social Security Bulletin*, **59-1**, 3–70.

Rust, J. and C. Phelan (1997): “How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets,” *Econometrica*, **64-5**, 781–832.

Todd, P. and K. Wolpin (2003): “Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility: Assessing the Impact of the School Subsidy Program in Mexico,” manuscript, University of Pennsylvania.

Tuma, N. (2001): “Approaches to Evaluating Induced Entry into a New SSDI Program with a $1 Reduction in Benefits for each $2 in Earnings,” manuscript, Stanford University.

Wheeler, P.M. (1996): “Message from the Associate Commissioner,” *Social Security Bulletin*, **59-1**, p. 2.