Dynamic Structural Models of Retirement and Disability

Moshe Buchinsky, Brown University and NBER
John Rust, Yale University and NBER
Hugo Benítez-Silva, SUNY–Stony Brook
October 6, 2000

Summary of Proposal:

We propose to use all available waves of the Health and Retirement Survey (HRS/AHEAD) to estimate a comprehensive dynamic programming (DP) model of behavior at the end of the life cycle that provides a detailed treatment of the Social Security Administration's (SSA) Old Age and Survivors (OASI), Supplemental Security Income (SSI) and Disability Insurance (DI) programs. Major changes to these programs are being contemplated. Yet, we currently lack a unified model of social insurance at the end of the life cycle that can help us evaluate the behavioral and distributional impacts of these policies. Particular attention is paid to developing, estimating, and testing a multi-stage dynamic programming (DP) model of the SSI and DI application, appeal, and award process, for (possibly) heterogeneous agents.

We are developing a tractable empirical model that captures an individual's decisions regarding (1) labor supply and retirement (3) application for OA, DI and SSI benefits, and (4) consumption and savings. The resulting model will allow us to derive predictions of the behavioral and welfare implications of policy changes.

While there is a large literature using reduced-form and static structural models that has investigated some of these issues, it suffers from two major shortcomings. First reduced-form models cannot be used for welfare analysis or to predict behavior responses to policy changes. Second, static structural models do not accurately reflect the level of complexity and uncertainty facing individual decision makers, nor do they capture the important dynamic elements of the decision processes. The DP model we will develop will circumvent these shortcomings, providing a tractable framework for analyzing individual behavior and well-being, and forecasting their response to a wide range of policy changes.

Our model could provide new insights into a number of puzzling aspects about disability in the U.S. One puzzle is to determine the factors responsible for the pronounced swings in DI incidence rates in recent years. Another puzzle is to determine why the fraction of Americans receiving SSI and DI benefits continues to increase despite overwhelming epidemiological evidence of steady improvements in various objective indicators of health status. The SSA is currently contemplating significant changes to the disability award process, in order to reduce delays, and reduce large unexplained state-level differences in award rates.

We will use detailed health and functional status indicators from the HRS to evaluate whether or not there are alternative screening rules that can reduce the level of classification errors in the DI award process. Our estimated DP model will produce detailed predictions of the behavioral and welfare effects of changes in benefit levels, delays, the probability of being awarded benefits, and the probability that a DI beneficiary will be audited. This framework will allow us to develop methodologies for characterizing *efficient* policies, i.e., those that minimize the expected discounted cost of providing a stream of social insurance benefits subject to the constraint that individuals' expected discounted utilities are at least as high as under the *status quo.*

1 Introduction

We propose to develop a unified empirical model of social insurance at the end of the life cycle using data from the Health and Retirement Survey (HRS). We will pay particularly close attention to accounting for the three key components of social insurance provided by the Social Security Administration (SSA): 1) Old Age and Survivors Insurance (OASI), 2) Medicare and Medicaid, and 3) Disability Insurance (DI) and Supplemental Security Income (SSI) benefits. Because of the rapid aging of the U.S. society, none of these programs are in long run actuarial balance, and sooner or later Congress will have to decide on whether significant tax increases or benefit cuts will be made to keep the program in the black. A number of important changes will begin to take effect in coming years as a result of the 1983 Social Security amendments, which was intended to create incentives for delayed retirement in order to restore the long run balance in the OASI program. In addition to increasing contribution rates, the 1983 amendments increased the normal retirement age (NRA) from 65 to 67, increased the delayed retirement credit (DRC) from 1% to 8%, and decreased the "retirement test" tax on post-retirement earnings from 50% to 33%. In recent years much more radical changes to the Social Security program have been proposed, including full or partial privatization, the introduction of individual accounts, and proposals that would invest a substantial share of the Trust Fund in equities. Another change that was enacted recently is the removal of the earnings test for those 65 and over years of age: the effects of this change can be easily examined in our framework.

Substantial changes to the SSI and DI programs are also being contemplated, resulting in part from the huge increases in processing delays and backlogs following the rapid growth in applications and appeals during the early 1990s. As part of its "Disability Process Redesign" (DPR) plan, the SSA is currently considering major changes in the multi stage application and appeal process. The proposed changes include simplifications designed to reduce delays in initial award and appeal decisions, and the use of standardized functional impairment indices to reduce large unexplained state-level differences in award rates. All of these proposed policy changes will have significant effects on the structure of the Social Security program that cannot be accurately predicted and examined using reduced-form methods that estimate behavioral relationships under the *status quo.* SSA currently does not have a unified behavioral model that would enable it to forecast the behavioral and welfare implications of any of these policy changes. The only comprehensive way to deal with these multitude of questions is by modeling explicitly the decision processes by both the individuals and the SSA, under realistic assumptions governing their behavior.

Fortunately, in recent years increasingly realistic dynamic structural models have been formulated and estimated in the academic literature. These models include Rust and Phelan (1997), which estimated a detailed dynamic programming model of the OASI and Medicare program. In contrast to reduced-form models in the literature, the innovation of Rust and Phelan is that their paper showed that a number of previously puzzling aspects of retirement behavior are simply artifacts of particular details of the Social Security rules. In particular, it showed that OASI and Medicare benefits have complex *interacting* incentive effects. For example, while many of the features of the OASI rules encourage early retirement, the fact that individuals do not qualify for Medicare benefits until the normal retirement age creates a substantial incentive for a significant fraction of "health insurance constrained" individuals to delay retirement. Their DP model accounts for wide variety of phenomena observed in the data, including the pronounced peaks in the distribution of retirement ages at 62 and 65. These results illustrate the importance of developing an integrated dynamic model of social insurance at the end of the life cycle. Such a model can help us understand a number of other puzzling aspects about individuals' decisions regarding retirement, disability application, and health insurance, and could have substantial practical value as a tool to assist policy evaluation and forecasting by government organizations such as the SSA. We will use our model to analyze a number of important policy-related questions and issues including:

1. Why does the fraction of Americans on the DI and SSI roles continue to increase when epidemiological studies find that health of older Americans has improved over time?

2. What is the relative importance of changes in award rates, unemployment rates, and social factors in the large swings in DI incidence rates in recent years?

3. What role do delays in the DI award process have on incentives to apply or appeal? Will proposals to speed up this process increase the number of applications and awards?

4. Will the 1983 Social Security Amendments, particularly the increase in the NRA and DRC, cause individuals to significantly delay the age at which they apply for OA benefits? How would individuals be affected if the Medicare eligibility age (MEA) were also changed?

5. Will the increase in the NRA increase the incentive to apply for SSI and DI benefits prior to the NRA? If so, to what extent will any reduction in the costs of the OA program due to the increased NRA be offset by an increase in the cost of the SSI/DI program?

6. How would retirement incentives and individual welfare be affected by an introduction of "individual accounts" similar to the plan Governor Bush has proposed?

Although our model can address a wide range of policy issues, we will devote particular attention to modeling the dynamics of disability, mortality and health, and the factors influencing decisions to apply for SSI and DI benefits since these are relatively volatile programs that have grown at unsustainable rates in the past.

2 Overview of Issues and Trends in the DI Program

There is a large empirical literature studying the factors affecting DI applications and awards, and a somewhat smaller literature on the SSI program. This literature, (see, e.g. Rupp and Stapleton, 1996 or Stapleton et al. 1994 for an overview) has identified a number of important factors: 1) benefit levels, 2) program leniency as measured by award probabilities and audit rates, 3) strength of the demand for labor (as measured by unemployment rates), 4) the availability (or lack thereof) of alternative sources of support, and 5) social attitudes, particularly those affecting any possible stigma associated with receiving DI benefits. However, the relative importance of these factors is still not well understood, hampering SSA's ability to do policy analysis and short and long term forecasting. Figures 2.1 and 2.2 illustrate some of the key historical and forecasted trends in the DI program.[1]

Figure 2.1 summarizes the historical and projected trends in the size and cost of the DI program, measured by the DI prevalence rate and by the ratio of DI expenditures to GDP. The right hand side of Figure 2.1 shows a rapid rise in the cost of the DI program since its inception in 1956 until the mid 1970s, interrupted only by a decrease in the cost of the program during a period of retrenchment from 1977 to 1990, and a decrease during the economic boom of recent years. The Actuary forecasts continued growth in the program over the next 75 years, topping out at roughly 0.9% of GDP by 2075. The left hand panel of figure 2.1 plots historical and projected prevalence of DI over the period 1988 to 2075. We see that prevalence rates have increased steadily over the period 1988 to 1996, pausing briefly in 1997 and 1998. The Actuary forecasts a particularly rapid increase in DI prevalence until 2030, by which time most of the baby boom generation will have reached normal retirement age. Thereafter prevalence continues to grow at a more moderate rate

[1] We thank Eli Donkar and William Ritchie of the Social Security Office of the Actuary for providing forecasts of GDP and DI roles and expenditures. These forecasts were used in the 1999 Social Security Trustee's Report and are discussed in more detail in an appendix to the Final Report of the 1999 Technical Panel on Assumptions and Methods of the Social Security Advisory Board, of which Rust was co-author.

reaching 7% of the insured population by 2075. The "adjusted" prevalence curve reflects a forecast based on a counterfactual assumption that the age distribution of the U.S. remains at its 1998 values. The unadjusted prevalence rates increase from 0.4 to 0.7 between 1999 and 2075, whereas the adjusted prevalence rate only increases to 0.5. Thus, in rough terms, population aging accounts for only one third of the projected increase in prevalence of DI in the next 75 years.

Figure 2.1: Historical and Forecasted Growth in SSDI Roles and Costs

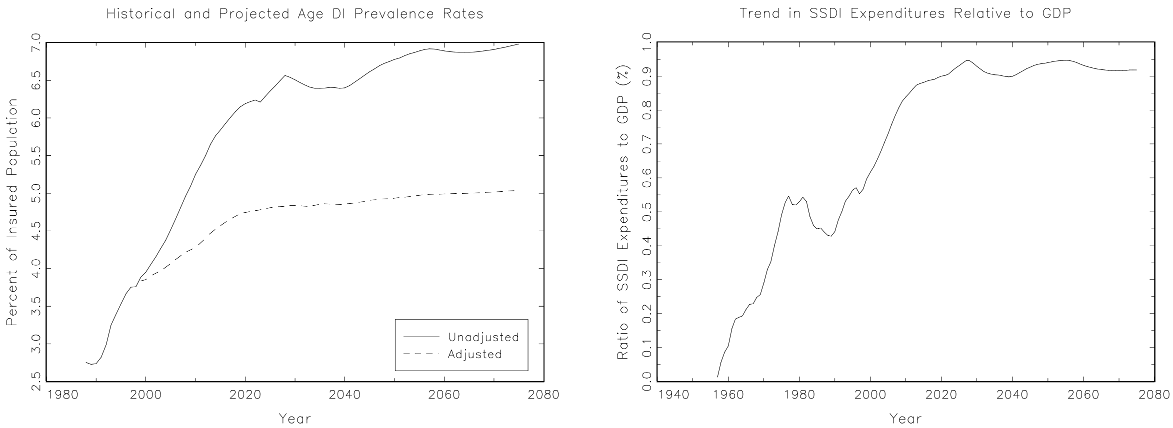
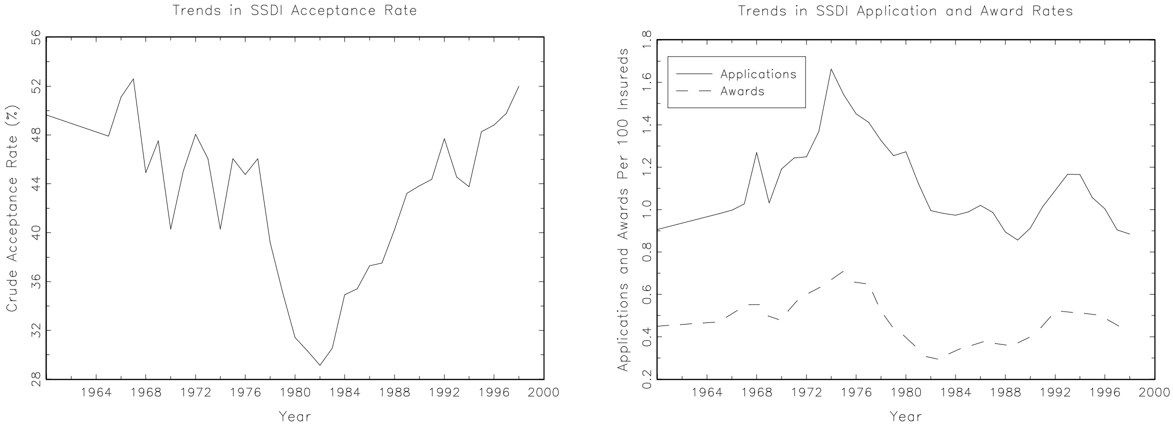


Figure 2.2 illustrates some of the historical volatility in application and award rates. The left hand panel plots the trend in the "crude" acceptance rate—the ratio of the number of new DI awards to the total number of applications and appeals files in a given year. The right hand panel plots the ratio of DI applications and awards to the DI insured population. The award rate reached its lowest level in 1982 during the Reagan Administration, during a clamp-down on the DI and SSI programs. There was a large increase in audits, also known as "Continuing Disability Reviews" (CDR), during this period. The combined effect was to strongly discourage individuals from applying for DI benefits, as is evident in the right hand panel of figure 2.2. On the other hand DI applications and awards peaked in 1974. This peak in applications was due to several factors: 1) a recession in the early 1970s, 2) a rapid increase in benefit levels due to an error in the 1972 Social Security amendments which resulted in an inadvertent double indexing of Social Security benefits to inflation, and 3) a lenient policy towards DI applicants reflected in the high award rates in figure 2.2. SSI was also introduced in 1974, so the public may have actually perceived that SSA as encouraging applicants, reducing perceived "hassle costs" to applying for DI, or helping to reduce the stigma associated with receiving benefits. DI application rates began growing rapidly again in the early 1990s following a sustained period of growth in award rates. The causes of this rapid burst of growth are not fully understood, but high unemployment rates in the early 1990s, and a cutback in state General Assistance (GA) programs are thought to be important contributing factors. The passage of the Americans with Disabilities Act in 1990 was designed to force employers to accommodate workers with disabilities and thus reduce the incidence and prevalence of individuals receiving DI benefits, however, it may have had an opposite effect by helping to reduce the stigma of being disabled.

On the other hand, application rates declined equally quickly after 1993. This was also the peak year for enrollments in the AFDC program, and a period of high social stigma towards welfare recipients may have been one of the most important factors motivating the tough 1996 Welfare Reform Act. While part of the decline in application rates might be ascribed to an increase in real or perceived stigma towards AFDC and SSI recipients, the years after 1993 have also constituted the longest peacetime economic boom in recorded history. Only within a structural model can one make an attempt to disentangle the relative contributions of the various possible explanations for changes in DI application rates.

Figure 2.2: History of SSDI Application and Award Rates, Roles, and Costs



Our own previous empirical work using the HRS data shows that self-reported health and disability status indicators are powerful predictors of which individuals choose to apply or appeal for DI benefits. Nevertheless, the patterns of fluctuation in aggregate DI award and application rates suggests that individual decisions are governed only partly by "objective" physical and health conditions. Social, political, and economic factors have just as significant impact on application and awards at the aggregate level.

The paradox that DI prevalence rates have grown while the objective health status of Americans has improved suggests that the concept of "disability" used by the SSA is not based on an absolute objectively determinable measure of physical status, but is rather more akin to a socially defined concept whose absolute standards may change over time with changes in the political, social, economic, and technological environment. Clearly, the nature of physical/mental conditions that are regarded as disabling is very different in today's "information economy" than they were in an industrial/agrarian economy in the 1800s. It is not surprising therefore that SSA's Actuarial Study 114 (1999) documents significant changes in the distribution of impairments that are listed as the primary reasons being awarded DI benefits. For example, mental conditions in recent years account for a much larger share of disabilities than in 1982; 20% compared with only 11% of all awards, respectively (See also Gruber and Kubik 1997 for further discussion).

Although the DI program was intended to be an insurance program and not a means-tested transfer program, the DI program effectively has many of the characteristics of a welfare program. Our previous research indicates that most DI beneficiaries are in the lowest income brackets. A possible explanation for the upward trend in the prevalence of DI is the fact that a richer, more technically advanced society can afford the luxury of supporting an increasing fraction of its least healthy and productive citizens.

We believe that the way SSA administers the DI award process can help to create a "social standard" that has a powerful impact on the public's perception of the thresholds for mental and physical impairments that are sufficiently severe to constitute "disability". Indeed, we have shown (see Benítez-Silva et al. 2000) that DI applicants satisfy an "unbiased reporting" hypothesis, i.e., their self-reported disability status is an unbiased indicator of SSA's ultimate award decision. This finding suggests the possibility that tightening or loosening of DI award rates may have a double effect. Its direct effect is on the individuals' incentives for applying for benefits since it affects their chances of success. The indirect effect is through the individuals' self-perceptions of whether or not they believe they are, in fact, disabled. Our model incorporates this fact and allows "disability" to be a social standard that evolves slowly over time. Whether the current social standard is just a "veil" that affects individuals' decisions (through calculations of costs and benefits to filing for DI or SSI benefits), or whether the social standard also has direct affects on their self-perceptions of their capacity to work, is an extremely important issue that we plan to analyze.

3 Reduced-form Models of the DI/SSI Award Process

The DI and SSI programs can be viewed as a game between applicants and the government. The outcome of this game depends on the objectives of the government (e.g. Social welfare maximization), and preferences of the individuals. There is, however, an additional complication: the "government" is not one decision maker, but rather a hierarchical bureaucracy with over 15,000 administrators. These are divided among the 54 state Disability Determination Services (DDS's) that process initial applications and reconsiderations, and the more than 1,000 Administrative Law Judges (ALJ's) and Appeals Board that process appeals. The complete DI application, award, and appeal process has been modeled econometrically, using the SIPP panel data, in Hu et al. (1997) and Lahiri et al. (1995), and Benítez-Silva et al. (1999) with the Health and Retirement Survey. The latter paper estimated a detailed reduced-form multi-stage model of an individual's decision to apply for SSDI and SSI benefits, the DDS "first stage" award decision, the decision by rejected applicants to appeal an initial rejection, and the "second stage" decision by SSA's Administrative Law Judges and Appeals Board whether to award or deny an appeal. The paper finds that there is a large apparent return to appealing an initial rejection by the DDS. The "ultimate award rate" rises from about 46% at the first stage decision made by the DDS, to about 73% when the option to appeal is considered. However, this increased award rate comes at the cost of substantial delays, which appear to function as an implicit type-dependent "application fee" that dissuades opportunistic behavior. Indeed, the study not only shows that an individual's self-assessed health status is one of the most powerful predictors of application and appeal decisions, but also that it is a very powerful predictor of SSA's award decision.

The Benítez-Silva et al. (1999) study also revealed the importance of opportunistic economic factors and policy incentives affecting an individual's decision to apply for DI benefits. For example, only very few individuals who are over 62 apply for DI benefits, presumably because they can get early Social Security retirement benefits at this age. Even though DI pays a benefit equal to the full Primary Insurance Amount (PIA) payable at normal retirement age, in contrast to the early retirement benefits, the uncertainties and opportunity costs associated with the long delays in the award process are not adequately compensated by the higher level of benefits.[2]

However, Benítez-Silva et al. (1999) is just an exploratory analysis into the overall disability award process, and their reduced-form results cannot be used to forecast the effects of policy changes, such as changes in award rates, audit rates, or benefit levels. The only way this can be done is via a structural econometric approach that models the individual's application decision. There are a number of static structural models of the DI application process, such as Halpern and Hausman (1978) and Kreider (1999). The problem with these models is that they are incapable of capturing the dynamic aspects of the application and appeal process. We use dynamic programming to model the sequential decision process of whether or not to apply for DI benefits, allowing us to explicitly account for the potential returns, uncertainties, and "hassle costs" associated with submitting an initial application for benefits, and deciding whether to appeal or re-apply upon denial. The success of the DP approach in modeling and predicting retirement behavior (see, e.g.

[2] Other econometric studies have revealed the importance of occupational and economic factors (e.g. the local unemployment rate) on disability application and award decisions. It is also well known (see Social Security Advisory Board, 1998), that there are large differences in the delays and award rates among the 54 state-run DDS's and the over 1,000 ALJ's who handle appeals of DI denials.

Rust and Phelan 1997) suggests that a DP model would produce more accurate forecasts of changes in the OA and DI programs than would be possible using static structural models.

4 Dealing with the Endogeneity of Disability and Health Status

A pervasive concern with the use of self-reported health and disability measures in behavioral models is that they are biased and endogenous. A commonly suggested explanation is that survey respondents exaggerate the severity of health problems and incidence of disabilities in order to rationalize labor force non-participation, application for disability benefits and/or receipt of those benefits. We re-examined this issue (see Benítez-Silva et al. 2000) using a subsample of HRS respondents who applied for DI or SSI benefits, and for which the SSA's award/deny decision could be ascertained. For these individuals we have two independent indicators of their disability status: the individual's self-reported disability status $\widetilde{d}$ and SSA's award decision $\widetilde{a}$. The indicator $\widetilde{d}$ equals 1 if the respondent reports that they have a health problem that prevented them from working entirely, and 0 otherwise. The indicator $\widetilde{a}$ equals 1 if the respondent was ultimately awarded DI or SSI benefits (including those who were initially denied but were successful upon appeal), and 0 otherwise. We tested the hypothesis—which we term the *unbiased reporting* (UR) hypothesis—that DI and SSI applicants are aware of the criteria and decision rules that SSA uses in making awards, and individuals act as if they were applying these same criteria and rules when reporting their own disability status. The UR hypothesis amounts to moment restrictions given by

$$E\{\tilde{a} - \tilde{d}|x\} = 0 \Longleftrightarrow Pr\{\tilde{a}|x\} = Pr\{\tilde{d}|x\}, \tag{1}$$

where x denotes a vector of objectively measurable health and socioeconomic characteristics. Using a variety of parametric and non-parametric tests (e.g. Horowitz and Spokoiny), that have optimal rate of convergence against a broad class of non-parametric alternatives, we were unable to reject the UR hypothesis, even when we used a large number of objectively measurable health and socioeconomic characteristics similar those the SSA uses in making its award decisions, as instruments. Thus our results provide evidence that DI applicants do not exaggerate their disability status—at least in anonymous surveys such as the HRS.

Given the relatively low sample sizes and the absence of functional form restrictions, the power of these conditional moment tests can be low. Hence, we also performed Wald and Likelihood Ratio tests of a parametric version of the UR hypothesis where the conditional probabilities are derived from the bivariate probit function:

$$Pr\{\tilde{a}|x\} = E\{I[x\beta_a + \epsilon_a \geq 0]\}, \quad Pr\{\tilde{d}|x\} = E\{I[x\beta_d + \epsilon_d \geq 0]\}, \tag{2}$$

where (ϵ_a, ϵ_d) have a bivariate normal distribution with mean 0, unit variance, and correlation coefficient ρ. Here, the UR hypothesis amounts to the restriction $\beta_a = \beta_d$. We employ both a one-group model and two-group model. In the latter we allow for two types of individuals in the population. In this model there are two sets of parameter, one for each group, β_{dj}, β_{aj}, and corresponding ρ_j, $j = 1, 2$. The UR hypothesis imply then that $\beta_{aj} = \beta_{dj}$ for $j = 1, 2$. Again, we are unable to reject the UR hypothesis at conventional significance levels.

The parametric model suggests the following interpretation of the UR hypothesis. Without loss of generality, SSA's ultimate award decision can be represented by an index rule depending on information x that is observed by the econometrician and other information ϵ_a that is observed by SSA, but unobserved by the econometrician. The coefficient vector β_a represents the weights SSA assigns to various health conditions and socioeconomic characteristics in coming up with an overall "disability score" $x\beta_a + \epsilon_a$. Individuals with sufficiently high disability scores are awarded benefits (represented by the cutoff rule $x\beta_a + \epsilon_a \geq 0$) and the rest are denied benefits. Similarly, the individual's self-reported disability status can also be represented by an index rule with the corresponding vector of weights β_d, and private information ϵ_d, that is observed by the individual but not by the econometrician. The UR hypothesis amounts to a rational expectations restriction that individuals use the same variables and weighting vector — i.e. the same criteria — as SSA uses to make DI determinations. However, individuals' self-reports also depend on private information ϵ_d that is unobserved by the SSA. Likewise, the SSA's award decision may be affected by "bureaucratic noise" ϵ_a that the individual does not observe. This implies that the indicators $\tilde{a}$ and $\tilde{d}$ are not perfectly correlated. However, if UR restriction holds then they have identical conditional probability distributions, a restriction that we are unable to reject empirically. Thus, while we conclude that self-reported disability status appears to be "accurate" and "exogenous" in the sense of conforming to the same standards SSA uses, from an aggregate perspective disability status is endogenous, since the UR hypothesis implies that changes in SSA's criteria for determining disability status affects individuals' self-perceptions of disability status. An important aspect of our research will be to examine the implications of this hypothesis for policymaking by SSA.

5 A DP Model of OASI, SSI, DI and Medicare

This section outlines our plans to estimate a dynamic programming (DP) model of male and female labor supply and social insurance application decisions at the end of the life cycle. The following key aspects of the U.S. Social Security program and private insurance and pensions will be modeled: (1) Old Age and Survivors (OASI); (2) Medicare/Medicaid (MC, MCA); (3) Disability, including Supplemental Security Income (SSDI, SSI); (4) private health insurance; (5) private pensions and annuities; (6) Unemployment and Worker's Compensation (UI, WC), and (7) joint decisions of couples in a household. Our initial work will focus on developing a model that incorporates items (1) to (4), and once this is successful we will go on to extend the model to incorporate items (5) to (7).

Our ultimate goal is to develop a DP model where decisions are made on a monthly basis, however our initial work will focus on building a DP model where decisions are taken at annual intervals. A monthly model is necessary for modeling certain details of the DI application and appeal process, since a sequence of application, award/denial, and appeal decisions can often occur within a single year. We plan to use the value functions (to be described below) from our annual model as terminal value functions for a series of nested monthly level DP "sub-problems". This extended model will incorporate the month by month decisions on labor supply, pensions, disability unemployment, workmen's compensation application decisions, etc. The DP model will be programmed in C, using the Parallel Virtual Machine (PVM) library to distribute the computation over several networked Unix workstations located at Yale, Brown, and SUNY. All data and source code developed from NIH funding will be freely available to the research community over the web.

Perhaps the best way to describe the DP model and illustrate the feasibility of our approach is to formulate, solve, and simulate a specific prototype of the DP model we are planning to extend, and empirically implement, over the course of this research project. We emphasize that the following model is preliminary and is presented to provide a concrete illustration of how a relatively simple and parsimoniously parametrized specification of a DP model can yield richly detailed, intuitively plausible solutions. We do not claim that this illustrative model is empirically "realistic" at this point. Indeed, the purpose of this proposal is to obtain the necessary funding to develop empirically realistic specifications whose unknown parameters will be econometrically estimated using the HRS/AHEAD data.

Although simplified in several respects, our illustrative model already constitutes one of the most ambitious and detailed computational models of life-cycle behavior that has ever been formulated and solved. We pay special attention to providing a fairly realistic treatment of the main features of the U.S. Social Security system including the Disability Insurance program. Except for the Phd thesis of Benítez-Silva (2000), no other study has attempted to model both the consumption/savings and labor/leisure decisions jointly in a framework that includes a realistic treatment of the U.S. Social Security system.[3]

We solve an 80 period 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. These choices are denoted by the three variables $\{l_t, c_t, ssd_t\}$. Here, l_t denotes *leisure,* that is, the amount of waking time devoted to non-work activities, where the amount of waking time during a year (assumed to be 12 hours per day, or a total of 12·365 hours over the year) is normalized to 1. Thus, $l_t = 1$ corresponds to not working at all, $l_t = .543 = (12 \cdot 365 - 2000)/(12 \cdot 365)$ corresponds to full time work (i.e., working 2000 hours per year), while $l_t = .817 = (12 \cdot 365 - 800)/(12 \cdot 365)$ corresponds to part time work (i.e., working 800 hours per year). The quantity c_t denotes consumption expenditures, which are, naturally, treated as a continuous variable. The quantity ssd_t denotes the individual's *Social Security decision,* where $ssd_t = 1$ denotes the decision to apply 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 under the early retirement age (denoted by ERA, currently set at 62) are not allowed to receive OA benefits. Hence, their choice set reduces to $ssd_t \in \{0, 2\}$. After the normal retirement age the OA and DI programs are formally merged, thus the choice set reduces to $ssd_t \in \{0, 1\}$. Also if a person is already receiving OA benefits they cannot re-apply for additional benefits, so they face no further choices unless ERA $\leq t <$ NRA, 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: current age t, net (tangible) wealth w_t, the individual's Social Security state ss_t, and average wage, aw_t. The ss_t variable has three mutually exclusive values: $ss_t = 0$ (not entitled to benefits), $ss_t = 1$ (entitled to OA benefits), and $ss_t = 2$ (entitled to DI benefits). As noted above, the OA and DI programs are merged after the NRA, so any individual with $ss_t = 2$ is automatically converted to $ss_t = 1$ when $t \geq$ NRA.

The average wage aw_t is a key variable in the DP model, serving two roles: 1) it acts as a measure of "permanent income" that serves as convenient "sufficient statistic" for predicting the evolution of annual wage earnings, 2) it is key to accurately modeling the rules governing payment of Social Security benefits. The Social Security rules are fairly complicated, but can be summarized as follows. An individual's highest 35 years of earnings are averaged, and the resulting *Average Indexed Earnings* is denoted (on an annualized basis) as aw_t. The potential Social Security benefit rate for retiring at the normal retirement age (NRA), the so-called *Primary Insurance Amount,* (PIA) is a piece-wise linear, concave function of aw_t whose value (on an annual basis) is denoted by $pia(aw_t)$.

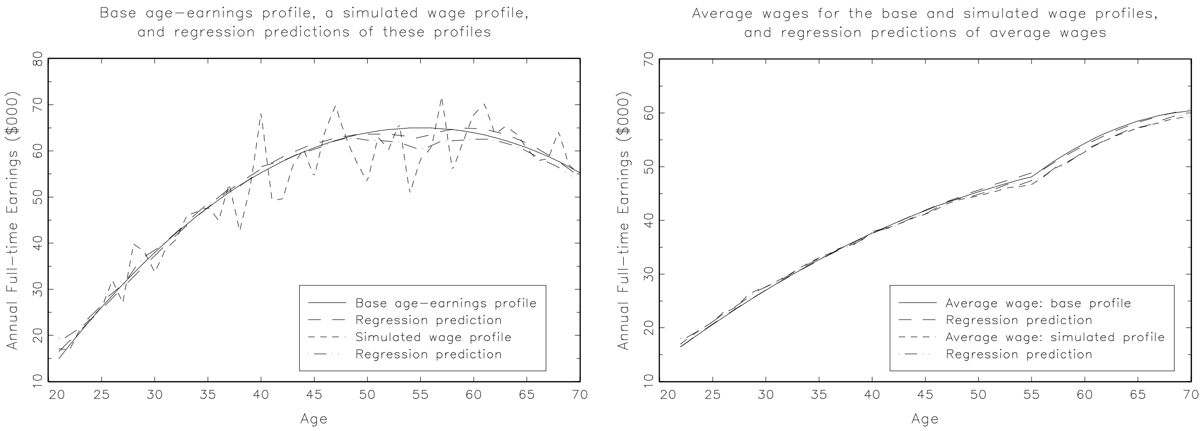
The fact that aw_t is an average of the 35 years of highest earnings suggests that to accurately model the U.S. Social Security system we would need to carry as state variables an individual's entire past earnings history. However, due to the "curse of dimensionality" of solving high-dimensional DP problems, this approach is not feasible. Fortunately, it is possible to approximate the evolution of average wages in a Markovian fashion: i.e. to find a formula that allows us to make highly

[3] Rust and Phelan (1997) and French (2000) have also estimated DP models with a detailed treatment of Social Security, but neither study models the consumption/savings decision explicitly.

accurate predictions of next period average wage aw_{t+1} using only age, t, current average wage aw_t, and current period earnings y_t.

Figure 5.1 illustrates this. We assumed that the potential earnings an individual could make from working full time are given by *IID* deviations about a quadratic age-earnings profile such as is shown in the left panel of figure 5.1 below. Assuming for the moment that this is the "true model" for individual earnings, we generated an artificial data set of simulated earnings histories assuming each individual works full time from age 21 until retirement at the MRA. For each simulated earnings history, we used the actual Social Security rules to compute the corresponding sequence of average wages. This is shown in the right panel of figure 5.1. Notice that since aw_t is a long moving average of past wages, the implied path for $\{aw_t\}$ is far less variable than the path of earnings $\{y_t\}$.

Figure 5.1: Age-earnings profiles and the implied path of average wages



We used the observed sequence of average wages as regressors to estimate the following (mis-specified) log-normal regression model of an individual's annual earnings:

$$\log(y_t) = \alpha_1 + \alpha_2 \log(aw_t) + \alpha_3 t + \alpha_4 t^2 + \eta_t \tag{3}$$

Even though this regression does not correspond to the true process by which earnings are generated (i.e. as *IID* innovations around the quadratic "base case" age-earnings profile in the left panel of figure 5.1), the R^2 of the above regression is .99940, and as we can see from the curves labelled "regression prediction" in the left panel of figure 5.1, the misspecified regression model in equation (**??**) tracks the true age-earnings profile and the conditional expectation of the wage distribution very closely. We also estimated a log-normal regression model for average wages:

$$\log(aw_{t+1}) = \gamma_1 + \gamma_2 \log(y_t) + \gamma_3 \log(aw_t) + \gamma_4 t + \gamma_5 t^2 + \epsilon_t \tag{4}$$

We also found of very high R^2 (.99998) for this specification. The estimated standard error for ϵ_t was also very close to zero, a result of low variability of the $\{aw_t\}$ sequences. As we can see in the right hand panel of figure 5.1, the misspecified regression predictions of average wages, using the Markovian approximation in equation (**??**) above, provides a highly accurate estimate of "true" average wages, i.e. the values that come out of the actual formula used by Social Security. This finding is extremely encouraging, since it allows us to make an important computational simplification without compromising our ability to accurately model the Social Security rules. In subsequent research we we will re-estimate these regression models using the restricted HRS data on Social Security earnings histories and average wages: we used simulated data here to avoid any possible confidentiality issues.

Our DP model also accounts for the other key details of the Social Security rules. For example, there is a penalty for retiring prior to the normal retirement age: an individual's 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 (to a maximum of NRA-ERA) that the individual first starts receiving OA benefits. Our DP model uses the actuarial reduction rate $g_1 = .0713$ that is currently in effect in the U.S. If a person is accepted into the DI program, they receive the full PIA regardless of their age. To increase incentives to delay retirement, the 1983 Social Security reforms gradually increased the NRA from 65 to 67 and increased the *delayed retirement credit.* This is a permanent increase in the PIA by a factor $\exp\{g_2 k\}$ where k 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, in the simulations below we use the current value $g_2 = 0.077$. 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. For this reason it is not optimal to delay applying for OA benefits beyond the MRA: due to mortality, further delays necessarily reduce the present value of OA benefits the person will collect over their remaining lifetime.[4]

The final aspect of the Social Security rules concerns taxation of benefits. Individuals whose combined income (including Social Security benefit) 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 the Federal income tax on wage earnings. 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 (currently $10,800) 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 (currently $17,000).[5] For individuals who are above the MRA, there is no earnings test.

We can embed all of the Social Security rules into a single function, the Social Security *benefit function* $ssb_t(aw, y, ss, ssd)$. This function specifies the benefits an individual of age t will get if they have average wage aw, have earned income of y, are in Social Security state ss and have taken Social Security application decision ssd. The ssb_t function includes as sub-functions the determination of PIA from average wages, the potential reduction in benefits due to earnings test, and the adjustment of the PIA for retirement before or after the NRA as discussed above. We have coded this function in C and Gauss. It is a simplified version of SSA's own ANYPIA program except that our version incorporates SSI and DI benefits as well as OA benefits.

Finally, our model provides a simplified account of the DI award/appeal process. We assume that if a person applies for DI benefits but continues to work, there is 0 chance of being awarded benefits due to the test for "substantial gainful activity". However, even if a person is not working, there is only a probability $p_1(t, w_t)$ that a person of age t and wealth w_t will be awarded benefits. Even if benefits are awarded there is a six month waiting period before they can be paid. Adding on the typical delays in the application and appeal process, we assume that if a person applies for DI at the beginning of year t that they will not receive any benefits during that year, but will be notified whether they are accepted into the program at the start of year $t + 1$ at which time

[4] This is not quite true: it is possible that if wages continue to rise with age, then continued work could increase the individuals' average wage and thus the present value of Social Security benefits from delaying retirement past the MRA, even accounting for mortality. However, if wages do not increase sufficiently fast beyond the MRA, then it will still be true that it will not be optimal to delay retirement beyond the MRA, in the sense that further delays necessarily reduce the present value of Social Security benefits.

[5] The earnings test provision has been recently eliminated for individuals 65 and over. However, it was still in place during the time the HRS data was collected and therefore we include it in our model. This policy change allows us to perform a policy experiment that we present in section 7.

they can start receiving benefits. We do not model the trial work period at this stage, and assume that if a DI recipient works full or part time, they will be immediately detected and removed from the program. We assume that SSA also randomly audits DI recipients who are not working, and with probability $p_2(t, w_t)$ a DI recipient can be removed from the roles. We currently set the audit/removal probability to a very low value, about 1%, reflecting the fact that audits have not been extensively used since the Reagan/Bush administration, although the incidence of CDR's has been increasing in recent years.

To complete the specification of the DP model, we need to make some assumptions about individuals' health, mortality, and preferences for leisure and consumption. In this prototype version of the model we have not included an extra state variable to describe health status. Instead we assume that all individuals of a given age are equally healthy, although we do implicitly account for a mild "aging effect" by assuming that the disutility of labor effort gradually increases as the person ages. We assume that the maximum possible age for any individual is 100, but individuals can also die prior to age 100 according to age specific death rates taken from the U.S. Decennial Life Tables (Table 2 in vol 1 no. 1 of 1997 Report of the U.S. Center for Disease Control).

If this proposal is funded, a major priority will be devoted to investigating the best way to incorporate health status into the DP model using the rich set of variables in the HRS/AHEAD data set. We will introduce one or more health state variables that can enable us to distinguish between temporary acute health problems and chronic health conditions, as well as measures that account for varying degrees of severity of health problems, with the most extreme of these (short of death) being interpreted as disabling health problems. We will also use HRS/AHEAD to estimate a richer model of mortality, including health, wealth, marital status, education, and other variables as covariates instead of just age and sex that standard life-tables provide.[6]

We assume that the individual's utility for consumption and leisure is given by

$$u_t(c, l) = \frac{[c^{\lambda_t} l^{1-\lambda_t}]^{1-\rho_t}}{(1-\rho_t)} \tag{5}$$

where we allow the parameters (λ_t, ρ_t) to depend on age, reflecting "aging effects". The ρ_t parameter is the coefficient of relative risk aversion, and it could potentially increase with age, although we fix $\rho_t = 1.5$ for all t in the simulations below. The λ_t parameter governs the relative weight the individual places on consumption versus leisure. Decreasing the value of λ_t corresponds to placing less weight on consumption and more weight on leisure, or stated differently, is equivalent to increasing the marginal disutility of labor supply. In the simulations below we assume $\lambda_t = .6/(1 + t/50)$, so that λ_t gradually decreases from $\lambda_{21} = .563$ at age 21 to $\lambda_{100} = .2666$ at age 100. This can be interpreted as a way of modeling the effect of gradually declining health with age. But as noted above, there is no "disability" in our model: we do not allow the marginal disutility of effort to suddenly increase, or the wage rate to suddenly decrease, although this would be easy to do in future extensions of the model that allow for a health state variable. We also assume that individuals have a bequest motive which is captured by a utility of bequeathed wealth w_t (assumed to occur at death) of $B(w) = .2w^{(1-\rho_t)}/(1-\rho_t)$, where $\rho_t = 1.5$.

Regarding wage earnings, we model the stochastic evolution of full-time wages for full-time workers via the regression model in equation (**??**) discussed above. These wages are based on 2000 hours of work per year. Part time workers are assumed to work 800 hours per year and at a wage rate that is only 83.75% as large as the full-time wage rate. Thus, the annual earnings paid to a

[6] We are well underway in pursuing this estimation using some parametric and semiparametric methods, which will enable us to predict how the probability of death varies as a function of a number of different observed and unobserved characteristics.

part-time worker is 33.5% ($.8375 * 800/2000$) of the wages paid to a full-time worker, as generated by the model in equation (**??**).

We account for time and financial costs of applying for DI benefits as a reduction of consumption and leisure. Initially we assume zero financial costs to applying for DI, but assume that there is a small cost in leisure time equal to .005, or a total of 22 hours of waking time to gather the information necessary to file an application for benefits. We assume there are no time or financial costs to applying for OA benefits. In subsequent work we will treat these costs as unknown parameters to be estimated. We use the notation $u_t(c, l, ssd, ss)$ to alert readers to the possibility of allowing for potential "stigma costs" to being on the DI roles. However, initially we assume there is no stigma.

Let $V_t(w, aw, ss)$ denote the individual's value function, the expected present discounted value of utility from age t onward for an individual with current wealth w, average wage aw and in Social Security state ss. We solved the DP problem via numerical computation of the Bellman recursion for V_t given by

$$V_t(w, aw, ss) = \max_{\substack{0 \le c \le w \\ l \in \{.54, .81, 1\} \\ ssd \in A_t(ss)}} V_t(w, aw, ss, c, l, ssd) \tag{6}$$

where

$$\begin{aligned} V_t(w, aw, ss, c, l, ssd) = [u_t(c, l, ssd, ss) \quad & + \beta(1 - d_t) EV_{t+1}(w, aw, ss, c, l, ssd) \\ & + \beta d_t EB(w, aw, ss, c, l, ssd)] \,. \end{aligned} \tag{7}$$

$A_t(ss)$ denotes the set of feasible Social Security choices for a person of age t in Social Security state ss (see discussion above) and d_t denotes the age-specific death rate for males, taken from the 1997 edition of the U.S. Decennial life tables as noted above. The functions EV_{t+1} and EB denote the conditional expectations of next period's value and bequest functions, respectively, given the individual's current state (w, aw, ss) and decision (c, l, ssd). Specifically, we have

$$\begin{aligned} EV_{t+1}(w, aw, ss, c, l, ssd) \quad = \quad & \int_{y'} \sum_{ss'=0}^{2} V_{t+1}(wp_t(w, aw, y', ss, ssd), awp_t(aw, y'), ss') \\ & \times f_t(y'|aw) g_t(ss'|aw, w, ss, ssd) dy', \end{aligned} \tag{8}$$

where $awp_t(aw, y)$ is the Markovian updating rule for average wages that approximates Social Security's exact formula for updating an individual's average wage to reflect current earnings, and wp_t summarizes the law of motion for next period's wealth:

$$wp_t(w, aw, y, ss, ssd) = R\left[w + ssb_t(aw, y, ss, ssd) + y - \tau(y, w) - c\right], \tag{9}$$

where R is the return on saving, and $\tau(y, w)$ is the *tax function,* which includes income taxes such as Federal income taxes and Social Security taxes and potentially other types of state/local income and property/wealth taxes (although the latter taxes are not included in this version of the model). The awp_t function was derived from the regression equation (**??**) and is given by

$$awp_t(aw, y) = \exp\left\{\gamma_1 + \gamma_2 \log(y) + \gamma_3 \log(aw) + \gamma_4 t + \gamma_5 t^2 + \sigma^2/2\right\} \tag{10}$$

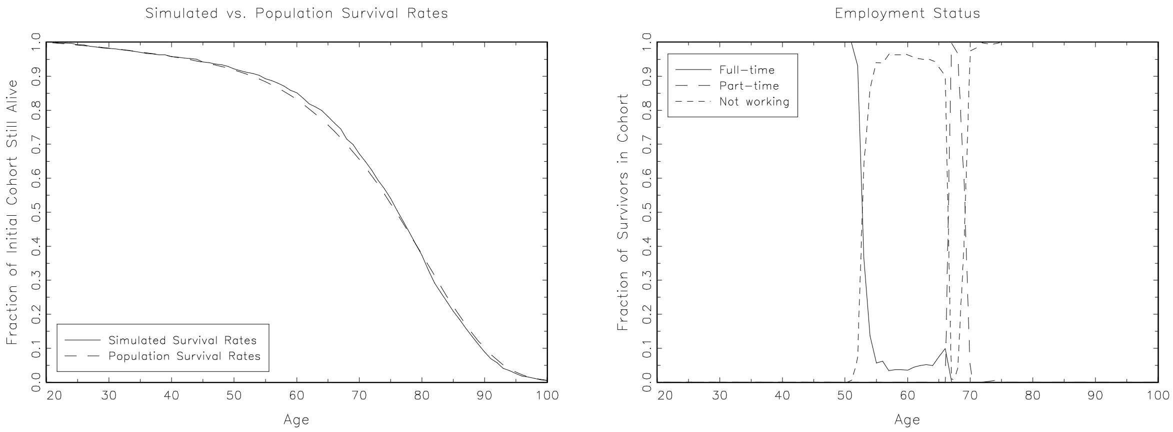
where σ is the estimated standard error in the regression (**??**). Finally, $f_t(y|aw)$ is a log-normal distribution of current earnings given current age t and average wealth aw that is implied by regression equation (**??**) under the additional assumption of normality of errors η_t. The discrete conditional probability distribution $g_t(ss'|aw, w, ss, ssd)$ reflects uncertainty in whether a DI applicant will be awarded benefits or a DI beneficiary will be audited and terminated from the roles. We allow this probability to depend on (w, aw) to reflect means-testing that occurs, for example, in the SSI program. Otherwise it reflects the deterministic transitions for OA benefits, e.g. if a person is over the ERA, is not receiving OA, and applies for OA benefits, then they become eligible with probability 1 (i.e. if $ss_t = 0$, $t \geq ERA$ and $ssd_t = 1$ then $ss_{t+1} = 1$ with probability 1).

The DP algorithm in equation (**??**) was programmed in Gauss and C although. Nevertheless, some routines, such as the zbrent nonlinear root finding algorithm, were based on modified code from Press et al. (1992). The zbrent algorithm used was used to solve for optimal consumption (treated as zero to the first order condition using numerical derivatives), and 15 points Gaussian quadrature was used to compute approximations to the integrals. At each time period the explicit optimizations in equation (**??**) were performed over a grid of 375 points in the (w, aw) state space (25 grid points for w and 15 grid points for aw), where w ranging from \$1 to \$1,000,000, and aw ranging from \$3,000 to \$72,600. Two-dimensional interpolation was used to compute approximate values for $V_t(w, aw, ss)$ at (w, aw) points that are not on the predefined grid. The entire DP problem can be solved in less than 5 minutes on a laptop computer (400Mhz AMD K6 processor). Moreover, our software can generate 1,000 *IID* simulations of the optimal solution in under 5 seconds. We expect to obtain further speedups via use of more efficient numerical methods (such as polynomial approximations to the value functions) and parallel processing techniques. It worth emphasizing that our initial experience make us confident that it is computationally feasible for us to estimate the types of models we are proposing.

Figures 5.2 to 5.4 illustrate the rich types of behavior that the DP model predicts. Each of the curves is an average of 1,000 *IID* simulations, with each simulation corresponding to a separate "person" followed from age 21 until their death. The left hand panel of Figure 5.2 compares the simulated versus population survival rates, where the latter are those implied by the age-specific mortality rates from the U.S. Decennial life tables. We see that the average of the simulations is very close to the population values, which provides a useful check of the validity of the simulations. The remaining curves in figures 5.2 to 5.4 present "mean paths" for various variables of interest, where we computed the average at each age for the subpopulation of survivors who lived until at least that age.

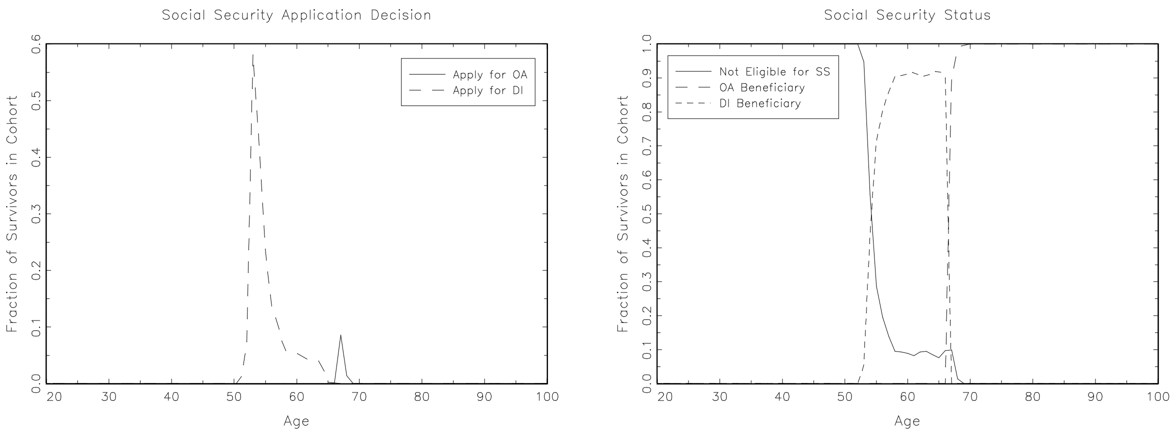
The right hand panel of figure 5.2 shows the employment status of the sample as a function of age. Between ages 21 and 50 virtually 100% of the sample works full time. Then at age 51 there is a precipitous drop in labor force participation until by age 60 only about 5% of the surviving subpopulation is still working full-time. Then at age 67 (the NRA), 99.7% of the simulation sample begins working part-time until age 73 when the remaining sample of survivors stops working for the rest of their lives.

Figure 5.2: Mean Paths from 1,000 *IID* Simulations of the DP Model



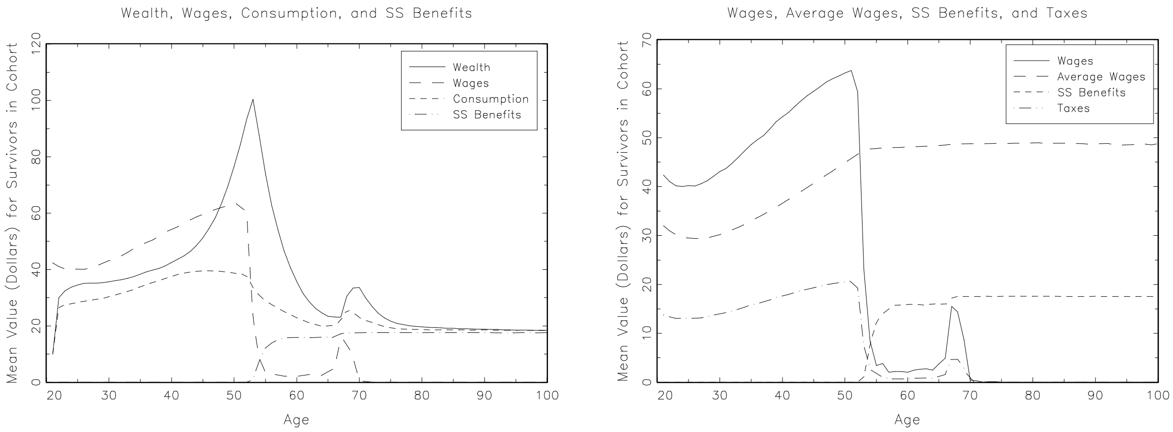
The left hand panel of figure 5.3 provides some insight into what at first glance might appear to be a very puzzling pattern of employment behavior implied by the DP model. The graph shows the Social Security decisions made by the members of the simulation sample. We see that starting at age 50 there is a very rapid increase in the number of people applying for DI benefits, peaking at 54.65% of the surviving sample at age 53, and falling off sharply thereafter until reaching 0 at age 65. The right hand panel of figure 5.3 shows that as a result of the burst in DI applications between age 50 and 65, 91% of this sample succeeded in being awarded DI benefits by the time they reached the Social Security ERA. The remaining 9% who did not apply for or were not awarded DI benefits applied for OA benefits at the NRA, 67. After that, with the automatic conversion of DI beneficiaries to the OA program, virtually 100% of the surviving sample were in the OA program for the remainder of their lives.

Figure 5.3: Predicted Paths for Social Security Decisions and States



The fact that virtually all of the simulation sample decided to work part time at age 67 may also appear puzzling, but the left hand panel of figure 5.4 provides some insight into why this is happening. The figure shows trajectories for wages, Social Security benefits, consumption and wealth over the life-cycle. Starting with wages, we see that wages increased from their initial value of $40,000 until peaking at $63,258 at age 51. During this first 30 years individuals consumed only 70% of their wage earnings, resulting in a rapid buildup of net worth that peaked at a value of nearly $100,000 at age 53. At this point a large number of the sample were quitting their full-time jobs and applying for DI benefits. However, since the average Social Security benefit is only about $15,000 and members of the sample kept consuming at a level considerably above $15,000 even while they were receiving DI benefits, the members of this sample began rapidly decumulating their savings until the NRA when average remaining wealth, $23,239, was only slightly higher than average consumption expenditures, $21,720. Thus, the explanation for the sudden rise in part-time employment can be explained by a need to replenish wealth stocks in order to have sufficient precautionary balances for the remainder of the life-cycle. The other explanation for the rise in part-time employment is that at age 67, the earnings test threshold increases from $10,800 to $17,000, but average part-time wage earnings are only $15,000. Thus, these individuals were able to work part-time without subjecting themselves to the additional 33% earnings surtax. This brief stint of part-time post-retirement work enabled the sample to rebuild wealth stocks to nearly $34,000 by age 70. Afterwards, consumption and wealth both declined to a level equal to the average Social Security OA benefit, which is slightly higher than $17,000 for the surviving members in this age range.

Figure 5.4: Earnings, Taxes, Social Security Benefits, Consumption, and Wealth



The right hand panel of figure 5.4 shows the evolution of average wages and tax payments. Average wages start out at \$32,000 and steadily climb to \$48,000 by age 60, reflecting gradual human capital accumulation over the life-cycle. Tax payments increase from \$13,792 at age 21 to a maximum of \$20,540 at age 51, the point at which most of the members of this population begin applying for DI benefits. We see that about this same time, payment of Social Security benefits starts increasing rapidly as more and more of the simulation sample succeeds in getting awarded DI benefits. It is straightforward to compare the expected discounted sum of tax payments to expected discounted sum of benefits: for this sample the internal rate of return on Social Security contributions is 4.2% if only the individual's tax contributions are counted, and 2.2% if we include the employer's matching contribution.

We conclude with some observations about this example. We have simulated a population of 1,000 individuals who are essentially identical, except for idiosyncratic stochastic variations in realized wages and Social Security DI awards and audits. Thus, although there is still a great deal of heterogeneity in individuals' simulated labor supply, wage, consumption, and wealth paths, it would be easy for us to add additional sources of heterogeneity into the DP model. This would allow us to simulate a wide range of alternative types of behavior, such as individuals who choose to work full time well into their 70s and 80s before retiring, as well as individuals who never work a significant amount over their entire life-cycle. Our results were "concocted" to emphasize opportunistic behavior in applying for DI benefits. Here we have a sample of individuals, none of whom is ever truly disabled. Instead, their disutility for working is increasing very slowly with age. With full knowledge of the Social Security rules, and betting on the odds of getting accepted into the DI program, this sample of individuals makes unabashed use of the DI program as a type of special, stochastically awarded, "early retirement" benefits. The fact that these individuals are not "truly disabled" is especially transparent from the fact that once they reach NRA and their DI benefits are automatically converted to OA benefits, nearly all of these individuals voluntarily choose to return to work on a part-time basis. They do not choose to do this prior to the NRA (even though they have only \$23,000 of their \$100,000 accumulation of wealth at age 53) because they know that if they start working before the NRA they will fail the DI test for substantial gainful activity and consequently their benefits will be terminated. Note also, that their health, as proxied by their disutility for work, is worse in their 60s than in their 50s, yet they were receiving DI benefits during their 50s and able to return to work in their 60s.

We do not present this example to suggest that all, or even a majority of real DI beneficiaries are behaving opportunistically as the agents in our simulated sample are behaving. However, there are a number of features of disability that do suggest the possibility of opportunistic behavior on the part of a significant fraction of DI applicants and beneficiaries, including the fact noted above that the propensity to apply for DI benefits declines rapidly as individuals approach the ERA. This suggests to us that rational maximizing behavior with full knowledge of the Social Security benefit rules may, in fact, provide a good description of the behavior of many of the people who apply for DI benefits. Our purpose is not to pass judgement on whether opportunistic behavior is necessarily good or bad, but rather to determine to what extent it occurs and to consider whether there are alternative rules for paying out Social Security benefits that might lead to higher welfare but at lower cost than under the current system. In particular, our results could suggest that it might be more fruitful to view DI and SSI as types of welfare or early retirement programs. Our model could be used to determine if the potential welfare gains provided by operating a "monitoring technology" designed to assess who is "truly disabled" (i.e. the huge bureaucracy devoted to judging DI and SSI applications and handling appeals, audits and so forth), might be outweighed by the high cost of operating this bureaucracy, especially if the "signals" it generates are very noisy.

Before we can start to answer these interesting policy questions we need to find an "empirically realistic" specification of the DP model, including potential unobserved heterogeneity in various parameters such as parameters governing earnings potential, preferences for leisure vs. work, risk aversion, subjective discount factors and mortality rates, etc. An "empirically realistic" model is one that would not be rejected by standard specification and goodness of fit tests. Informally, it is a model that could pass the "Turing test": i.e. stochastic simulations of the DP model would appear statistically indistinguishable from real data such as we have in the HRS/AHEAD. We have considerable room for experimentation to provide empirically realistic DP models and are very optimistic about the prospects of developing a satisfactory model. However, with this degree of freedom there is always a concern about "overfitting", i.e. of selecting particular functional forms that enable us to fit the data well in-sample. The concern is that such "specification searching" will not result in a model that forecasts well out-of-sample. As we will discuss below, we plan to subject our model to one of the most demanding tests by using the DP model to predict the effect of several significant changes in the rules for payment of Social Security SSI and DI benefits, and comparing how closely the DP model's predictions correspond to the responses of human subjects who will be participating in demonstration experiments conducted by SSA as mandated by the 1999 Ticket to Work and Work Incentives Act. This is demanding not only because we will be working with a completely different sample of individuals than those in the HRS/AHEAD, but also because the subjects in the demonstration projects will be facing a completely different set of policies than the HRS/AHEAD subjects were facing and which were used to estimate the unknown parameters of the DP model. Our predictions of the effect of policy changes will be made prior to the experimental results being known and will thus represent an honest and demanding test of the accuracy of the DP approach to policy forecasting.

6 Econometric Implementation

We briefly describe how our DP models will be estimated and tested econometrically. We partition an individual's state at time t, s_t, into *observed* and *unobserved* components $s_t = (x_t, \epsilon_t)$. The individual observes the full state s_t but the econometrician only observes x_t, and the remaining components ϵ_t are *unobserved state variables.* We account for unobserved states not only because it is more realistic to do so, but also because a model that has no unobserved state variables would be *statistically degenerate,* i.e. the DP model would predict that certain observed combinations of states and decisions would have zero likelihood of occurring.

Assume for the moment that all choices are discrete, i.e. the individual chooses a decision d_t from a finite set of alternatives $C_t(x_t)$. In this case there is a well-developed econometric approach (summarized in Rust, 1994) for incorporating unobserved state variables into the DP model in a tractable fashion, resulting in a non-degenerate statistical model with an associated likelihood function that enables estimation of the unknown parameters of the DP model. Let the number of components of ϵ_t be the same as the number of choices in the choice set $C(x_t)$. Then $\epsilon_t(d)$ can be interpreted as a component of utility that depends on the decision d and other unobserved states of the agent. For reasons of computational tractability, we impose the assumption that x_t and ϵ_t are *conditionally independent*

$$p(x_{t+1}, \epsilon_{t+1}|x_t, \epsilon_t, d_t) = p(x_{t+1}|x_t, d_t)q(\epsilon_{t+1}). \tag{11}$$

We also assume that $\{\epsilon_t\}$ is an *IID* Type III extreme value process. Under these assumptions Rust (1994) showed that the DP recursions take the following form:

$$V_t(x, \epsilon) = \max_{d \in C(x)} \left[u_t(x, d) + \epsilon(d) + \beta \int_{\epsilon'} \int_{x'} V_{t+1}(x', \epsilon') p_t(dx'|x, d) q(d\epsilon') \right]. \tag{12}$$

We can rewrite equation (**??**) as follows:

$$V_t(x, \epsilon) = \max_{d \in C(x)} \left[V_t(x, d) + \epsilon(d) \right], \tag{13}$$

where

$$V_t(x, d) = u_t(x, d) + \beta \int_{x'} \sigma \log \left[\sum_{d' \in C(x')} \exp\left\{ \frac{V_{t+1}(x', d')}{\sigma} \right\} \right] p_t(dx'|x, d), \tag{14}$$

with the terminal condition $V_T(x, d) = u_T(x, d)$, where T is the maximum life span.

Equation (**??**) constitutes the basic recursion equation that we will be calculating in order to solve the DP problem with unobserved state variables. The value functions $\{V_t(x, d)\}$ resulting from these recursions imply a corresponding sequence of *conditional choice probabilities* $P_t(d|x)$ given by

$$\begin{aligned} P_t(d|x) &= Pr\left\{ d = \text{argmax}_{\,d' \in C(x)} \left[V_t(x, d') + \epsilon(d') \right] \right\} \\ &= \frac{\exp\{V_t(x, d)/\sigma\}}{\sum_{d' \in C(x)} \exp\{V_t(x, d')/\sigma\}}. \end{aligned} \tag{15}$$

The conditional choice probabilities can be used to form a likelihood function for the data. Let $\theta = (\theta_1, \theta_2)$ denote a vector of unknown parameters, where θ_2 are the unknown parameters of the transition probabilities (beliefs) $\{p_t(x_t|x_{t-1}, d_{t-1}, \theta_2)\}$ and θ_1 are the unknown parameters in individuals' preferences $\{u_t(s_t, d_t, \theta_1)\}$ (the discount factor β could also be one of the components of θ_1). The full likelihood function for an (unbalanced) panel where individual i is followed for periods $t = 1, \ldots, T_i$ is given by

$$L(\{x_{it}, d_{it}\}_{t=1}^{T_i}, i = 1, \ldots, I|\theta) = \prod_{i=1}^{I} \prod_{t=1}^{T_i} P_t(d_{it}|x_{it}, \theta) p_t(x_{it}|x_{it-1}, d_{it-1}, \theta_2). \tag{16}$$

The HRS is conducted at (approximate) two year intervals whereas the time interval in the DP model is annual — and in some of the specifications we will be considering — monthly. We briefly indicate how we can account for the difference in time scales econometrically. Consider first the annual DP specification. In this case assume that we observe individuals at even time periods, $t = 0, 2, 4, \ldots, 2T$ whereas the DP model yields predictions at all even and odd time periods between

$t = 0$ and $t = 2T$. In this case it is feasible to "integrate out" the unobserved $(x_i t, d_i t)$ observations at the odd time periods. The full likelihood will still be a product of transition probabilities similar to equation (**??**), but it will be a product of two year transition probabilities, $\rho(d_{t+2}, x_{t+2}|d_t, x_t)$. These can be computed via the *Chapman-Kolmogorov equation:*

$$\rho(d_{t+2}, x_{t+2}|d_t, x_t) = \int_{x_{t+1}} \sum_{d_{t+1}} P(d_{t+1}|x_{t+2}) p(x_{t+2}|x_{t+1}, d_{t+1}) P(d_{t+1}|x_{t+1}) p(x_{t+1}|x_t, d_t). \tag{17}$$

The corresponding "Chapman-Kolmogorov maximum likelihood estimator" will involve a greater computational burden than a standard ML estimator when the individual is observed at all time periods, but we believe it is computationally feasible to implement. In the case of a monthly model, the computations become infeasible using standard numerical methods since we would need to integrate over the unobserved state and control variables for the 23 intervening months between successive binannual surveys. In this case we would resort to simulation estimation methods. We do not have space to describe this approach in detail but the essence of the idea can be described quite simply. Even though we only observe individuals at two year intervals, we can still form a large number of sample moments from the HRS, such as mean wealth at each survey, the fraction of individuals working full-time, part-time and not at all, the fraction of each age group receiving OA or DI benefits, etc. For each trial value θ we can very rapidly generate thousands of independent simulations of the DP model (corresponding to thousands of simulated individuals, similar to our illustrative simulation in section 5). Then we can discard the simulated observations for the intervening months that we do not observe in the HRS, and using the remaining data to form a vector of simulated moments corresponding to the moments for the variables we observe in the HRS. Then we iterative update θ to find vector $\hat{\theta}$ that best fits the observed moments in the sense of minimizing a quadratic form between the simulated and observed moments. Under very general conditions, the resulting *simulated minimum distance estimator* (SMD) can be shown to be consistent and asymptotically normally distributed, although it will generally be less efficient than the maximum likelihood estimator. We believe that the SMD estimation is computationally feasible and will make it possible for us to formulate and estimate DP models on annual, quarterly and monthly time scales.

We conclude this setion with a brief discussion of two other econometric issues: how to account for continuous control variables such as consumption, and how to account for unobserved heterogeneity.

In our model the Social Security application decision *ssd* is naturally modeled as discrete, and we also treat the individual's labor supply decision as discrete (e.g. not working, part-time, full-time). However our DP model also includes consumption, which is more naturally modelled as a continuous control variable. Furthermore, we do not directly observe consumption in the HRS/AHEAD data. Our solution to this problem is to observe that consumption does not enter directly as an argument in any of the value functions that determine the likelihood function (**??**). Instead, it is "substituted out" and optimal consumption and the associated value functions can be written as functions of observed state variables such as wealth, average wages, Social Security status, health status, age, etc. The DP model's prediction for optimal consumption does affect the evolution of net worth, and thus enters indirectly into the corresponding component of $p_t(x_t|x_{t-1}, d_{t-1}, \theta_2)$, but as long as there is sufficient flexibility in the distribution of wages and returns to savings, the DP model will imply a positive density for any positive value of net worth. Thus, even though consumption is not observed, the parameters in the utility function specifying the individual's preferences for consumption can be identified due to the fact that changes in these parameters affect the DP model's predictions for labor supply and wealth accumulation, both of which are observed and enter into the likelihood function (**??**). We also plan to develop DP models that allow for limited amounts of borrowing. This will enable us to deal with the fact that some individuals have negative net worth. Allowing for the possibility of borrowing will require us to impose bankruptcy constraints, an issue that we do not have space to describe here, but an important side benefit is that it will allow us to account for the possibility that uninsured medical expenses might exceed an individual's net worth.

Although the DP model allows for a great degree of heterogeneity in potential outcomes due to our rich specification of observed and unobserved state variables, we also plan to investigate the possibility that there are additional sources of *unobserved heterogeneity* entering into individual's preferences and beliefs. Unobserved heterogeneity can be viewed as the limiting case of an unobserved state variable that does not not vary over time. In such a case it can be regarded as a time-invariant characteristic or *type* of the individual. If we denote the vector of types by the symbol τ, we obtain a DP with preferences given by $\{u_t(x_t, d_t, \theta_1, \tau)\}$ and beliefs given by $\{p_t(x_t|x_{t-1}, d_{t-1}, \theta_2, \tau)\}$. For any given value of τ, we can solve the recursions given above and obtain a corresponding likelihood function $L(\{x_{it}, d_{it}\}_{t=1}^{T_i}, i = 1, \ldots, I|\theta, \tau)$. To obtain the overall likelihood, we must "integrate out" with respect to the distribution of the unobserved types. We will use the semi-parametric approach of Heckman and Singer (1984) and approximate the distribution of types by a discrete distribution. If $p(\tau|\theta_3)$ denotes the probability that an individual is of type τ, where θ_3 is a vector of parameters specifying these probabilities, then we can account for unobserved heterogeneity by maximizing the likeihood function given by

$$L(\{x_{it}, d_{it}\}_{t=1}^{T_i}, i = 1, \ldots, I|\theta_1, \theta_2, \theta_3) = \sum_{\tau} L(\{x_{it}, d_{it}\}_{t=1}^{T_i}, i = 1, \ldots, I|\theta_1, \theta_2, \tau)p(\tau|\theta_3) \tag{18}$$

Of course, accounting for unobserved heterogeneity imposes a computational cost since we must solve a separate set of DP recursions (**??**) for each type τ. However in most practical applications of the Heckman-Singer approach, only a small number of types (typically only 3 or 4) are required to "soak up" any residual unobserved heterogeneity in the sample, i.e. the additional improvement in the likelihood from adding more than 3 or 4 types is typically negligible. In the case of SMD estimation, allowing for unobserved heterogeneity is straightforward. We simply simulate over $\{x_{it}, d_{it}\}$ for different types τ chosen at random according to the probability $p(\tau|\theta_3)$. By matching simulated moments to observed moments, the SMD estimator will search for a best fitting specification of unobserved heterogeneity in the same way that it searches for best-fitting values of other parameters of the model. Once again, the main additional computational burden is the need to solve separate versions of the DP model for each type τ. However this is a type of calculation that is highly amenable to efficient solution via parallel processing techniques.

We believe the approach we have outlined for estimating the DP model is computationally feasible and, in many previous empirical applications, it has resulted in reasonable parameters estimates. That is, the maximum likelihood estimation algorithm succeeds in finding a "best fitting" DP model whose stochastic simulations are typically very difficult to distinguish from the real data. However this represents only the first step in estimating and testing a candidate specification. The more demanding next step is to see if the estimated DP model yields accurate predictions of how individuals respond to changes in policy. A key part of our research plan is to subject our models to rigorous specification and goodness of fit tests, including tests of the ability of these models to predict actual responses to policy changes out-of-sample. One interesting set of out-of-sample predictive tests will be conducted by John Rust, who is an expert advisor to SSA's demonstration project for the 1999 "Ticket to Work and Work Incentives Improvement Act" (PL 106-70). This legislation is designed to create increased incentives for DI and SSI beneficiaries to return to work and to ultimately leave the roles via a variety of policies, including 1) altering the 24 month waiting period for Medicare benefits, 2) lengthening the trial work period, 3) gradually reducing benefits paid to workers similar to the OA "earnings test" rather than cutting off benefits entirely, and 4) expanding the use of job training, vocational rehabilitation, and placement services. The legislation requires SSA to conduct controlled experiments with human subjects to gauge which of these policies (or combination thereof) would be most effective in encouraging DI beneficiaries to return the work and ultimately leave the roles. This sort of experiment provides an ideal opportunity to test the predictions of our DP model: for each proposed policy change we can solve the DP model and generate predictions of how individuals will react to it. Then we can compare these predictions to the outcome of the demonstration experiments to see whether the responses of the humans are sufficiently close to the predictions of the DP model.

If the DP model fits well in-sample (using the HRS/AHEAD data) and is able to accurately predict individuals' response to policy changes out-of-sample (and as discussed above, there is strong theoretical reason to believe that a structural DP is the most promising approach for predicting and analyzing policy changes), then it should be clear that our model could have a wide array of practical applications including evaluating the fiscal, welfare, and distributional implications of alternative proposals for reforming various parts of the U.S. Social Security system. Below we provide an example of how our model can be used to evaluate even fairly "radical" policy changes including partial or complete "privatization" of the Social Security system, as well as less radical policies such as the recent elimination of the earnings test for individuals over the NRA.

7 Policy Analysis and Forecasting using DP Models

Policy forecasting using DP models is conceptually straightforward. We can represent a *policy* by the government abstractly by the symbol π. This is a vector of parameters specifying tax rates, ages of early and normal retirement, disability award rates, audit rates, benefit levels and so forth. The DP model will be estimated by maximum likelihood, using observations on the states and decisions of a sample of HRS/AHEAD respondents over a period when Social Security policy is assumed to be fixed at a given *status quo* value which we denote by π_{sq}. Now consider a new alternative policy π. We can use the DP model to predict how individuals will react to, and be affected by the new policy π by re-solving the DP problem, but with the new policy π in effect instead of the *status quo* policy π_{sq}. The new decision rules $\{f_t(s_t, \pi)\}$ constitute the DP model's prediction of the behavioral response, and the value functions $\{V_t(s_t, \pi)\}$ can be used to assess the welfare and distributional effects of the policy change.

We provide several concrete examples of how DP models can be used for policy forecasting and analysis. We emphasize that these results are intended only as illustrations and are not meant to be empirically realistic. Both of the examples are taken from the Ph.D. thesis of Benítez-Silva (2000). The examples also show how we can incorporate private pensions and annuities into the DP model.

Figure 6.1 illustrates how DP models can be used to analyze "radical" policy changes. It compares how individuals behave in two different scenarios: 1) a world in which there is no government Social Security program (i.e. individuals save for their own retirement), 2) a world where there is a Social Security program similar to the one currently in effect, except that we initially assume that the earnings test is in effect for individuals over 65 (as noted previously, recent legislation has repealed the earnings test for individuals over 65). The left panel of figure 6.1 shows that the introduction Social Security leads to a dramatic drop in labor supply after age 65. The right hand panel shows that Social Security also leads to correspondingly dramatic declines in wealth accumulation over most of the life-cycle. The left hand pandel also provides an example of how the DP model can be used to analyze less radical policy changes. It shows the DP model's forecast of how individuals will react to the repeal of the earnings test: the model predicts that it will result in a significant increase in labor supply between ages 65 and 69. Note that the model reflects the fact that there already is no earnings test for individuals over 70. Individuals take advantage of this opportunity to return to the labor force and rebuild their stock of wealth. This explains why there is a large increase in labor supply at age 70 when the earnings test is in effect. When the earnings test is repealed individuals are working more during ages 65-69 so there is less incentive to return to the labor force at age 70.

Figure 6.1: Labor Supply and Wealth Effects of S.S.

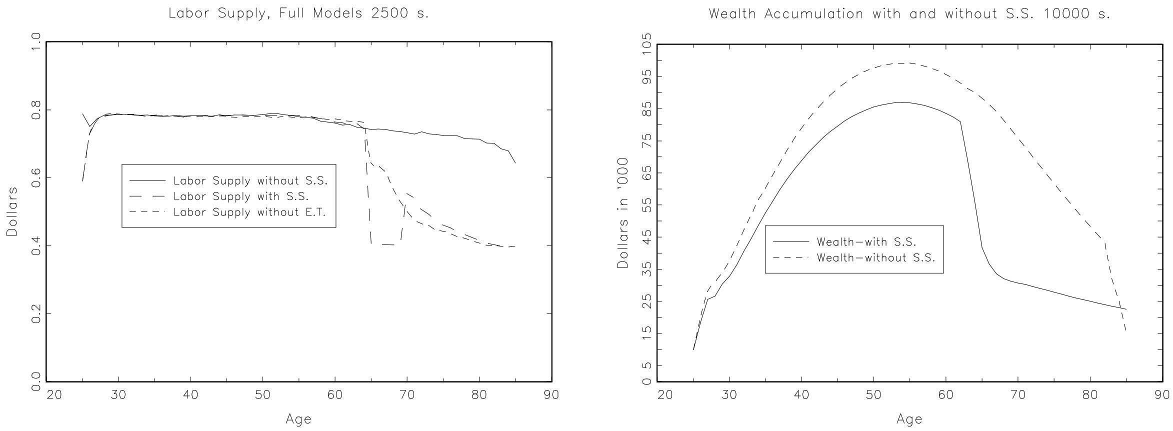
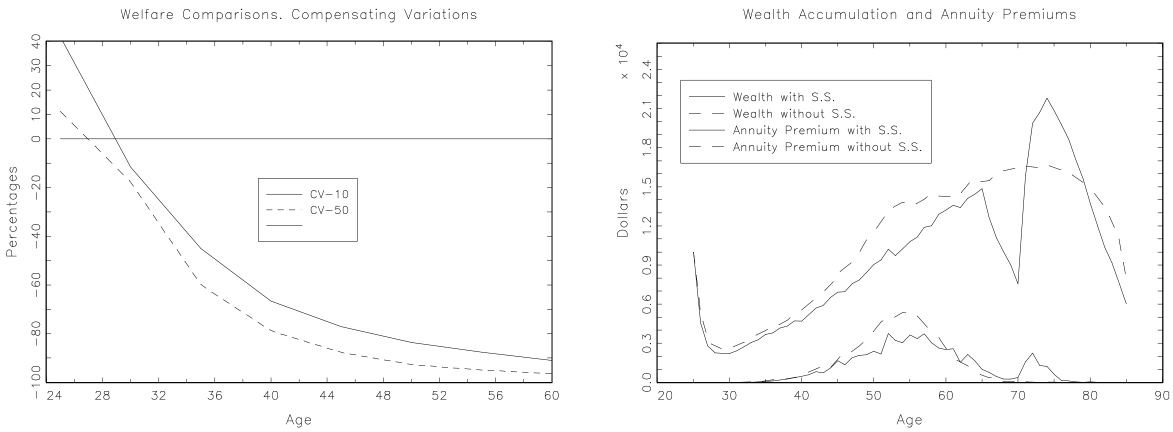


Figure 6.2 illustrates how the DP model can be used to conduct welfare and distributional analyses. For example, we could ask "which individuals would be hurt or helped by a sudden privatization of the Social Security system?" The left panel of figure 6.2 presents the welfare effects for individuals of different ages, measured in terms of compensating variation, i.e. the fraction of their current wealth an individual would have to be compensated if Social Security were abandoned. We see that young individuals are better off without Social Security, but those older than 40 years of age are willing to pay an increasing fraction of their wealth, approaching 100% for those approaching retirement age — to keep the existing Social Security system in place.

The right hand panel of figure 6.2 shows the wealth accumulation profiles of individuals in a version of the DP model that includes an endogenous choice over private annuities. The addition of annuities creates a complicated set of options to individuals: when should they purchase an annuity, and if and when they do, how much of their wealth should they annuitize? The figure shows the paths of the optimal level of wealth and annuity purchases (i.e. the amount of wealth converted into an annuity) with and without Social Security in the model. We see that most annuity purchases are predicted to occur in mid life, just prior to retirement age. However far less than 100% of bequeathable wealth is annuitized. This holds even when annuities are "actuarially fair". Similar to figure 6.1 we see that the introduction of Social Security reduces the level of wealth accumulation. However Social Security also reduces the incentive to purchase annuities. The zig-zag pattern of wealth accumulation under Social Security is a consequence of the earnings test: after indviduals reach age 65 they apply for Social Security and stop working, but maintain consumption at pre-retirement levels, leading to a rapid decumulation of wealth. However after age 70 there is no earnings test, so individuals re-enter the labor force leading to a reaccumulation of wealth until age 75 when most begin to premanently retire. We see that at this time there is a final rapid decumulation of wealth. The presence of annuities tends to exaggerate the changes in wealth: by accumulating sufficient wealth individuals can purchase an annuity to help smooth their consumption for the remainder of their lives. A side result of the right panel of figure 6.2 is that annuity purchases mainly happen late in life and in relatively small amounts, on average. This sheds new light on what it has been called the "annuity puzzle," the question of why the annuity market is so small. These simulations indicate that it may not be necessary for individuals to purchase large amounts of annuities if there is a Social Security system in place, especially if they can insure against residual mortality risk via variation in labor supply. This example also demonstrates how a DP model can incorporate both defined benefit and defined contribution pension plans: in this case Social Security can be regarded as a defined benefit plan and annuities can be regarded as an endogenously chosen defined contribution plan.

Figure 6.2: Welfare Comparisons and the introduction of Annuities.



We conclude this section with a brief discussion of "systematic" approaches to policy forecasting and analysis. We can view policy analysis conceptually as a "Stackelberg game" between the government and individuals. We assume that the government is interested in maximizing welfare of the citizens, but it may weight the welfare of different citizens differently, i.e. it may be interested in redistribution of income or wealth. Given any policy π individuals form their "best-replies," i.e. they maximize their expected discounted utility subject to the assumption that the government will commit to implementing the given policy π. Let $V_\pi(s)$ denote the expected discounted utility of an individual in state s under policy π (for notational simplicity we also let s also incorporate the person's age). If we identify the states s with different individuals (i.e. each individual is in a given state s at the time that the government is determining its policy) we can let $\mu(s)$ denote the population distribution of the different states (individuals). Finally, let $w(s) \geq 0$ denote the weight the government places on an individual in state s. For example, if a component of s is income or wealth, this could reflect the government's relative concerns towards rich vs. poor people. Then a "utilitarian" government would choose a policy π^* that maximizes social welfare, i.e. π^* is the solution to

$$\pi^* = \underset{\pi}{argmax} \int V_\pi(s)w(s)\mu(ds) \quad \text{subject to:} \int c_\pi(s)\mu(ds) = 0. \tag{19}$$

The function $c_\pi(s)$ represents the net discounted cost of the policy for a person in state s, i.e. the expected present discounted value of benefits less tax receipts for a person in state s. If this cost is negative, it means that for person s the present value of their taxes exceeds the present value of the benefits they will receive. The constraint specifies that any policy π must have a balanced budget: this prevents the government from transferring funds from other uses to make individuals arbitrarily well off. It is computationally feasible to solve "social planning problems" such as in equation (**??**), however, the difficulty is how to specify a weighting function $w(s)$ that represents the government's "preferences". An alternative, more ethically neutral approach that does not require the specification of the subjective weighting function $w(s)$ is the *dynamic mechanism design* approach. We discuss a simplified version of this approach here, which we refer to as *computational mechanism design* because we intend to actually implement it in order to characterize efficient Social Security policies.

Let π_{sq} denote the policy of the government under the *status quo.* Then an *efficient policy* π^e is the solution to

$$\pi^e = \underset{\pi}{argmin} \int c_\pi(s)\mu(ds) \quad \text{subject to:} \quad V_\pi(s) \geq V_{\pi_{sq}}(s) \quad \forall s\,. \tag{20}$$

Thus, an efficient policy π^e minimizes the net costs of providing government benefits to individuals subject to the constraint that no individual is worse off under the efficient policy than they are under the *status quo.* We can compute the expected cost of a government program under the *status quo* and compare it to the cost of an efficient policy: the difference represents a dollar valued estimate of the inefficiency in the current system. The current aggregate cost of the U.S. Social Security system is in the range of $10 trillion dollars: this is the expected discounted value of Social Security's unfunded benefit obligations. The most ambitious part of our research will be to attempt to solve the mechanism design problem (**??**) numerically and provide a dollar estimate of the inefficiency in the current Social Security system. To make this computationally feasible, we will use a "parametric" approach to the mechanism design problem whereby we search over a parametrized subclass of policies π rather than the space of all possible policies Π. The latter approach requires much more sophisticated methods including recursive linear programming and the use of the *revelation principle* and imposition of *individual rationality* and *incentive-compatibility* constraints. By looking at a subclass of parametrized policies and explicitly solving for best-replies by dynamic programming, we avoid having to impose the incentive-compatibility constraints, and the individual rationality constraints are satisfied by virtue of the fact that an efficient policy guarantees that nobody is worse off under an efficient policy π^e than under the *status quo.* The main drawback of our approach is that by restricting our search to a parametric subspace of the space Π of all policies, we will only be able to provide a lower bound on the inefficiency of the current Social Security system. Stated differently, there may be alternative policies that we have not considered in our parametric sub-family of policies that could result in even higher cost savings than the policy we determine. Nevertheless we believe that a sufficiently flexibly parametrized set of policies will provide a good approximation to the "non parametric" solution to the mechanism design problem that does not place an *a priori* restriction on the space of possible policies. We believe our research in this area could lead to important theoretical as well as practical insights.

8 References

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

Benítez-Silva, H., Buchinsky, M., Chan, H-M. Rust, J. and S. Sheidvasser (2000): "How Large is the Bias in Self-Reported Disability Status?" under revision for *Review of Economic Studies.*

Benítez-Silva, H. (2000): *Dynamic Life Cycle Models of Labor Supply, Consumption/Saving, Annuities, and Job Search Behavior with Empirical Applications*, Ph.D. Dissertation, Yale University.

French, E. (2000): "The Effects of Health, Wealth, and Wages on Labor Supply and Retirement Behavior," manuscript. Federal Reserve Bank of Chicago.

Gruber, J. and J. Kubik (1997) "Disability Rejection Rates and the Labor Supply of Older Workers" *Journal of Public Economics* **64** 1–23.

Halpern, J. and J.A. Hausman (1986): "A Model of Applications for the Social Security Disability Insurance Program," *Journal of Public Economics*, **31** 131–161.

Heckman, J.J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica,* **46-6**, 931-959.

Heckman, J.J. and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, **52-1**, 271–320.

Horowitz, J.L. and V.G. Spokoiny (1999): "An Adaptive, Rate-optimal Test of a Parametric Model Against a Nonparametric Alternative," manuscript, Department of Economics, University of Iowa.

Hu, J., K. Lahiri, D.R. Vaughan, and B. Wixon (1997): "A Structural Model of Social Security's Disability Determination Process," ORES Working Paper No. 72, Office of Research and Evaluation Statistics, Social Security Administration, 500 E Street SW, Washington, D.C.

Kreider, B. (1999): "Disability Applications: The Role of Measured Limitation on Policy Inferences," manuscript, Department of Economics, University of Virginia.

Lahiri, K., D.R. Vaughan, and B. Wixon (1995): "Modeling SSA's Sequential Disability Determination Process Using Matched SIPP Data," *Social Security Bulletin*, **58-4** 3–42.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery (1992): *Numerical Recipes in C*, Cambridge University Press, Reprint of 1999.

Rupp, K. and D. Stapleton (1995) "Determinants of the Growth in the Social Security Administration's Disability Programs" *Social Security Bulletin* **58** 43–70.

Rupp, K. and C. Scott (1996) "Trends in the Characteristics of DI and SSI Disability Awardees and the Duration of Program Participation" *Social Security Bulletin* **59** 3–21.

Rust, J. (1994): "Structural Estimation of Markov Decision Processes" in R. Engle and D. McFadden (eds.) *Handbook of Econometrics* Volume 4, 3082–3139.

Rust, J. and C. Phelan (1997): "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," *Econometrica*, **65-4** 781–831.

Social Security Administration (1999) "Social Security Disability Insurance Program Worker Experience" Actuarial Study 114, Office of the Chief Actuary, Baltimore, Maryland.

Social Security Advisory Board (1998) "How SSA's Disability Programs Can be Improved" Report 6, Social Security Advisory Board, Washington, D.C. (online copy available at `http://www.ssab.gov/Report6.html`).

Social Security Advisory Board (1999) *Final Report of the 1999 Technical Panel on Assumptions and Methods* Social Security Advisory Board, Washington, D.C. (online copy at `http://www.ssab.gov.Rpt99.pdf`).

Stapleton, D., B. Barnow, K. Coleman, K. Dietrich, and G. Lo (1994): *Labor Markets Conditions, Socioeconomic Factors and the Growth of Applications and Awards for SSDI and SSDI Disability Benefits: Final Report*, Lewin-VHI, Inc. and the Department of Health and Human Services, The Office of the Assistant Secretary for Planning and Evaluation.