Technical Paper Series
Congressional Budget Office
Washington, DC

OVERVIEW OF THE CONGRESSIONAL BUDGET OFFICE LONG-TERM (CBOLT) POLICY SIMULATION MODEL

Josh O'Harra (e-mail: josh.oharra@cbo.gov)
John Sabelhaus (e-mail: john.sabelhaus@cbo.gov)
Michael Simpson (e-mail: michael.simpson@cbo.gov)
Congressional Budget Office
Washington, DC

January 2004
2004-1

Abstract

The Congressional Budget Office Long-Term (CBOLT) policy simulation model was developed to answer budgetary and distributional questions about Social Security, Medicare, and other long-term policy issues. CBOLT has three distinct solution modes for making projections: static simulations with a fixed macro environment and actuarial projection modules like those used by the Social Security Administration (SSA) and Center for Medicare and Medicaid Studies (CMS); a macro growth model environment with SSA/CMS-style actuarial projection modules; and an integrated micro/macro model with economic and policy outcomes based on a representative population sample. The first mode gives answers that approximate those of SSA and CMS actuaries, the second indicates the first-order effects of considering macroeconomic feedbacks, and the third provides the opportunity to consider behavioral responses and conduct detailed distributional analysis. CBOLT can be solved using either fixed or stochastic values for the key exogenous input variables, so the model is capable of generating confidence intervals for budgetary and distributional outputs.

Contents **Page**

1. Introduction[1]

This paper describes the Congressional Budget Office Long-Term (CBOLT) policy simulation model. CBOLT is used to analyze how sensitive Social Security, Medicare, and other long-term projections are to demographic, economic, and behavioral assumptions. The model is designed to examine the effects of various federal budget policy options, including detailed alternative Social Security tax and benefit rules and Medicare eligibility and cost growth assumptions. In addition, it can be used to simulate the effects of trust fund investment, private accounts, and other fundamental policy changes that have been suggested during reform discussions. Standard simulation outputs include macroeconomic and budgetary projections along with various types of distributional tables.

CBOLT integrates several modeling techniques into one tool for analyzing long-term policy issues. The methods employed in a particular CBOLT simulation include various combinations of actuarial algorithms for projecting population and entitlement program finances, a macro growth model with consumption and labor supply feedbacks, a detailed federal budget accounting framework, and micro simulation models that operate on samples of the population. Different combinations of these techniques produce three distinct solution "modes": a static macro environment with actuarial projections like those used by the Social Security Administration (SSA) and Center for Medicare and Medicaid Studies (CMS); a macro growth framework with actuarial projection modules; and a fully integrated dynamic micro/macro growth model that aggregates economic and policy outcomes for a

[1]The authors would like to thank Amy Harris, Kevin Perese, Steve Lieberman, Noah Meyerson, Joel Smith, and Julie Topoleski of CBO's Health and Human Resources Division for their support and helpful comments.

representative sample of the population. Any of these three types of projections can be implemented in either deterministic (fixed-input) or stochastic (Monte Carlo-generated input) simulation environments.

Each of the three CBOLT solution modes plays a role in CBO's analysis of long-term budget policy; the first mode provides answers that approximate the SSA and CMS actuaries' methods. The second can be used to examine the first-order effects of considering macroeconomic feedback from policy changes. The third provides the best data for considering more detailed behavioral responses and implementing detailed distributional analysis. Interrelationships between CBOLT solution modes — modes one and two share the same core actuarial projections modules, and modes two and three share the same macro growth model framework — are useful for disentangling the differences in policy conclusions produced by different models. In particular, by solving the model in the various modes and with various input assumptions, CBO can determine the sensitivity of projections to exogenous assumptions and to modeling techniques.

This paper provides an overview of the capabilities of CBOLT. It is organized around the three solution modes, beginning with a more careful description of how each of the these modes differs and then explaining each solution mode further. The description first includes a discussion of which exogenous inputs and policy parameters the user can set in a given type of simulation and then provides details about the simulation modules themselves. In many instances, although the discussion of how (and why) certain modules operate is surface level only, readers are referred to the detailed analysis of those simulation modules that is available in other CBO papers. The concluding section of this paper provides a brief introduction to CBOLT's stochastic capabilities, which apply to all three solution modes in roughly the same manner.

2. CBOLT's Three Solution Modes

CBOLT has evolved through three distinct phases, each time retaining the simulation capabilities established in previous versions. Its three evolutionary steps have created a static model with a fixed macro environment and actuarial projection modules that mimic the approach used by SSA and CMS in their long term projections, a macro growth model environment with actuarial projection modules and detailed federal budget rules, and a macro growth environment with detailed federal budget rules and an integrated dynamic micro simulation model. This section describes the relationships between the three solution modes. In the most general sense, the first two share many of the same actuarial projection techniques but differ in terms of the macroeconomic environment. The second and third solution modes basically share the same macroeconomic environment but differ in terms of how demographic, economic, and program outcomes are projected.

2.1 Actuarial Projections in the Static Macro Environment

Completed in late 2000, the first version of CBOLT was focused on the Social Security system and designed to generally mimic SSA's Office of the Chief Actuary's (OCACT) methodology. The model replicates some OCACT techniques and calibrates around remaining differences so that it generates a baseline that matches intermediate SSA projections (given the same input assumptions). More importantly, it has sufficient detail to ensure that responses to policy changes or input assumptions are very close to those reported in OCACT's annual *Trustees Report*.

The actuarial modules of this version of CBOLT are "cell-based," meaning that they operate on the U.S. population through detailed age, sex, and marital status groups. In addition, statistics such as population counts, average benefits, the number of eligible beneficiaries, the number of disabled, and the number of workers are associated with each cell. Aggregating cell statistics for each year results in annual totals of revenues and expenditures. (Further manipulation produces statistics such as the combined trust fund balance). The flow of year-by-year cell-based projections is straightforward — first it solves for population by age and sex, then labor force by age and sex, then Social Security covered worker counts, payroll taxes, beneficiary counts, benefit amounts, and finally trust fund outcomes (see Figure 1).

Just as in the SSA actuaries' approach, some exceptions to the cell-based structure in CBOLT actuarial modules are present. For instance, some variables are tracked only in aggregate over time rather than by age and sex in each year. Within that class of aggregate variables, some (such as Social Security cost and income rates, trust fund balances, and so on) are appropriately tracked only as totals since there is no further detail to exploit. To keep the model simple, however, other variables are modeled as aggregate concepts. For example, although separate earnings distributions for each age and sex group could have been included in the model, CBOLT relies instead on a univariate earnings distribution for the entire workforce in each year. The disadvantage of adopting this simplifying assumption is that the actuarial modules do not automatically generate a distribution of payroll tax contributions by age and sex, which is data that would be useful for cohort-level distributional analysis.

Another exception to the cell-based rule in CBOLT's actuarial solution mode is one that adds detail instead of simplification. That exception relates to the computation of OAI and DI worker benefit awards, which are computed using a micro sample of recent beneficiaries. The limited micro simulation is then used within the actuarial framework to capture historical heterogeneity in workers' earnings histories. What results is a model that allows the effect of policy changes to generate a more representative stream of simulated future benefits. The cohort of recent beneficiaries is "re-sampled" in each projection year, with work and earnings histories adjusted to compute future benefit awards that are consistent with other model assumptions.

The static model incorporates a stochastic simulator for nine major demographic and economic input assumptions. This feature allows the model to generate probability distributions as well as point estimates for the actuarial projections. The nine assumptions — the same assumptions the Social Security Trustees vary for their annual low- and high-cost scenarios — are the fertility rate, rate of mortality improvement, number of immigrants, real wage growth, inflation, unemployment rate, real interest rate, disability incidence rate, and disability termination rate. For each input, mean values are set equal to the OCACT intermediate input assumptions specified in the annual *Trustees Report*. Deviations from those means are modeled on the basis of historical variability.[2]

Embedded in all versions of CBOLT is a simple dynamic micro model that simulates outcomes for the same "example" workers analyzed in the report issued by the President's Commission to Strengthen Social Security (PCSSS) and in other SSA distributional analyses.[3] Future earnings of the example workers analyzed are derived by applying fixed age-earnings factors to the projected average wage index, which is a variable in the actuarial model. To solve for benefits, benefit formula parameters are applied directly to predicted example worker earnings histories. An advantage of

---

[2]This version of the model was used in producing, for example, *Uncertainty in Social Security's Long-Term Finances: A Stochastic Analysis* (Congressional Budget Office 2001).

[3]See, for example, President's Commission to Strengthen Social Security (2001) and Social Security Administration Office of the Chief Actuary (2001).

CBOLT's integrated framework is that the stochastic simulator values for variables, such as real wage growth and inflation, that determine the average wage index also act directly on the example workers. Thus, variability in example-worker outcomes occurs automatically. That feature is useful when analyzing both expected outcomes and variability in outcomes for example workers under alternative benefit rules.[4]

2.2 Actuarial Projections in the Growth Model Framework

Completed in mid-2001, the second version of CBOLT integrated the core actuarial modules into a macro growth framework with detailed federal modules, including a detailed Medicare actuarial model developed to mimic the CMS actuaries' approach. It also replaced the baseline macroeconomic assumptions provided by OCACT for the first ten years of the projection with those forecasted by CBO. In this version, the macro model generates endogenous values for wage growth and interest rates that replace the exogenous values specified in static simulations. Through this mechanism, the model can be used to estimate the macroeconomic effects of potential policy changes. The macro predictions are in turn influenced by the actuarial and policy assumptions. Detailed federal budget modules are included, and various tax and spending assumptions are controlled by user specifications across dozens of options about policymaker behavior. Simulation options range from making tax and spending levels constant as a share of GDP to specifying movements in various budgetary parameters

[4]See, for example, Harris, Sabelhaus and Simpson (2003) or Sabelhaus (2003).

to close deficits or stay within GDP-based debt ceilings. This version of the model is used to produce the long-term budget projections in several CBO analyses.[5]

It is instructive to view the growth model solution (mode two) as simply adding more structure around the SSA static model solution (mode one). (See Figure 1). The same steps needed to solve the static model each year are also needed to solve the growth model — in fact, the models share much of the same computer code. Still, the growth model adds several steps that are not included in the static simulations. For instance, although population and labor force are solved for identically in both models, the growth model uses a Cobb-Douglas production function with exogenous total factor productivity growth to solve for total output. Given exogenous price indexes, output is allocated between income to labor and income to capital using standard first-order conditions. Outcomes of this process are real wage growth and interest rates, which replace the exogenous values used in mode-one static solutions.

Solving the growth model requires estimating values for the capital stock because capital (along with labor and exogenous productivity) determines total production. Solving for capital requires first solving for public and private saving, which together determine total investment. Thus, in addition to solving for Social Security budgetary outcomes (as in the static model), the growth framework also computes outcomes for Medicare and other federal revenues and spending, state and local budgets, and foreign transactions, especially capital flows. Those additional modules solve for government and foreign saving, which combine with the private saving decision to determine aggregate investment, and

[5]See, for example, CBO's January 2003 *Budget and Economic Outlook* and December 2003 *Long-Term Budget Outlook*.

thus next-period capital stock. The nature of the private saving decision is a key element of model design that affects dramatically the long-run growth model solution and, therefore, the model allows for analysis based on various user-determined saving responses.

A significant part of the effort involved with adding the macroeconomic environment was introducing a comprehensive federal budget sector, including actuarial Medicare modules that mimic the long run projection techniques used by the CMS actuaries. CMS takes the SSA population projections as given, so those CBOLT modules overlap. Given population, the CMS approach is to project Medicare outlays by category (and across Parts A and B) using age-sex indexes and assumed "excess" cost growth–the amount by which per capita Medicare spending growth exceeds per capita GDP growth.[6] In subsequent work, the Medicare actuarial modules were refined to include a "time until death" factor for spending and more careful tracking of eligibility and participation.

Although mode-two growth model solutions add significant economic structure to mode-one static macro solutions, the ability to run stochastic simulations and operate on example workers is not affected. The only noticeable change for stochastic simulations reflects the change in inputs; rather than set expected values and stochastic processes for real wage growth and real interest rates, the expected level of total factor productivity growth and the expected gap between the interest rate and the return on capital are set exogenously, and then actual values for those inputs vary stochastically. Finally, results for example workers are produced in a mode-two simulation just as they are in a mode-one static macro solution.

---

[6] Like the OASDI actuarial modules the Medicare modules do not attempt to mimic the short run projections produced by the CMS actuaries, focusing only on the long run techniques.

2.3 Dynamic Micro Simulation in the Growth Model Framework

The third version of CBOLT was initially run in Fall 2002, and work refining the micro modules continues as of this writing. This version differs from the second model in that actuarial estimates for crucial variables are replaced with values from a dynamic micro simulation on a 1:1000 representative population sample. For each of the 300,000+ observations in the longitudinal micro sample, CBOLT simulates birth, death, immigration, marital transitions, marital pairings, labor force participation, hours, earnings, Social Security benefit claiming, and Social Security benefit levels.

The dynamic micro model replaces the outputs from many of the actuarial modules, substituting aggregated individual outcomes for group-level (cell-based) projections (see Figure 2). For example, labor supply and earnings are determined person-by-person, rather than group-by-group. Still, much of the growth model structure is the same as in mode two, because, for example, it does not make sense to track aggregate budgetary outcomes at the individual level. As a result, just as modes one and two share much of the same actuarial computer code, modes two and three share many growth model components.

Although details of the micro modules are discussed below, it is worthwhile to note the reasons for replacing the actuarial modules with a dynamic micro simulation while retaining modes one and two. One criticism of the actuarial modules used in solution modes one and two is the lack of detail in certain parts of the model; for example, in solution modes one and two, taxable payroll is not tracked by age and sex. Ultimately, a representative micro sample allows the most flexibility in terms of distributional analysis and the ability to capture micro level behavior.[7]

Although the fully integrated micro/macro environment is in many ways the culmination of the initial CBOLT development effort, some computational costs must be considered. The first two solution modes are computationally intensive, requiring nearly 250 megabytes of RAM and about 90 seconds for a 100 year simulation. The third solution mode, however, is a world apart. A CBOLT run with the dynamic micro model requires about 1.4 gigabytes of RAM (more than was available in most off-the-shelf PCs offered for sale in 2003) and about 20 minutes to run (depending on the processor). This difference becomes particularly important in a stochastic environment; modes one and two can produce 400 runs with random draws stochastically overnight. The fully integrated micro/macro model (mode three) requires about 150 hours of processing time for the same simulation. Even with this amount of processing time, these runs are feasible. By breaking the simulation across multiple machines a run of 400 draws can be (and are) completed over a weekend or during a week of evenings.   Still, the computational intensity of solving the full micro model underscores the utility of preserving several solution modes.

3. CBOLT Static Solutions

In its original conception, CBOLT was designed to mimic, but not necessarily to replicate, SSA projections. To meet that goal, varying levels of detail were included in the model. The idea was

[7]It is important to note here that the dynamic micro simulation in CBOLT is still evolving; some micro behaviors are in the model, but others, such as pension coverage, health status and health spending, and life-cycle consumption behavior, are still under development.

to build a tool that would provide answers similar to SSA's about the budgetary effects of changes in Social Security tax and benefit rules, as well as changes in key demographic and economic input assumptions. This version of CBOLT is referred to as the static version because it does not allow for feedbacks resulting from the federal budget, interest rates, or wage growth.

Many of the projection modules in CBOLT's actuarial (mode one and mode two) simulations are either greatly simplified versions of the SSA actuaries' approach or ad hoc ratio calculations derived directly from their projections. For example, the CBOLT actuarial modules do not project how the number of auxiliary (spouse, children, widow) beneficiaries will change over time or across different policy regimes (though most auxiliary outcomes are projected directly in the dynamic micro modules). Rather, such beneficiary estimates are generated by applying fixed ratios of auxiliary beneficiaries to relevant population subgroups. The ratios themselves are derived from detailed estimates provided by the SSA actuaries. One convenient property of CBOLT is, however, its ability to change auxiliary beneficiary counts according to changes in population mortality, fertility, or immigration — only its ratios to population are fixed. CBOLT's reliance on OCACT population projections underscores the important point that in modes 1 and 2 CBOLT is not an alternative to the SSA actuaries' models; rather, it is dependent on SSA actuaries for annual input files.

3.1 Exogenous Parameter and Policy Settings in CBOLT Static Solutions

A good starting point for describing how CBOLT operates in its static version is to focus on what input variables are required in those simulations (see Figure 3). The list of Social Security policy parameters includes most of the options discussed when policy changes are under consideration: OASI and DI payroll tax rates, benefit formula parameters, normal retirement age, and so on. Some policy parameters, such as tax rates, operate on aggregate values in the static model, others, such as benefit formula parameters, operate on the aged micro sample, thereby capturing the important interaction between heterogeneity in the population and the benefit formula.

The basic demographic and economic input required for CBOLT static solutions is the same as that listed at the front of the annual Social Security *Trustees Report*. The three demographic assumptions (rates of mortality improvement, fertility, and immigration) determine population by age and sex for each year of the projection. The four economic assumptions (inflation, unemployment, real wage growth, and the real interest rate) and the two behavioral assumptions (disability incidence and termination) affect various parts of the model. Some inputs, such as real wage growth, unemployment, and inflation, affect benefit outcomes because they play a role in the aging process for the micro sample used in the "cohort re-sampling" benefit projections described further below. Other inputs only impact aggregate outcomes. For example, the interest rate interacts with aggregate trust fund outcomes to determine interest paid or received by the system.

CBOLT can also simulate many different types of trust fund investment or individual account policies. The model can simulate the effects of investing trust fund assets in equities or corporate bonds controlling for the assumed rate of return on the investments. When the model is solved stochastically, it can project the expected value and uncertainty of those returns.[8] Six sets of parameters go into an individual account simulation; users set contribution rates, participation rules and behavior,

---

[8]CBOLT actually has a number of ways for using historical Ibbotson (2000) data to project returns on financial assets. For a discussion, see Sabelhaus and Smith (2003).

administrative costs and general fund transfer rules, portfolio allocation and annuitization assumptions, and a selection of benefit offset rules. Finally, users can specify behavioral responses to trust fund shortfalls. The options include accumulating debt (often called the "current law" scenario), raising payroll taxes, reducing new benefit awards, or reducing all benefits.

3.2 Actuarial Population Projections

The first step in a CBOLT simulation is the population projection. Population evolves each year through births, deaths, and immigration. In the CBOLT actuarial modules, population is tracked as of January 1st of each year by single year of age (0 to 100), by sex, and by four marital status groups (never married, married, widowed, and divorced). As a result, more than 800 cells are included in the population matrix; 101 for age, multiplied by two for sex, multiplied by four for marital status groups.

The mortality process in CBOLT uses a standard stationary population calculation, with baseline rates of mortality improvement fixed across 20 age and 2 sex groups, consistent with the Social Security actuaries' approach. Users can set both the initial and ultimate rates of mortality improvement by age and sex for the simulation, or set cohort life expectancy by sex in each year. Fertility is also a standard actuarial calculation, with exogenous age-specific annual fertility rates determining the number of births. Immigration is set exogenously in annual levels, and the distribution by age and sex is taken from SSA projections. Marital behavior is the only demographic process in the static solution mode that does not closely approximate the SSA approach. CBOLT actuarial modules make no attempt to model marital distributions; instead, they are set equal to SSA projections for each year. Thus, the distribution of the population across marital states within each age and sex group is invariant to overall population size.

3.3 Labor Force and Employment

As with the population modules, the goal of CBOLT actuarial labor force modules is to mimic the approach of the SSA actuaries. To that end, CBOLT predicts labor force participation for 103 separate groups by age, sex, marital status, and presence of children. The participation rate for each group is based on a time-series equation estimated using March Current Population Survey data for the period from 1968 through 1999. Explanatory variables differ across groups, but they include factors such as percentage in school, percentage in the military, real GDP, disability prevalence, spouse labor force participation rates, number of children, and age- and sex-specific time trends. For the group aged 62 to 70, Social Security policy affects labor force participation insofar as benefits (relative to past earnings) and the Social Security earnings test affect the percentage of people in the labor force.

Applying these estimated labor force participation rates to the population matrix (actually, the "noninstitutional population" matrix, which is fixed by age and sex relative to the main model "SSA Area Population" concept) yields an estimate of labor force by age and sex. Resulting labor force estimates are then multiplied by exogenous unemployment rates (the overall rate is set by the user, and age- and sex-specific rates vary automatically) to solve for the employment matrix by age and sex. Finally, employment numbers are used to estimate the number of workers covered by Social Security in a given year.

3.4 Social Security Payroll Taxes, Beneficiaries, Benefits, and Trust Funds

Social Security modules pick up where labor force and employment leave off. The first step is to apply fixed covered worker rates by age, sex, and year. These rates vary by age and sex for two reasons: first, some age/sex groups are more likely to have uncovered employment; and second, the concepts of covered workers and employment differ systematically because employment is weighted across months of the year, and covered workers is an "any job during the year" concept. Estimates of covered workers by age and sex feed through to other Social Security modules.

The second step is to compute aggregate taxable payroll by estimating the portion of earnings above the taxable maximum using a fixed univariate earnings distribution and taxable maximum policy parameter, aged forward using real wage and covered employment growth. The univariate earnings "look up table" comes from SSA actuaries. The SSA actuaries build in a slight shift in the distribution for the first 10 years, then fix the (real) values for the remainder of the simulation. In this table, as the taxable maximum varies, the underlying total taxable payroll changes. Given taxable payroll, the model multiplies by the (exogenous user-specified) OASI and DI payroll tax rates to solve for aggregate payroll tax contributions during the year.

OAI worker beneficiary counts are computed using insured worker rates, by age and sex, which in principle are based on past covered worker counts, but in practice are nearly 100 percent and constant for eligible age groups. DI worker beneficiary counts by age and sex are based on (exogenous user-specified) DI incidence and termination rates applied to underlying age- and sex-specific probabilities. Counts for dual beneficiaries (those beneficiaries entitled to benefits based both on their own earnings and someone else's earnings) are computed using an SSA formula that predicts such "duals" on the basis of the relative earnings of men and women. Finally, other auxiliary beneficiary counts (CBOLT includes a total of 25 outlay categories) are determined by applying exogenous ratios to population subgroups derived from the main population matrix. For example, widow beneficiaries are a fixed fraction of the number of widows ages 60 and older, where the fraction in each year is computed using SSA actuarial projections.

New OAI and DI worker benefit awards in the actuarial model are computed using a static micro simulation that operates on a sample of beneficiaries taken from the 2001 Continuous Work History Sample (CWHS), the same data set that plays an important role in the CBOLT dynamic micro model. For the actuarial model, CBOLT loads earnings and labor force participation histories of about 15,000 beneficiaries who began claiming OASDI benefits in 2001. For each of the 15,000 beneficiaries in the sample, earnings and labor force participation rates are adjusted to be consistent with the actuarial/macro economic labor force and earnings outcomes that CBOLT generated over those beneficiaries' working lives. For example, new benefit awards for 62-year-olds in 2050 are in principle consistent with the actual labor force participation of 61-year-olds in 2049, 60-year-olds in 2048, and so forth. The technique of reusing the same sample each year to compute new benefit awards is referred to here as "cohort re-sampling" to distinguish it from the techniques used in the integrated micro/macro model (mode 3).[9]

The CBOLT actuarial benefits module uses this micro sample together with the relevant policy parameters to compute new awards. Yet, even with its ability to capture the interaction between

---

[9]For a detailed discussion of the differences between cohort re-sampling and dynamic micro-simulation, see Harris, Sabelhaus, and Simpson (2004).

individual heterogeneity and detailed policy rules, CBOLT's static micro simulation approach has shortcomings; first, for any given cohort of beneficiaries, retrospective micro earnings may not be consistent with the aggregate, or economywide, earnings for the year in question, and, because of the approach, it is impossible to reconcile aggregate and micro earnings (across all working cohorts) in the static model. Second, although distinct shifts in the profiles of earnings of the 2001 beneficiary group by age or by sex have already been observed, the model does not alter its profiles for future cohorts. Both of those problems are resolved in the CBOLT dynamic micro simulations described below.

While the static micro simulation generates new benefit awards, existing OAI and DI worker average benefits by age, sex, and age at entitlement are simply aged forward using the cost of living adjustment and fixed ratios derived from SSA data. Those ratios control for the effects of differential mortality and re-computations on average benefits within a cohort. Auxiliary benefit levels are based on fixed SSA ratios to OAI and DI worker benefits. Thus, those levels move up and down as policy parameters for the new OAI and DI worker benefits change.

Using payroll taxes derived from the univariate earnings distribution and benefits derived by multiplying beneficiary counts by average benefits, CBOLT "trust fund" modules generate the few additional inflows and outflows through the use of simple ratios. These include administrative costs relative to system costs; benefit tax relative to benefit levels; and exogenous general fund transfers and transfers to Railroad Retirement. The exogenous interest rate on new trust fund issues is then used to "grow" the trust funds forward over time, which completes the computations for a given simulation year.

4. CBOLT Macro Growth Model Solutions

One of the simplifying assumptions of the SSA-style (mode one) actuarial projection is that the macroeconomy does not respond to changes in system finances. All of the major economic variables are specified exogenously by the user in a mode-one CBOLT run. In mode two and mode three key economic determinants of OASDI finances are generated using a macro growth framework. The difference in modeling strategy can be thought of as embedding the SSA actuarial model within a comprehensive model of the macroeconomy, one that by necessity considers the behavior of all government and private actors to determine how much the economy grows from one year to the next. In that sense, Social Security affects the macroeconomy directly through its impact on overall federal government deficits, and potentially by inducing behavioral reactions to changes in payroll tax and benefit rules. One crucial component of the comprehensive federal budget sector is Medicare; just as the actuarial modules in the first version of CBOLT were designed to mimic the approach used by the SSA actuaries, the Medicare modules were designed to be consistent with the approach used by the CMS actuaries.

The 10-year model described most recently in the paper *CBO's Method for Estimating Potential Output: An Update* and used by CBO for many years serves as the starting point for the CBOLT macro growth model.[10] Although the CBOLT model uses the same five-sector production decomposition and the same production function and capital input specifications as the 10-year model, it builds on the 10-year framework by adding more detail in the federal budget and by introducing

[10]See Congressional Budget Office (2001).

endogenous labor supply and private saving behavior. Even with those differences, however, the description of the 10-year model is an excellent starting point for readers who wish to learn more about the CBOLT macro growth model environment.

4.1 Exogenous Parameter and Policy Settings in CBOLT Macro Growth Solutions

As with static (mode-one) macro simulations, it is instructive to begin with the list of parameters and exogenous values that are specified in a CBOLT macro growth simulation (see Figure 4). The list of parameters includes most of the options in a static (mode-one) macro simulation, except the real wage differential and the real interest rate. In a mode-two or -three solution, those variables are generated by the production function and associated first-order conditions that determine wage rates and the return on capital. In addition to user settings that overlap with static runs, several options are offered in the economic inputs and other policy rules categories. The "other policy rules" and "behavioral assumptions/parameters" are entirely new.

The sole exogenous determinant of output (real GDP) at any time is total factor productivity. The other two inputs to the production function — total hours worked and the capital stock — are determined by the model itself. Also on the list of economic inputs are several assumptions that determine real wage growth, given values for GDP and labor input. Those inputs include the gap between the consumer price index, CPI-W, and GDP price indexes; the average hours growth rate; and the growth of taxable earnings as a share of total compensation. These real-wage determinants are dealt with explicitly by SSA actuaries in the section of the annual *Trustees Report* that explains their estimate of long-run real wage growth. The CBOLT macro framework also considers three other economic inputs. First, because investment is differentiated by type, the trajectory for computer prices (which have fallen dramatically in recent years) is specified. Second, the capital share of output (the exponent in the Cobb-Douglas production function in the nonfarm business sector) is set exogenously. It should be noted that this feature also affects the historical levels of total factor productivity. Finally, the gap between the 10-year interest rate (which is solved for in the growth model) and the 5-year interest rate (which best approximates the rate on new issues of OASDI trust fund assets) is specified as an exogenous parameter.

Additional "other" policy rules in the macro growth framework generally address non-Social Security government budgeting behavior (see Figure 5). There are a few non-Social Security categories of federal spending that get detailed attention in CBOLT. Medicare is modeled at a very detailed level (more below) while Medicaid and Supplemental Security Income (SSI) have spending indexes that vary by age and sex. Also, the medical spending categories are interacted with exogenous "excess" cost growth parameters that specify how much per person costs outpace per person GDP growth. Thus, as the population ages, spending on those programs will rise relative to GDP automatically, but the rate at which the gap widens (which depends on the "excess" cost growth) is effectively exogenous.

Detail on other federal revenues and spending is fairly minimal. For example, no attempt is made to model federal personal or corporate income tax. CBOLT simply applies an effective tax rate to aggregate taxable income. Many options are available for setting those effective rates, some of which involve endogenous policy responses. Similarly, spending rates are generally fixed percentages of GDP, and various rules set the exact trajectory of spending percentages. Exceptions to fixed spending rules involve a few age- and health-related categories (Medicare, Medicaid, and Supplemental Security Income) for which the age composition of the population and assumed health cost growth affect outcomes.

CBOLT's detailed federal budget modules also allow users to specify numerical debt ceilings and tax or spending responses as those limits are approached. This element is necessary to ensure plausible paths of future economic growth and model sustainability. Users can set debt limits ranging from 0 to 500 percent of GDP, and they can specify whether the government will respond to those limits with tax increases or spending cuts. Debt limits influence behavioral responses (private saving and labor supply) because the labor supply is sensitive to changes in tax rates and because saving is affected by both tax and interest rates (and interest rates are affected by the level of debt in the growth model).

Finally, several behavioral options must be specified in a growth model simulation. Currently, CBOLT includes several types of private savings equations, each differing in terms of how government debt affects private savings (see Figure 6). Some of the savings equations have user-specified parameters that determine the magnitude (elasticity) of responses, generally with respect to changes in the interest rate. The default rule targets the capital-output ratio, which effectively fixes the real interest rate, and thus creates a stable baseline economy such as the one projected by SSA. Also included is an option that allows for a response of labor input (aggregate hours worked) to changes in tax rates. This result implies an important interrelationship between policy and economic activity. If mounting debt leads to a tax increase, labor supply falls, because the substitution effect is assumed to dominate the income effect.

The final behavioral assumption is for the excess cost growth in the component programs of

Medicare and Medicaid; these inputs set gaps between GDP growth and health care costs that determine the extent to which health program costs will outpace growth in GDP. Medicare costs grow with age and sex (offset to some extent because, as life expectancy rises, people have longer "time until death" and thus lower costs in a given age/sex group) which means the aging population will raise program costs relative to the size of the economy. However, based on historical experience, a larger share of the growth in health costs is "excess," meaning that it cannot be explained by simple demographics and GDP growth.

4.2 Production, Incomes, and Factor Prices

The first two steps in the growth model macro solution are the same as those employed in the CBOLT static (mode-one) model macro solution (see Figure 1). That is, in a given year, the model first solves for population, then for labor force and employment, using the same computer code described in modes one and two throughout. CBOLT also includes an hours worked response to changes in personal tax rates. This innovation is one of the two routes (together with private savings, discussed below) by which policy affects economic growth.

Significant differences between mode one and mode two solutions begin after employment is calculated. A static simulation moves directly to solving for Social Security budgetary outcomes. The growth model, however, takes the derived labor input, combines it with capital input (solved for at the end of the previous year) and exogenous total factor productivity, then solves for real output. Once real output and exogenous prices have been determined, the model distributes nominal output through various types of income.

As noted, in a macro growth solution, total labor input (hours worked) is solved for by using the derived employment numbers from the actuarial modules and the assumed average hours growth rate, set as an exogenous parameter. CBOLT allocates total hours across the four productive sectors of the economy. Government — federal, state, and local — hours worked are based on the amount of real purchases in those sectors, where real purchases are determined by budget policy. Given productivity in the government sector, the model solves for the labor input needed to generate the specified level of spending. Hours worked in the farm and household service sectors are relatively small and are set constant relative to overall labor input. The residual, after allocations to other sectors are subtracted, goes to the nonfarm business sector, which accounts for the lion's share of output. The Cobb-Douglas production function in the nonfarm business sector is then used to determine real output in that sector. Output in all other sectors is determined by the hours worked allocation. The sum of real output across the four sectors is total real GDP.

Given aggregate production (real GDP), the basically exogenous GDP price index determines total nominal GDP, which is in turn broken down into income components using the first-order conditions from the production function and other National Income and Product Account (NIPA) accounting identities. The first step is to subtract capital consumption (depreciation) from nominal GDP, a straightforward process with fixed depreciation rates applied to the existing capital stock. The second step is to subtract indirect business taxes, such as excise and property taxes, that are accounted for in total sales (nominal GDP) but not included in calculations of national income. Similar to the hours worked allocation, the level of indirect taxes is a function of government budget assumptions. In the final step before allocating incomes to labor and capital, a statistical discrepancy between production and income measures that involves a significant net addition to output before solving for incomes must be resolved.

Once those subtractions and additions have been accomplished, the growth model has generated the level of aggregate income that will be allocated between labor and capital income. The procedure is simple and consistent with the Cobb-Douglas production function used to identify the initial determinants of output. Total earnings grow with output; capital income is the residual after labor earnings are subtracted. Then, the rate of real wage growth (used in the Social Security actuarial modules) is total earnings divided by hours worked, adjusted for earnings as a share of compensation, average hours growth, and the gap between GDP and CPI-W. The rate of return on capital is the residual capital income divided by the capital stock. As a result, the 10-year government interest rate varies with the derived return on capital, and the real interest rate on new OASDI issues (used in the Social Security actuarial modules) moves one-for-one with the 10-year rate because (as described above) the gap between the two interest rates is an exogenous parameter. Thus, the two exogenous inputs in a static CBOLT simulation (real wage growth and the interest rate on OASDI assets) are now generated endogenously by the growth model.

Referring back to the schematic that details differences between static and growth model solutions (see Figure 1) shows the two solutions coming back together at this point. Indeed, the next step in solving for federal budget outcomes is to run all Social Security modules, exactly as they exist in the static solutions but with a different set of real wage growth and interest rate values because those are now generated by the growth model. In the growth model, however, the entire federal budget, not just the Social Security budget, is important because overall federal taxes, spending, and deficits affect economic growth.

4.3 Medicare

After Social Security, Medicare is modeled with the most detail in CBOLT. As with the original OASDI actuarial modules, the initial goal was to produce projections which look similar to those produced by the responsible actuaries, in this case, CMS. Medicare is modeled at the sub-program level, with individual calculations made for inpatient hospital, skilled nursing facilities, home health care, hospice, outpatient hospital and other Part B services, physician, group plans and prescription drugs. In CBOLT, Medicare outlays can be projected using spending indexes that are fixed by age and sex as CMS does, or, there ia an alternative in which the parameters are further adjusted for "time-until-death" in each age-sex group. The additional detail of time-until-death modeling is needed because as people live longer they can expect to have additional years of healthy, low medical cost living. Modeling without time-until-death will overstate the costs of the Medicare system.[11]

The other similarity between Medicare and Social Security is a detailed tracking of trust fund balances and reconciliation with CBO's ten-year baseline. Thus, the model predicts not only spending on Medicare beneficiaries on a NIPA basis but also tracks trust fund inflows, outflows, and balances for Parts A and B. In late 2003, the model was upgraded to include the new prescription drug benefits and financing.

[11]For details, see Sabelhaus, Simpson, and Topoleski (2003).

4.4 Other Federal Budget Outcomes

Few non-Social Security categories of federal spending receive detailed attention in CBOLT. Medicaid and Supplemental Security Income all have fixed age/sex spending indexes tied to excess cost growth parameters that specify the amount by which per-person costs outpace wage growth. Thus, as the population ages, spending on those programs automatically increases relative to GDP, but the rate at which the gap widens (which depends on excess cost growth) is fixed exogenously.

In CBOLT, most federal budget inflows and outflows are solved for by using simple spending and revenue rules (see Figure 5). These rules have descriptive identifiers such as "balance non-trust fund budget," which means adjust taxes as a share of GDP to meet that target. This type of targeting behavior causes the only true "looping" in a CBOLT run; federal budget modules first compute what would happen with no change in taxes or spending, then re-solve using tax and spending rates that are consistent with whatever target the user sets. Again, those rates are usually simple ratios to GDP, not detailed manipulation of (for example) personal income tax brackets and rates.

The ultimate goal of federal budget calculations (including the initial step when Social Security finances are calculated) is to compute the size of the federal deficit (both on and off budget). CBOLT uses both NIPA and unified accounting to solve for total investment; thus, government deficits (along with net foreign investment and private savings) determine the level of investment in capital for next period production.[12] CBOLT does have rules that maintain national debt (accumulated deficits) below

[12] One of the key model refinements underway is adding more unified budget detail, which will make it possible to show CBOLT results using standard CBO budget table formats.

a specified critical value relative to GDP above which federal debt-ceiling rules mandate tax increases or spending cuts.

4.5 State and Local Budgets

State and local budgets are treated even more simply than the federal budget. Most inflow and outflow calculations are based on fixed ratios of taxes and outlays to GDP. The exception is Medicaid, because states must pick up part of the tab as costs rise. Thus, the state and local spending module works differently than the federal sector module; however, because the goal is to keep state and local deficits a constant fraction of output, taxes are adjusted when Medicaid expenditures change. This convention keeps the focus of the model on federal government behavior and is consistent with state budgeting practices.

4.6 Foreign Transactions

Just as the focus of government modules is on determining deficits, the foreign transactions focus is on international capital flows, especially net foreign investment, which subtracts directly from the pool of other savings used to finance capital investment. Although in principle net foreign investment should be affected by the (endogenous) interest rate in the model, in its current version, CBOLT ratios of capital flows to GDP are effectively exogenous. CBOLT tracks gross and net assets held in the United States and other countries and solves for net exports using NIPA identities on capital flows. That net exports value is a key piece of the NIPA product side- identity in CBOLT, which is the textbook GDP condition — it equals consumption, plus investment, plus government purchases, plus net exports. Therefore, in the current model, the income to foreign capital ratio does not result in higher future foreign investment.

At this point in the model, total output (GDP) is known, government purchases and net exports have been solved for, thus the only computation required to complete the model solution is to divide what remains of GDP between consumption and investment.

4.7 Private Savings

The CBOLT private savings module determines how consumption and investment respond to different policy regimes. It is, therefore, key in determining how the growth rate of the economy will vary under different policy regimes (see Figure 6). After government and foreign transactions modules have run, CBOLT solves for aggregate potential consumption, which is GDP minus government purchases and net exports. Potential consumption is then allocated between actual consumption and investment using a private savings rule. Because it is assumed that consumers will allocate some income between consumption and saving, NIPA income-side identities are used in conjunction with product-side identities in CBOLT. The amount of income that they save is added back to the pool of savings for investment, and it becomes capital in the next period.

Before describing CBOLT's alternative equations, it helps to begin with a quick review of how different savings equations operate. All of CBOLT's savings equations operate on aggregate variables in a myopic framework. Currently, the only dimension along which alternative savings equations differ is in the extent to which private savings offset government deficits.

At one extreme, CBOLT allows a "targeted capital to output ratio" option which effectively neutralizes the impact of government deficits on economic growth — there is no "crowding out." In that case, private savings has a one-for-one response to government deficits. Under this equation, as Social Security and other aging and health-related programs follow trajectories into deficit in future decades, private savings will increase such that the allocation of potential consumption between investment and actual consumption will be unaffected by those deficits. This extreme behavior is often assumed for comparability to SSA projections, because it generates a stable baseline economy with fixed real interest rates.

CBOLT also allows for other alternatives in how government deficits affect private saving where the impact is less than one-for-one and acts through changes in the interest rate. Initially, when government deficits rise, the amount of investment falls because the pool of funds for investment shrinks. However, reduced investment slows growth in the capital stock, which, given the first-order conditions of the production function, raises interest rates. That increase in interest rates triggers savings responses in three of CBOLT's five savings equation options.

The simplest option for linking interest rates and private savings is the "constant savings elasticity," in which users set the responsiveness of the private savings rate to changes in the interest rate. One variant is the "variable savings elasticity," in which the responsiveness of the savings rate changes as the level of the interest rate itself rises — think of this as the economy becoming more "Ricardian" as debt grows and interest rates rise. The third variant operates on the same principle but ties back to the complete offset model above; in the "variable deficit offset" savings model, the percentage of any given deficit offset by private savings varies with the level of the interest rate. At very high rates, the offset is one-for-one (the Ricardian case), while at lower rates the offset is much more modest. In all three of those variants there is some user-defined parameter that determines the degree of savings responsiveness to interest rates and thus the degree to which private savings offsets government deficits.

To some extent, the choice of which savings equation to use (and how to set parameter values on those with parameters) cannot be made independently of assumptions about long-run fiscal policy. If assumptions about effective tax and spending rates caused government debt to skyrocket in the future, the savings equation that was selected would matter tremendously. If, however, budgets were to be balanced every year, the various CBOLT savings equations would predict very similar paths for private savings.

Under the constant savings elasticity model, rapid growth in debt pushes interest rates well out of their historical ranges. This solution highlights whether constant elasticity is a reasonable response under very high-debt conditions. Indeed, under some elasticity values and some fiscal policy assumptions, CBOLT will crash because the underlying economic environment (and/or the specified policy, depending on one's viewpoint) is unrealistic.

4.8 Next Period Capital

At the end of each simulation year, CBOLT solves for total investment, which is the sum of private savings, federal government savings, state and local government savings, and net foreign investment. Total investment is then allocated across several types of capital, and each capital stock grows by the amount of investment minus depreciation. By definition, some types of capital —

residential, for example — do not contribute to output in the nonfarm business sector. Types of capital that affect output are weighted by service flows to determine aggregate capital input for the Cobb-Douglas production function in the next period.

5. CBOLT Integrated Micro/Macro Solutions

With CBOLT, the move from static to macro growth model solutions involved adding economic structure around SSA-style actuarial modules. The innovation introduced in CBOLT's third solution mode is to replace most of the actuarial methods with a dynamic micro-simulation model. That model starts by using longitudinal demographic and economic data for a representative sample of the population as of some base year. In each year of the projection, the model simulates demographic and economic life events for each member of the sample. Ultimately, those micro level events include calculations of payroll tax liabilities and Social Security benefits. Thus, the dynamic model provides an alternative to actuarial modules for computing OASDI and other program finances — directly aggregating outcomes across the representative micro sample.

Dynamic micro-simulation is both conceptually and computationally more difficult than actuarial modeling. A dynamic micro model must go beyond projecting average outcomes across age/sex groups — the technique used in a cell-based model — and capture heterogeneity within each group. Capturing this type of heterogeneity at the individual level involves adding random components to the estimated equations that explain demographic and economic transitions. Thus, the typical dynamic micro module involves looping over the entire sample, computing the probability of transition for each person in the sample, then using a computer-generated random number to determine the actual outcome.

It is easy to see why, when the micro sample is large, the number of computations in a given simulation is much higher than in an actuarial model. If, for instance, the representative sample includes about 300,000 observations, and approximately 10 demographic and economic outcomes are being produced per person, then about 3 million micro calculations must be made per year for each of the 75 years of a given simulation. This increased number of calculations shows up in model solution times; a CBOLT static mode-one or growth model/actuarial (mode-two) simulation takes about 90 seconds; an integrated micro/macro solution (mode three) takes about 20 minutes, but the exact time depends on the clock speed of the computer.

5.1 Integrating a Dynamic Micro Model into the Growth Framework

The best place to begin describing CBOLT's dynamic micro model is with the demographic and economic variables tracked for each person in the representative micro sample (see Figure 7). For each person in the 1:1000 sample, the micro model tracks date of birth (and thus implicitly age), immigration year, sex, education, marital status, marital partner, earnings, labor force participation, Social Security benefits and beneficiary status, and a host of intermediate variables needed to compute transitions. That list of variables is consistent with the goals of CBOLT. Still, it suggests the limitation of replacing the actuarial model with the micro model. For example, because there is currently no link between parents and children in the micro model, children's auxiliary benefits cannot be computed on the basis of their parents' earnings. The list of variables tracked for the micro sample also suggests what sort of transition processes are needed to dynamically age the sample forward each year. On the demographic front, each person in the existing micro sample faces some probability of death, and individuals ages 16 and older face the probability of marital transitions, depending on their current marital status. In this approach, individuals who transition into marriage are matched to potential spouses in a simulated "marriage market." Person-level economic outcomes include labor force participation, hours worked, unemployment spells, and ultimately, earnings. Given demographic and economic variables, a final set of equations computes beneficiary status (claiming behavior) for OAI and DI worker benefits. Longitudinal earnings histories combined with detailed program rules are used to compute benefit levels.

In some cases the dynamic micro and actuarial modules work together to determine outcomes. The most noticeable example of this is the basic demographic processes: births, deaths, and immigration. For these transitions the actuarial modules determine how population evolves — year-by-year and by age and sex — and the micro modules simply allocate outcomes across the representative sample of the population. This feature has the immediate advantage that basic demographic processes are still controlled through the same parameter settings as in mode-one and mode-two simulations. Nonetheless, future development plans for CBOLT include improving how it models basic demographics.

Other than basic demographics, the dynamic micro modules provide a distinct alternative to the actuarial modules. Indeed, when the dynamic micro model runs, most of the actuarial modules used in mode-one and mode-two simulations are turned off. A good example of how this works is illustrated in the labor supply and employment modules. As described above, the actuarial modules — in both static and growth model modes — use 103 time-series equations to project group-level labor force participation. Fixed unemployment rates determine employment by age and sex and in the growth model, multiplication by exogenous average hours determines total hours worked. In the micro model, each person's labor supply is determined through a sequence of equations for labor force participation, hours, and unemployment, and total hours worked is summed directly in the micro sample. From the perspective of aggregate production, the choice between the underlying labor supply modules is irrelevant because the only important variable in the aggregate production function is total hours worked.

5.2 Preparing the CWHS Micro Base File

Because no single data set includes all of the demographic and economic data needed for a CBOLT dynamic micro simulation, a much more complex effort, using several data sets, is required to create a base data file with all of the required information. Moreover, because micro data are generally available only with a lag, CBOLT uses the dynamic micro model to project micro outcomes for the years between when the data set ends and the simulation begins.

One of the first choices in dynamic micro simulation is where to start when developing the base file. CBOLT uses administrative data from SSA's Continuous Wage History Survey — a 1:100 sample of all Social Security numbers ever issued. As of 1998, CWHS included several million observations.[13] Because it is based on administrative data records rather than surveys, it provides the

[13]One of the on-going parts of the CBOLT project is updating the micro base file as the CWHS data set is updated. As of this writing, the model is being updated to the 2001 input file.

best earnings and benefit data available to researchers. In addition, its sample size far exceeds any survey data set. (CBOLT actually reduces the CWHS sample by a factor of ten — to a 1:1000 sample — and even with that reduction, the sample is still orders of magnitude larger than alternatives such as Panel Study of Income Dynamics (PSID) survey data.) Still, the administrative nature of the data set has one serious drawback: it contains only the most basic demographic, labor supply, and earnings data. The CWHS also suffers from missing values for some crucial data, particularly dates of death for people who never received Social Security benefits. Moreover, earnings before 1978 are only reported up to the Social Security taxable maximum, and the reported earnings above the taxable maximum through the first few years of the 1980s are thought to be suspect.

Before its use in CBOLT, CWHS data must undergo a number of procedures to prepare it for use as a base file. The goal is to develop a representative sample of the U.S. population as of the end of 2002, with all the micro variables listed above (see Figure 7). The starting point is a raw CWHS file with limited demographics, earnings, and beneficiary data for a 1:100 sample of all Social Security numbers issued through 1998. The end product is a 1:1000 sample that represents the U.S. population during the 1984 to 1998 period (historical information on earnings back as far as 1951 is retained for individuals alive in 1984). Seven steps are required to complete the transformation:

Step One. The first step in working with the CWHS sorts the individual records randomly. This guarantees that in later steps, the process of sampling from the data or matching mates will be truly random and not reflect the systematic order of the CWHS. The original extracts of micro data appear to be in order of year of SSN assignment, thus the older observations are listed first. In addition, the active and inactive CWHS files are concatenated together in the original extract program, thus observations with observed deaths are listed last.

Step Two. The second step in base-file preparation focuses on creating a representative one percent sample as of the CWHS sample year (currently 1998). As noted, the CWHS does not have accurate dates of death for many people in the sample because the nature of the data collection makes that a priority only for people already receiving benefits (SSA wants to ensure they stop making payments at death). Although CWHS is a 1:100 sample of Social Security numbers, the failure to accurately record deaths leaves a file with observations that total more than one percent of the population in 1998. For the base file to be representative, dates of death must be imputed for the "extra" people, and those people are then removed from the sample. Because the CWHS also does not record immigration status, an individual who immigrates to the United States at age 40 appears as part of the population for the previous 40 years, which also inflates historical population totals. To solve this problem and better target the CWHS to match historical population counts, dates of immigration are imputed. Earnings and beneficiary status are used as indicators of either death or immigration. For instance, a worker who stops earning but does not go on OAI or DI is more likely to have died. And a worker who starts earning abruptly is more likely to be newly immigrated. In years prior to the mid-1980s, when only covered earnings were recorded, these indications could be confused with those of people moving in to or out of uncovered work, which is one reason why the micro base file is developed to be representative only starting in 1984. A demographics program removes excess observations from the CWHS by targeting historical population, worker beneficiary and covered worker counts from SSA. Using micro earnings and beneficiary status, the program targets people ages 15 to 100 to fill age and sex cells, including observations with positive earnings — counted as covered workers — and then observations with OAI or DI worker beneficiary status. Finally, population age and sex cells are filled with individuals who are most likely to still be alive.

Step Three. The third step in preparing the CWHS completes the earnings histories for the population. As noted above, data on earnings above the taxable maximum are not reliable until the mid-1980s. As a result, earnings above that maximum must be imputed for all sample members who were in the work force before 1984. In addition, deferred compensation must be imputed for 1984 to 1998 along with late postings in 1998; although the CWHS was released in 1999, it often takes two years to accurately record all FICA earnings because of "late postings."

Step Four. Once the 15-year representative longitudinal sample with complete earnings histories is complete, the 1:1000 micro sample is selected at random by choosing one out of 10 observations from the full micro sample.

Step Five. The fifth step imputes additional demographic and economic characteristics onto the CWHS sample, including marital status as of 1983 and education. The approach often involves working backward with equations estimated from supplemental data sets. The same statistical relationships used in the dynamic micro model projections (labor force participation depends on marital status, especially for younger and middle-aged women; earnings depends on age, sex, and education status) are reversed when imputing characteristics (younger and middle-aged working women are less likely to be married; higher earners are more likely to have high education). Marital histories back to age 16 are also imputed, including the number of previous marriages.

Step Six. The sixth step is to assortatively match married individuals in the population.

Because the CWHS does not include information on marital status, it also lacks information on spouses. Linking spouses is critical for projecting auxiliary Social Security benefits. Thus, extensive work was undertaken to produce an algorithm that matches mates. The mate-matching project considers all individuals married in 1983 and matches them together on the basis of earnings, age, and imputed education. (This same algorithm is used in all future projections.[14]) Once individuals are matched, it is necessary to align marital histories such that current spouses entered their most recent marriage in the same year. This project takes about an hour to run because it creates more than 56,000 unions. A married individual's record contains a link to the record of the spouse that can, in turn, be used to determine future marital transitions and auxiliary benefit awards and amounts.

Step Seven. The final step is to fill in missing micro data for 1984 through 1998 and then to simulate the entire micro sample forward to the end of 2002. This historical simulation uses much of the same computer code employed in the CBOLT dynamic micro model during the projection years, but differs at points where aggregate outcomes for certain processes are known from other data sources. For example, other data sources provide aggregate information on marital distributions, labor force participation, and hours worked that are then used as targets to calibrate micro processes. Those calibration ratios also have information about how well the micro processes are tracking history. This information about possible systematic biases endemic in the data or techniques helps to resolve "jump-off" problems associated with simulating forward in time. The result is a micro file that has the intended heterogeneity but is also consistent with known aggregates.

---

[14]See Perese (2002).

5.3 Demographic Transitions in the Micro Sample

Person-level demographic variables in the CBOLT dynamic micro model (see Figure 7) include years of birth and immigration, sex, marital status, and (if married) a linked spouse. The basic demographic processes (birth, immigration, death) are simple extensions of the actuarial techniques used in other CBOLT simulation modes. Marital transitions and mate matching are much more complicated, however, and rely on several estimated transition equations as well as a simulated "marriage market" that unites brides and grooms each year.

The basic demographics in a CBOLT dynamic micro simulation are controlled by the corresponding actuarial processes — mortality, fertility, and immigration (see Figure 8). In the case of mortality, the probability that someone in the micro sample dies in the current year is a function of their age and sex and is solved for as the ratio of expected deaths to beginning population. For example, if 1,000 individuals were in a particular age-cohort group in the micro sample at the end of year one, and the actuarial model predicts 990 people in the same age-cohort group at the end of year two, then the probability that someone in the group will die is (1000 - 990)/1000, or 1 percent. Each person in that group gets a random number — distributed uniformly between zero and one — in the mortality module. If that random number is less than .01, that person dies. One improvement is the addition of earnings as a determinant in the mortality probability equation. This controls for observed "differential mortality" in death rates across income while still preserving the number of deaths.[15]

[15]One of the oft-noted determinants of mortality (and other demographic processes to follow) that is missing in CBOLT is race. Although race is a strong determinant of several demographic and economic outcomes, there are several problems with including race, most notably the lack of good racial identifiers on the underlying CWHS data set used in the model.

The result of this process is that the expected number of deaths in a given age-cohort group will always match the actuarial model, although random variation will cause actual outcomes to differ. Still, CBOLT has a built-in solution that prevents random variation from affecting long-run results. If too many or too few people die in a particular year, the model will self-correct in the subsequent year, because the underlying probability of death is generated by comparing the actual age-cohort count in the micro sample with the value in the main population matrix, which is generated by the actuarial model.

Immigration and births are even simpler to calculate. (Note that as of yet, CBOLT does not attempt to link parents and children in the micro model. Consequently, the number of births remains exogenous to that model and is still determined by the actuarial modules.) The only transition code needed for new births is a random number that determines sex. The number of immigrants is also fixed in aggregate, as is the expected distribution by age and sex.

The first set of transition processes that rely on estimated micro transition equations are marital events. As with most of the transition equations, what follows is a brief discussion of those processes, but details can be found in a series of Technical Papers.[16] CBOLT includes four marital status categories: never married, married, divorced, and widowed. Thus, four transitions are modeled: first marriage, divorce, widowing, and remarriage. In the micro model, widowing occurs automatically, through a combination of the mortality probabilities and links between spouses. The other three processes are affected in various ways by age, sex, education, earnings, cohort effects, and marital history up to that point.

[16] In the case of marital transitions, see O'Harra and Sabelhaus (2002).

The 1996 Survey of Income and Program Participation (SIPP) panel data, linked with SSA earnings records, are used to estimate all of the marital transition equations for CBOLT. SIPP's retrospective marriage history module allows for construction of marital histories for each person in the sample ages 16 or older through 1996. Patterns that emerged from those marital histories motivated the strategy for estimating actual transition equations used in CBOLT. The approach CBOLT employs for marital transitions is to estimate separate equations for each age/sex group, where the effects of cohort can be separated from the effects of age and where the impact of other independent variables — earnings, education, marital history — are allowed to vary by age.

The first pattern observed from SIPP marital history data is that distinct differences are evident across cohorts in event probabilities at a given age. For example, the probability of a first marriage before the age of 25 has decreased dramatically over time. The probability of a first marriage after the age of 25 has, however, remained stable or even increased during the same time, such that the cumulative probability of a first marriage by age 40 is only marginally lower for later cohorts. The fact that an important determining variable — simulation year — has different effects on different age groups within the sample suggests that a flexible approach is required when estimating transition equations.

In CBOLT, the key to such flexibility is estimating separate transition equations for each single-age group. The alternative is to use many interactions between variables and potential nonlinear effects for a given variable. That, however, imposes unnecessary structure on the relationships and may actually introduce significant multi-collinearity. One problem with the single-year-of-age approach is that SIPP sample sizes get small, therefore the actual estimations use an "age-centered" strategy that includes data for the age group being analyzed plus groups within two years of age on either side, but weighted slightly less than the group being analyzed. As a result, CBOLT has, for example, separate first-marriage equations for every age/sex group between the ages of 16 and 70. When considering these issues, allowing for the effects of different explanatory variables across groups is important. Consider, for example, the effects of education: educational attainment affects first-marriage probabilities much more at younger ages (when people are still in school) than at older ages.

The other two marital transitions — divorce and remarriage — use a similar age-centered estimation strategy. Again, the value of flexibility is evident. In addition to basic demographics and earnings, equations for these transitions use marriage history to capture important longitudinal differences in marriage patterns; for example, the probability of divorce for people who are already once-divorced is higher than for those who have never been divorced. In addition, findings about divorce and remarriage have implications for OASDI outlays; men with low earnings and women with high earnings are both more likely to divorce, which suggests a self-sufficiency aspect to divorce decisions.

A key decision in micro marriage equations, and in most micro transitions, is how to extend unexplained trends into the future. As noted, the flexible age-centered approach is good at separating age and cohort effects, but it does not provide particular guidance about what to do with unexplained cohort trends. An example is the unexplained decrease in first-marriage probabilities among the young: should CBOLT assume the downward trend will continue further, stop where it is, or reverse? Here the usual CBOLT solution is to assume no further trends beyond identified cohort effects for young age groups and allow those to work their way through future age groups. Interestingly, this assumption generates answers about future marital status distributions that are very similar to SSA actuarial projections.

The final demographic process that is modeled entails solving a complicated algorithm that links brides and grooms.[17] A brief description of the process follows. Each year, a few thousand brides and grooms are simulated to experience CBOLT wedlock either through a first marriage or a remarriage transition. The goal of this CBOLT simulated marriage market is to unite couples in such a way that the joint distribution of husband/wife characteristics in the CBOLT micro world resembles the joint distribution observed in actual data.

The data set used to analyze the joint distribution of husband/wife characteristics is the same matched SIPP, with SSA earnings records used for marital transition equations. In the case of mate matching, the estimation period is restricted to marriages that occurred in 1994, 1995, or 1996. In those years, about 1,200 marriages were observed in SIPP, about two-thirds of which were first marriages (sorted by husband's marriage order); the remainder were second or higher-order marriages.

Characteristics considered in the mate-matching estimation are straightforward; basic demographics such as age difference between husband and wife are key, but determinants such as education and earnings are also important for simulating distributional and program outcomes. In particular, the process of "assortative" mating suggests that couples will tend to have appropriately correlated ages, educational attainments, and earnings, which has strong implications for distributional analysis when programs (such as Social Security) have rules based on both individual and spousal characteristics. The goal for the CBOLT algorithm is to develop a means to simulate the joint

[17]See Perese (2002).

distributions of age, education, and earnings observed in actual data without imposing too much uniformity on the resulting couples. Algorithms that directly impose maximized matches between husbands and wives (by finding each person's most suitable mate) cause too many perfect marriages (for example, husbands who are exactly one year older than their wives).

The CBOLT solution is to estimate logistic equations where the probability of the union between two individuals depends on the indicated variables: age gap, education gap, and earnings gap. The estimation phase involves considering every woman (bride) or man (groom) in SIPP a potential match for every other man or woman, where actual couples get successful outcomes and every other potential couple gets a value of zero. Thus, the estimation is made within a huge data set with few successes and many failures.

CBOLT's marriage market uses estimated probabilities of husband/wife unions in conjunction with random numbers to generate matches whose joint distribution of characteristics is consistent with SIPP data. Each year, the model proceeds groom-by-groom, first computing the probability of union for all potential brides, then proceeding bride-by-bride (where the brides are randomly ordered) using a random number compared with the probability of union to determine whether a match occurs. One of the tricks that makes this approach efficient is that the probability of union for any given potential bride is normalized by the probability of union for the most likely bride; thus, perfect matches happen without randomness (if the groom gets that far in his list) but there is still, for example, a 50 percent chance of union between a groom and a bride who is (statistically) half as likely to marry that groom as his ideal mate. This statistical randomness eliminates the tendency of mate-matching algorithms to produce too many perfect marriages. This mate-matching algorithm successfully reproduces the joint distributions of husband/wife characteristics exhibited in the underlying SIPP data.

5.4 Individual Labor Supply and Earnings

One of the most striking features of individual labor supply and earnings patterns is the tremendous amount of idiosyncratic (unexplained) but highly autocorrelated variation in the data. The extensive unexplained heterogeneity means that standard control variables such as education, marital status, cohort, and age may all have statistically significant effects on labor supply and earnings, but the overall fraction of the variation in labor supply or earnings explained by those determinants is quite low. The autocorrelated nature of the series means that if a person's outcomes diverge from the equation-based expected value in one direction or another in a given year, they are likely to continue to do so in subsequent years.

The sequence of equations CBOLT uses to predict labor force participation, the part-time/full-time decision, and the level of earnings are all structured to accommodate these observations.[18] In participation and hours equations, the probability of working full-time this year is strongly correlated with whether one worked full-time last year. In earnings equations, each person is assumed to have a permanent earnings differential that measures the extent to which their expected earnings differ from expected earnings for someone with their same characteristics. Because of "permanent shocks," that earnings differential evolves over time so that it can capture variations in lifetime earnings patterns observed in the data. Even so, current earnings are highly correlated with lagged earnings.

[18]For more details on these equations, see Harris and Sabelhaus (2002).

CBOLT micro labor force participation equations are estimated separately for men and women using pooled cross-sections from the annual (March) supplement to the Current Population Survey for 1976-2001. This estimation strategy uses the same age-centered technique employed for marital transitions. Use of this approach provides the most flexibility in terms of how independent variables (age, marital status, beneficiary status, cohort, and lagged participation) affect the probability of working. The equations reflect expected correlations (married women are less likely to work, for example) but also capture important cohort effects (younger female cohorts have higher participation in the labor force and are likely to continue to do so as they age).

CBOLT's full-time participation equation is similar to the overall participation equation. Because the Census Bureau's Current Population Survey (CPS) does not provide longitudinal details about hours worked, the full-time equation is estimated using data from the 1968-1992 Panel Survey of Income Dynamics (PSID). Again, the model is estimated separately for men and women, and explanatory variables include age, marital status, cohort, lagged participation, and lagged full-time participation. As with labor force participation, the lagged values introduce the observed autocorrelation (persistence over time) in workforce attachment.

In the CBOLT micro model, unemployment is modeled very simply; because the aggregate unemployment rate is an exogenous variable, the only task is to distribute unemployment events across groups. Unlike the actuarial model, the micro model has a problem allocating unemployment. In the aggregate, a 4 percent unemployment rate for a given group implies that the labor supply should be lowered by 4 percent for that group to generate a measure of labor input. At the micro level, however, this does not imply that 4 percent of the people should be unemployed, because individuals are not usually unemployed for a full year. To accommodate this phenomenon in the micro model, a simple regression that relates the fraction of people experiencing an unemployment spell to the aggregate unemployment rate is run on CPS summary statistics. Thus, for example, a 4 percent unemployment rate is consistent with 7.5 percent of the workforce experiencing unemployment spells of varying lengths in a given year. CBOLT allocates those spells differently across part- and full-time employees on the basis of CPS historical data.

Because of the approach CBOLT takes in estimating earnings equations it is necessary to solve for hours of labor input per person. Under this approach, earnings equations are estimated separately for men and women using CPS data for 1976 through 2001. The dependent variable in those equations is the log of real, full-time equivalent, earnings — that is, what the person would have earned in 1993 dollars if he or she had worked full-time. Before the equations are estimated, each part-time worker's hourly earnings are adjusted upwards to be consistent with the full-time equivalent concept. The historical values are then adjusted for overall nominal wage growth over time, so that the units are consistent over the 25 years used in the estimation.

Creating full-time equivalent earnings before estimating earnings equations is a means for isolating idiosyncratic differences across people. Consider, for example, two 30-year-old females with the same educational attainment who both work full-time for salaries of $30,000 per year. If one of those women was assigned part-time status in the following year, her earnings would be adjusted downward to account for the reduction in hours worked. Moreover, a full-time wage premium would be deducted. The woman who continued to work full-time would keep earning $30,000. If the part-time worker was to return to work full-time, her earnings would be reinstated to the level of her continuously full-time counterpart. Without the full-time equivalent adjustment, such a sudden increase would register as an unexplained earnings "shock" when, in reality, it is completely explained by changes in the labor force decision.

Explanatory variables in male and female full-time equivalent earnings equations include age (single year of age dummies), education, education-age interaction, and cohort trends. These equations produce the typical rising concave patterns of earnings — by age — that are observed in cohort-level data. Overall, however, the explanatory power of the equations is fairly low; only a fraction of variance in earnings is explained by control variables. In a sense, this implies that estimating earnings equations is actually only part of the overall process of predicting earnings. Indeed, modeling the unexplained residuals is at least as important as capturing the explained variation.

The pattern of errors estimated from the earnings equation suggests a useful strategy for predicting individual earnings. First, for a given person, the errors are highly autocorrelated. That is, if a person's earnings are above the equation-predicted value in one year, they are likely to be above in subsequent years. This correlation can be considered as distinguishing between two people who have the same measured characteristics — age, sex, cohort, education — but unmeasured differentiating characteristics such as skill or occupation. The equation predicts the average outcome for the two people at any point in their lives, but the higher earner generally remains above the predicted line, and the lower earner generally remains below it.

The second observation is that, although the tendency to be a high or low earner is evident in the data, some statistically random movement occurs in people's relative standing. Put simply, this means that a person's unexplained earnings — referred to as his or her "permanent earnings differential" in the CBOLT micro model — will evolve randomly over time. The random evolution of the unexplained earnings component can be captured using a number of statistical methods; the method used in CBOLT is to specify "permanent" shocks to the earnings differential terms. This works in conjunction with the (also random) "transitory" earnings shocks to determine actual earnings in a given year.

To summarize, in the CBOLT micro model, every individual's full-time equivalent earnings are composed of three components: the first is predicted by the earnings equation; the second is the permanent earnings differential, which evolves randomly over time; and the third is the current year transitory shock, which captures other unexplained variation. This decomposition of the earnings process is fairly complicated, but it is essential to generating longitudinal earnings profiles that show variation within and across people that are consistent with actual data.

5.5 Social Security Benefit Claiming and Awards

The primary goal of building the CBOLT micro model was to create a method for projecting Social Security benefits under baseline and alternative policies. The emphasis placed on the various processes speaks to this goal; we need realistic longitudinal labor force and earnings patterns and marital histories in order to simulate benefit awards under all the various OASDI programs.[19] In addition to having the underlying micro data to which rules are applied, the model also needs a set of behavioral rules that control the initiating of benefit claiming.

---

[19] In the current version of CBOLT, the only benefits not computed directly using the micro sample are for children and parents, because those linkages do not yet exist at the micro level. That part of the model is currently under development.

In the case of OAI and DI worker and spousal auxiliary benefits the CBOLT micro model has all the information needed to compute benefit amounts given claiming status. In the dynamic micro model, actual benefit calculations are made using the same code that was developed for CBOLT's static-aged micro model mode-one and mode-two actuarial simulations. That code already computes worker benefit awards using detailed policy rules. The only innovation of the dynamic micro model is that the sample of beneficiaries is now representative. Otherwise, inputs to the benefits calculator — age, current year, and earnings histories — are identical.[20]

CBOLT's micro model captures both aggregate and individual-level factors in the claiming behavior of DI and OAI workers. With respect to DI, the overall rates of incidence and termination by age and sex are controlled by the same input parameters utilized in the actuarial model (which incorporates expected trends in DI). The micro model is, however, careful to differentiate who — within an age-sex group — is likely to go on DI in a given year. The important variable by which to differentiate within an age/sex group is earnings. Because low earners are more likely to claim disability insurance, the expected benefits for DI recipients are lower than if one took a random sample of earners from a given age/sex group and assigned them to the DI program.

In the case of OAI worker claiming behavior**,** the initial approach tested for CBOLT was a structural retirement model that would predict retirement behavior by assuming people rationally compute potential benefits under alternative retirement dates, then choose a retirement date by making

---

[20]A preliminary version of this module was used by Harris and O'Harra (2001) to investigate the outlook for women's Social Security benefits.

a trade-off between benefit levels and the lure of leisure.[21] Unfortunately, structural models are not very good predictors of people's behavior in this area. In particular, important reference ages seem to be statistically more significant for explaining retirement than benefit trade-offs. With Social Security, the classic example is that most people retire at age 62 when early retirement is available, even though actuarial reductions associated with early retirement make retirement at any time between age 62 and 65 a proposition that is basically neutral. Absent self-selection, such "irrationality" makes structural modeling based on a rational agent's hypothesis difficult.

The OAI worker claiming module in CBOLT takes this age-referencing behavior as a starting point to which a policy response is added. The approach is to use the retirement behavior of the most recent observed cohorts — those who retired in the late 1990s — as a reference to determine a baseline probability of claiming at all ages greater than or equal to the early retirement age. The second step is to hypothesize that decreases in benefits relative to average economywide earnings — the so-called replacement rate — will cause people to work marginally longer, which is implemented by decreasing their claiming probabilities. Because the degree of responsiveness to benefit changes is an exogenous parameter that is under the user's control, sensitivity analysis is possible.

5.6 Distributional Analysis

In addition to the goal of producing better budgetary projections, a distinct benefit of using dynamic micro simulation is the ability to undertake distributional analysis, defined here as the tracking of tax and benefit outcomes across groups and time. In the standard SSA actuarial approach

[21]For details on this research, see Harris (2002).

distributional analysis is not an intrinsic feature of the model. Policy makers are provided with results for hypothetical example workers who bear no necessary relationship to real people; they are effectively historical averages adjusted for average wage growth. The dynamic micro model, on the other hand, automatically produces a representative sample in every forecast year, so that distributional results across a number of dimensions are possible.

The limits on distributional analysis in any given model are determined by the lowest level of detail in that model. Thus, in the group-level actuarial modules, it is possible to track benefit outcomes by age, sex, and birth year. However, because taxes are computed using the aggregate univariate distribution, comprehensive distributional analysis by age, sex, and birth year is not possible. This is why hypothetical example workers, with earnings fixed relative to the average wage index, are used by the SSA actuaries in their distributional tables.

The dynamic micro model allows a much more comprehensive look at distributional effects. The basic demographic and economic variables tracked at the person-level allow the sample to be divided up in many imaginable ways. Standard CBOLT distributional tables consider taxes and benefits by sex, birth year, lifetime earnings, marital status, and other basic variables. Also, the "representativeness" of the CBOLT micro sample is not suspect like it is for example workers. CBOLT tables for the "average" worker truly reflect an average outcome, not an outcome fixed relative to the average wage.

6. Stochastic Simulation Using Monte Carlo Draws for Inputs

All of the CBOLT simulation techniques described thus far assume a deterministic solution where the user defines a model solution mode and associated options or behavioral parameters, policy rules, and values for the exogenous input variables. For stochastic solutions the first two steps are the same, but the actual values for the exogenous inputs are drawn from a probability distribution using a Monte Carlo simulation technique. In CBOLT, users of stochastic runs can alter the expected values for exogenous inputs, after which actual inputs vary symmetrically around those expected values.[22]

Any given stochastic simulation yields a viable scenario, which means some chance exists that the particular set of exogenous inputs chosen in that set of random draws will be observed in the real world, even though it may not be the most likely scenario. The utility of stochastic simulations becomes apparent after the model has been solved many times using different sets of draws for the inputs each time. Because the probability distribution for each input has a bell-shaped distribution where values are likely to bunch around the expected value, repeated stochastic simulations will generate a distribution of system financial outcomes that also exhibits bunching. After a sufficient number of simulations have been run, it is possible to make inferences about the probability distribution of the same budgetary and distributional outcomes generated by a deterministic simulation.

---

[22]The SSA actuaries also introduced stochastic analysis in their latest (2003) *Trustees Report*, using a Monte Carlo simulation approach that was largely patterned on CBOLT. When it comes to generating ranges, however, their main focus is still on low and high cost projections.

6.1 Beyond Low-, Medium-, and High-Cost Projections

Each year, Social Security actuaries produce three sets of projections for their annual *Trustees Report*. Of those, the projection that generates the most attention is the intermediate one, because it is considered the most likely scenario. To produce the intermediate forecast the actuaries use values for the nine major input assumptions provided by the Trustees — fertility rate, rate of mortality improvement, number of immigrants, real wage growth, inflation, rate of unemployment, interest rate, disability incidence rate, and disability termination rate.

To characterize uncertainty about the projections, the actuaries produce two other simulations — the low-cost and high-cost scenarios. For those simulations, all inputs are set to some alternative value, each consistent with either improved or deteriorated system finances. For example, in the low-cost scenarios, actuaries use a lower rate of mortality improvement, higher fertility rate, higher number of immigrants, higher real wage growth, lower inflation, lower unemployment rate, and lower rate of DI incidence, because those values for the inputs all push the OASDI system toward increased solvency. In the high-cost scenario, the reverse is true (for example, mortality is assumed to improve more rapidly).

Several problems are inherent in using the high- and low-cost scenario approach to measuring uncertainty about projections. First, the gap between high-/intermediate- or low-/intermediate-values for a given input has no probabilistic interpretation based on evaluation of historical data. Second, in the alternative scenarios, all inputs move together, even though no statistical basis exists for expecting that type of correlation. Finally, rather than adding both high frequency (annual) and long-run variation, the actuaries only alter long-run values for the input assumptions. CBOLT stochastic simulations address all of those problems; the estimation phase involves identifying the time-series properties of each input and the correlations between inputs. Actual implementation involves drawing annual values for each input using estimated equations.

6.2 Time-Series Processes for Exogenous Inputs

The estimation phase of the stochastic simulations involves empirical investigation of each input assumption using time-series analysis. The focal point of the time-series analysis is measuring the dynamics of a given input assumption relative to a benchmark of white noise variation; this variation is characterized as the completely unexplained, uncorrelated movement around some central tendency. Time-series analysis examines each input assumption variable with the belief that a white noise process exists and can be isolated after all empirically identifiable correlations (over time and with other variables) have been removed.

Time-series processes for CBOLT's input assumptions are described here briefly (see Figure 9).[23] In looking at these processes, the first noteworthy point is that the list of stochastic inputs differs in static and growth model solutions. In a static run, the model needs input values for real wage growth and interest rates; however, those input values are effectively replaced by total factor productivity growth and the gap between 5- and 10-year rates in a growth model simulation. In addition, because of correlations between variables, equations for some of the other inputs also change when moving from mode-one to mode-two solutions.

---

[23] The equations are described in detail in the December 2001, Congressional Budget Office paper, *Uncertainty in Social Security's Long-Term Finances: A Stochastic Analysis.*

The descriptions of time-series processes (see Figure 9) for the inputs use standard econometric shorthand notation. For example, AR-1 implies an autoregressive equation with one lagged value. Consequently, mortality equations, the rate of improvement in each age group (actually the rate of improvement differenced from its own mean), depend on one lagged value and a white noise process. The ARMA(4,1) equation for fertility implies both autoregressive and moving average components to the dynamic process. Finally, the VAR notation indicates a vector autoregression. In those equations, each variable depends on its own lagged values and the lagged values of other variables in the VAR.

6.3 Properties of Stochastic Simulations

CBOLT's stochastic simulator is, in effect, a separate model that runs at the beginning of each simulation year. Using the time-series equations and other necessary information (usually lagged values of the stochastic variables), the simulator generates input values for the current year. The main model then solves for the current year with that stochastic set of inputs, just as if the inputs had been set that way by the user in a deterministic run. A standard CBOLT stochastic simulation involves 500 to 1,000 complete CBOLT simulations, each with a different trajectory for the inputs.

Although computationally intensive, stochastic simulation provides valuable lessons about the long-term projections. For example, one finding that was surprising at first is that the expected financial outcomes in a stochastic simulation differ from a deterministic solution even when the inputs are set to the same expected values. This occurs because of asymmetries in OASDI system responses with respect to symmetric variation in the inputs. That is, if inflation is expected to be 3 percent, the time-series equations imply that 2.5 percent or 3.5 percent are equally likely outcomes in a stochastic simulation. The positive financial effect of raising inflation from 3 percent to 3.5 percent does not, however, necessarily match the negative effect of lowering inflation from 3 percent to 2.5 percent.

Stochastic simulations also provide interesting findings with respect to overall levels of uncertainty about projections. First, although the ranges for inputs used by the Social Security actuaries in their high- and low-cost scenarios seem reasonable, conclusions about overall system finances do not. For instance, the low-cost scenario suggests that through the 75-year evaluation period system solvency is within the range of plausible outcomes because the summary balance is actually slightly positive. However, the base-case CBOLT stochastic simulation in the static (mode-one) solution with the expected values of the inputs set to the actuaries' intermediate values suggests only a small chance of solvency through 2077.[24]

In addition, CBOLT stochastic simulations suggest that uncertainty grows dramatically as simulations are done farther into the future and the reason for that uncertainty can be tracked back to a handful of input assumptions. The range of plausible outcomes for measures such as the gap between cost and income rates rises dramatically as one projects farther in time, with standard confidence intervals swamping the size of the gap itself by the end of the evaluation period. In the near term, however, system finances seem quite predictable because many of the important determinants of variability in near-term finances are already known — population characteristics, earnings histories for the soon to be retired, and so on. In the future, those determinants are, of course, much more

[24]See Harris, Meyerson, and Smith (2001) or Sabelhaus and Meyerson (2000) for an application to uncertainty about trust fund investment policies.

uncertain. For instance, the effect of variability in fertility on the size of the population is much lower at the beginning of the simulation than it is 30 years later because the size of the population at the beginning is known, whereas the size of the population that can give birth 30 years from now can only be projected. As a result, future fertility will be the product of cumulative random fertility, mortality, and immigration. One of the important lessons from stochastic analysis is that the current OASDI system is self-correcting with respect to some input assumptions, such as inflation, because both costs and incomes are affected. Other variables only cause movements in one direction — mortality improvement is a good example. Thus, reforms aimed at improving the expected financial status of the system should also be evaluated in terms of their impact on the variability of outcomes.

Stochastic analysis is also important for distributional analysis. For example, although it is widely acknowledged that policies involving investment in private assets will yield uncertain returns, it is not clear that the uncertainty about individual benefit levels would change dramatically for individual accounts created using a modest carve-out with a simple benefit offset mechanism.[25] Some analysts contend that individual accounts should be analyzed using just the "risk free" interest rate, while others contend the risk premium should be included. Stochastic analysis offers an alternative: the risk premium is included on an expected basis, but the historical variability in equity returns provides an estimate of the range of possible outcomes.

[25]See Harris, Sabelhaus, and Simpson (2003).

References

Board of Trustees, Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds, *The 2003 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds* (Washington, D.C.: U.S. Government Printing Office, March 2003).

Congressional Budget Office, *The Long-Term Budget Outlook* (Washington, D.C.: Congressional Budget Office, December 2003)

Congressional Budget Office, *The Budget and Economic Outlook: Fiscal Years 2004-2013* (Washington, D.C.: Congressional Budget Office, January 2003).

Congressional Budget Office, *Uncertainty in Social Security's Long-Term Finances: A Stochastic Analysis* (Washington, D.C.: Congressional Budget Office, December 2001).

Congressional Budget Office, *CBO's Method for Estimating Potential Output: An Update* (Washington, D.C.: Congressional Budget Office, August 2001).

Harris, Amy Rehder, "Modeling Retirement Behavior: A Test of the Option Value Model Using the Health and Retirement Study." (Draft) Washington, D.C.: Congressional Budget Office, (April 2002).

Harris, Amy Rehder, and Josh O'Harra, "The Impact of Marriage and Labor Force Participation Trends on the Social Security Benefits of Women." *Proceedings of the National Tax Association* (November 2001)

Harris, Amy Rehder, Noah Meyerson, and Joel V. Smith, "Social Insecurity? The Effects of Equity Investments on Social Security Finances," *National Tax Journal*, LIV(3) (2001), pp. 645-668.

Harris, Amy Rehder, and John Sabelhaus, "Projecting Longitudinal Earnings for Long-Term Policy Analysis," Technical Paper 2003-2. Congressional Budget Office (April 2003).

Harris, Amy Rehder, John Sabelhaus, and Michael Simpson, "Uncertainty About OAI Worker Benefits Under Individual Accounts." (Draft) Congressional Budget Office (May 2003).

Harris, Amy Rehder, John Sabelhaus, and Michael Simpson, "Projecting Long-Run OASDI Benefits: Implications of Adopting a Dynamic Micro-Simulation Approach." (Draft) Congressional Budget Office (January 2004).

Ibbotson Associates, *Stocks, Bonds, Bill and Inflation 2000 Yearbook: Market Results for 1926-1999* (Chicago: Ibbotson Associates, 2000).

O'Harra, Josh, and John Sabelhaus, "Projecting Longitudinal Marriage Patterns for Long-Term Policy Analysis," Technical Paper 2002-2. Congressional Budget Office, (October 2002).

Perese, Kevin, "Mate Matching for Micro-Simulation Models," Technical Paper 2002-3, Congressional Budget Office (November 2002).

President's Commission to Strengthen Social Security, *Strengthening Social Security and Creating Personal Wealth for All Americans* (Washington, D.C.: Report of the President's Commission to Strengthen Social Security, December 2001), available at www.csss.gov/reports/Final_report.pdf.

Sabelhaus, John and Noah Meyerson, "Uncertainty in Social Security Trust Fund Projections." *National Tax Journal*. 53:3, 515-530. (September 2000).

Sabelhaus, John, Julie Topoleski, and Michael Simpson, "Incorporating Longevity Effects in Long-Term Medicare Projections." (Draft) Congressional Budget Office (November 2003).

Sabelhaus, John, "For Whom the Bill Tolls: Distributing the Cost of Maintaining Social Security Solvency." (Draft) Congressional Budget Office, (May 2003).

Smith, Joel V. and John Sabelhaus, "Alternative Methods for Projecting Equity Returns: Implications for Evaluating Social Security Reform Proposals" Technical Paper 2003-8. Congressional Budget Office (August 2003).

Social Security Administration, Office of the Chief Actuary, "Internal Real Rates of Return Under the OASDI Program for Hypothetical Workers" (Washington, D.C.: SSA Actuarial Note No. 144, June 2001).

Figure 1.
Year-by-Year Solution Sequence for Static and Growth Model Simulations

Solution Mode 1 (Static Macro)	***Solution Modes 2 and 3 (Growth Model)***
1. Population	1. Population
2. Labor Force	2. Labor Force
	3. Production and Output
	4. Incomes and Factor Prices
3. Social Security Covered Workers	5. Social Security Covered Workers
4. Social Security Payroll Taxes	6. Social Security Payroll Taxes
5. Social Security Beneficiary Counts	7. Social Security Beneficiary Counts
6. Social Security Benefit Amounts	8. Social Security Benefit Amounts
7. Social Security Trust Funds	9. Social Security Trust Funds
	10. Medicare
	11. Other Federal Revenues and Spending
	12. State and Local Budgets
	13. Foreign Transactions
	14. Private Saving
	15. Next Period Beginning Capital Stock

Figure 2.
Actuarial Versus Dynamic Micro Simulation Modules in Growth Model Solutions (Modes 2 and 3)

	Solution Mode 2 ***Actuarial Module Simulations***	***Solution Mode 3*** ***Dynamic Micro Simulations***
1. Population	C Apply mortality rates by age and sex to population matrix, solve for number of deaths. C Apply fertility rates by age to female population matrix, solve for number of births. C Add immigrants; overall count is exogenous, distribution by age and sex uses fixed percentages. C Apply fixed marital distribution to population by age and sex; solve for counts of never married, married, divorced, and widowed.	C Use actuarial modules to update overall population matrix by age and sex. C Apply mortality, fertility, and immigration rates to micro sample; add and drop observations in micro sample through births, immigrations, and deaths. C Apply marital transition equations across four possible marital states. C Put observations entering marriage (from never married, divorced, or widowed) through mate-matching algorithm.
2. Labor Force	C Solve for labor force participation rates across 103 age, sex, and marital status groups; independent variables include trends and Social Security benefit replacement rates. C Apply labor force participation and exogenous unemployment rates to population matrix to yield employment matrix by age and sex; multiply by exogenous average hours to solve for total hours worked.	C Predict labor force participation for each person in the micro sample; determinants include age, sex, cohort trends, marital status, benefit levels, and taxes. C Predict part versus full time and hours worked for each observation; same determinants. C Apply exogenous unemployment rates to micro sample. C Sum hours worked in micro sample to solve for total hours worked.

Figure 2, Continued

	Solution Mode 2 *Actuarial Module Simulations*	*Solution Mode 3* *Dynamic Micro Simulations*
3. Production and Output	• Solve for aggregate production based on total hours worked (from actuarial modules), total capital stock, and exogenous total factor productivity.	• Solve for aggregate production based on total hours worked (from micro sample), total capital stock, and exogenous total factor productivity.
4. Incomes and Factor Prices	• Solve for total earnings and capital income (given aggregate production and exogenous prices) using first-order conditions from production function. • Compute rate of real wage growth, which is earnings divided by hours worked, adjusted for earnings as a share of compensation and other factors. • Solve for interest rate by dividing net capital income by capital stock.	• Perform same first three steps as in the actuarial model. • Compute individual earnings for observations in the micro sample with positive labor supply; apply permanent and transitory shocks and overall real wage growth.
5. Social Security Covered Workers	• Apply fixed covered worker rates by age and sex to aggregate employment matrix solved for in labor force module.	• Assign covered worker status to micro sample; randomly set small fraction of earnings for workers not covered by Social Security.
6. Social Security Payroll Taxes	• Compute aggregate taxable payroll on the basis of fixed univariate earnings distribution and taxable maximum policy parameter, aged forward using real wage and covered worker growth rates.	• Compute aggregate taxable payroll by summing taxable earnings (covered earnings below the taxable maximum) across all observations in the micro sample.

Figure 2, Continued

	Solution Mode 2 *Actuarial Module Simulations*	*Solution Mode 3* *Dynamic Micro Simulations*
7. Social Security Beneficiary Counts	• Compute OAI worker beneficiary count using insured worker rates by age and sex for people ages 62 through 70, which in principle is based on past covered worker counts but in practice is nearly 100 percent and constant. • Compute DI worker beneficiary counts by age and sex on the basis of exogenous DI incidence and termination rates applied to underlying age- and sex-specific probabilities. • Compute dual beneficiaries using an SSA formula that predicts duals on the basis of earnings of men versus women. • Compute other auxiliary beneficiaries largely on the basis of exogenous ratios of population subgroups; for example, widow beneficiaries are a fixed fraction of the number of widows ages 60 and older.	• Compute OAI worker beneficiary count by applying OAI worker claiming equation to all observations in the dynamic micro model with age greater than or equal to early retirement age. The main explanatory variable in the equation is the level of benefits relative to earnings (replacement rate) with one parameter controlling response to benefit changes. • Compute DI worker beneficiary counts by applying micro DI incidence and termination equations; exogenous DI incidence and termination rates by age and sex determine overall flows, with person-level *relative* probabilities of DI affected by individual *relative* earnings. • Compute dual and other auxiliary beneficiary counts on the basis of actuarial modules, though implementation of micro-level nonworker beneficiary status is under way.

Figure 2, Continued

	Solution Mode 2 ***Actuarial Module Simulations***	***Solution Mode 3*** ***Dynamic Micro Simulations***
8. Social Security Benefit Amounts	C Compute new OAI and DI worker beneficiary awards using “static” micro simulation on CWHS samples of 1996 beneficiaries with earnings and labor force participation adjusted to be basically consistent with actuarial/macro labor force and earnings outcomes. C “Age” forward existing OAI and DI average worker benefits by age, sex, and age at entitlement using fixed ratios derived from SSA data; those ratios control for the effects of differential mortality and recomputations on average benefits within a cohort. C Compute auxiliary benefits are based on fixed (SSA) ratios to OAI and DI workers.	C Compute new OAI and DI worker benefits using dynamic micro simulation with earnings and labor force participation automatically consistent with actuarial/macro labor force and earnings outcomes (since those are based on the same micro sample). Existing OAI and DI worker average benefits are the result of tracking benefits for those who remain alive; that implies the effect of differential mortality will not be observed in benefits unless mortality rates are adjusted for earnings. C Compute auxiliary benefits in the same way as actuarial benefits, though implementation of nonworker benefits is under way.
9. Social Security Trust Funds	C Apply OASI and DI tax rates to taxable payroll, add benefits tax, and solve for total system income. C Multiply average benefits by number of beneficiaries across age/sex groups and for each worker/auxiliary outlay category, add administrative costs, and solve for total system costs. C Set increments for OASI and DI trust funds using costs, income, and interest rates.	C Apply OASI and DI tax rates to individual earnings up to taxable maximum, add benefits tax, and solve for total system income. C Sum benefits paid to OAI and DI worker beneficiaries, sum actuarial-based auxiliary outlays, add administrative costs, and solve for total system costs. C Set increments for OASI and DI trust funds using costs, income, and interest rates.

Figure 2, Continued

	Solution Mode 2 ***Actuarial Module Simulations***	***Solution Mode 3*** ***Dynamic Micro Simulations***
10. Other Federal Revenues and Spending	C Solve for aggregate outlays in Medicare and other "age-related" programs using population matrix and fixed age/sex spending indexes. C Solve for other (non-age-related) federal spending using fixed ratios to GDP. C Solve for aggregate personal and corporate income taxes, indirect business taxes, and non-Social Security social insurance taxes using fixed ratios to aggregate taxable income. C Adjust tax and/or spending rates if a deficit rule is activated and recompute outlays and revenues to meet deficit rule.	C Perform the same steps as in the actuarial modules; no micro detail is available for taxes and spending (except Social Security).
11. State and Local Budgets	C Use fixed ratios of taxes and outlays (except age-related) to GDP to set most values. C Adjust taxes when expenditures on age-related programs change such that state and local deficit is constant as a share of GDP.	C Perform the same steps as in the actuarial modules; no micro detail is available for state and local taxes or spending.
12. Foreign Transactions	C Use fixed ratios of capital flows to GDP to determine net exports and net foreign investment. C Calculate net exports to close out product side of GDP accounts and net foreign investment to help determine next period capital.	C Perform the same steps as in the actuarial modules.

Figure 2, Continued

	Solution Mode 2 *Actuarial Module Simulations*	*Solution Mode 3* *Dynamic-Micro Simulations*
13. Private Savings	C Determine private savings using one of several options: complete savings offset (Ricardian), constant saving elasticity, variable savings elasticity, and variable deficit offset. With complete savings offset, as federal debt increases, private savings matches one for one. Constant savings elasticity means that as debt rises, the capital stock initially falls, then interest rates rise, causing savings to rise and offset part of initial loss. For the other options, the degree of deficit offset or response to interest rates varies with the level of debt; if debt gets very high, economy is effectively Ricardian.	C Perform the same steps as in the actuarial modules. Implementation of micro-level savings responses (based on forward-looking behavior) is under way.
14. Next Period Beginning Capital Stock	C Solve for total investment as the sum of private savings, federal government saving, state and local government saving, and net foreign investment. C Allocate investment across several types of capital; each capital stock grows by investment minus depreciation, overall stocks are then weighted to determine aggregate capital input for production function.	C Perform the same steps as in the actuarial modules.

Figure 3.
Policy Parameters and Exogenous Input Assumptions
in CBOLT Static Macro (Mode 1) Simulations

Social Security Basic Parameters

Payroll Tax Rate
COLA Differential
Benefit Formula Replacement Factors
Number Years AIME Computation
Normal Retirement Age
Actuarial Reduction Rate
Delayed Retirement Credit
Bend Points
Taxable Maximum

Demographic Input Assumptions

Mortality
Fertility
Immigration

Exogenous Trust Fund Assumptions

Administrative Expense Rate
Benefits Tax Rate
General Fund Transfers
Transfers to Railroad Retirement

Economic Input Assumptions

CPI-W Inflation
Unemployment Rate
Real Wage Differential
Real Interest Rate
Disability Incidence Rate
Disability Termination Rate
Equity and Corporate Bond Return Rates

Individual Account Parameter Settings

IA Administrative Cost/Transfer Rules
IA Contribution Rates
IA Participation Rules and Behavior
IA Portfolio Allocation
IA Annuitization Assumptions
IA Benefit Offset Rules

Other Policy Rules

Trust Fund Investment Policy
Low-Wage Supplementation
Trust Fund Shortfall Responses

Figure 4.
Policy Parameters and Exogenous Input Assumptions
in CBOLT Macro Growth (Modes 2 and 3) Simulations

Social Security Basic Parameters

- Payroll Tax Rate
- COLA Differential
- Benefit Formula Replacement Factors
- Number Years AIME Computation
- Normal Retirement Age
- Actuarial Reduction Rate
- Delayed Retirement Credit
- Bend Points
- Taxable Maximum

Demographic Input Assumptions

- Mortality Improvement
- Fertility
- Immigration

Exogenous Trust Fund Assumptions

- Administrative Expense Rate
- Benefits Tax Rate
- General Fund Transfers
- Transfers to Railroad Retirement

Other Policy Rules

- Trust Fund Investment Policy
- Low-Wage Supplementation
- Trust Fund Shortfall Responses
- Federal Tax/Spending Policy
- Debt-Ceiling Response Policy

Economic Input Assumptions

- CPI-W Inflation
- Unemployment Rate
- Disability Incidence Rate
- Disability Termination Rate
- Equity and Corporate Bond Return Rates
- Total Factor Productivity
- Gap Between CPI-W and GDP Prices
- Gap Between Computer/Other Prices
- Gap Between 10- and 5-Year Interest Rates
- Average Hours Growth Rate
- Growth of Taxable Earnings/Compensation
- Capital Share of Output

Individual Account Parameter Settings

- IA Administrative Cost/Transfer Rules
- IA Contribution Rates
- IA Participation Rules and Behavior
- IA Portfolio Allocation
- IA Annuitization Assumptions
- IA Benefit Offset Rules

Behavioral Assumptions/Parameters

- Private Savings Functional Form
- Private Savings Response Elasticity
- Hours Worked Elasticity and Marginal Rate
- Medicare Excess Cost Growth

Figure 5.
CBOLT Federal Tax, Spending, Debt-Limit, and Trust Fund Shortfall Rules

Federal Tax Rules

- Fix revenues at year 10 percentage of GDP for all years after year 10.
- Fix revenues at year 1 percentage of GDP for all years after year 10.
- Temporarily spend non trust fund surpluses after year 10.
- Permanently cut taxes/raise spending to balance 2011 non trust fund budget.
- Temporarily spend non trust fund surplus in all years.
- Spend non trust fund surplus and fix percentage of GDP after.
- Balance non trust fund budget after year 10.
- Balance non trust fund budget in all years.
- Temporarily spend non trust fund surplus years 1 to 10, use year 10 percentage of GDP after.

Federal Spending Rules

- Fix spending at year 10 percentage of GDP for all years after year 10.
- Fix spending at year 1 percentage of GDP for all years after year 10.
- Temporarily spend non trust fund surpluses after year 10.
- Permanently cut taxes/raise spending to balance 2011 non trust fund budget.
- Temporarily spend non trust fund surplus in all years.
- Temporarily spend non trust fund surplus years 1 to 10, use year 10 percentage of GDP after.

Federal Debt-Ceiling Rules

- Choose no debt ceiling or ceilings of 0, 50, 100, 200, or 500 percent of GDP.
- Choose whether to use tax increases or spending cuts to stay under the ceiling.

Trust Fund Shortfall Rules

- Accumulate debt (current law benefits).
- Raise payroll taxes.
- Reduce new benefit awards.
- Reduce all benefits.

Figure 6.
CBOLT Aggregate Private Savings Equations for Growth Model Simulations

Savings Equation	*Description*	*Properties*
Targeted Capital to Output Ratio	Private savings adjusts basically dollar for dollar to government deficits in current year, in order to keep capital to output ratio fixed over time.	Complete offset eliminates any impact of government deficits on investment and thus economic growth. This is the growth model version of the “static” solution.
Constant Savings Elasticity	Private savings is a fixed percentage of GDP, adjusted (on a fixed elasticity basis) for changes in interest rates. Default parameter = 0.2.	As government debt rises, interest rates rise, and private savings starts to rise with a lag. The smaller the elasticity is, the less the increase in private savings and the higher the crowding out.
Variable Savings Elasticity	Modified version of the constant savings elasticity in which the responsiveness of private savings depends on the level of the interest rate through a logistic function. Default parameter = -0.33.	Similar to the constant elasticity, but the degree of responsiveness changes as debt rises, so the variable elasticity actually becomes more “Ricardian” as interest rates rise.
Variable Deficit Offset	Variable version of the complete savings offset; rather than have private savings offset every dollar of government debt, the fraction offset varies with the interest rate. Default parameter = -2.0.	Like variable savings elasticity, the degree of responsiveness changes as debt rises, so the variable elasticity actually becomes more “Ricardian” as interest rates rise.
Partial Savings Offset	Similar to complete offset model in that the fraction of debt offset does not vary, but here the fraction is below one. Default parameter = 0.4.	Interest rate has no effect on offset parameter, so equation can actually lead (under high debt levels) to unstable economy, because private savings becomes too low to fund any new investment.

Figure 7.
Person-Level Variables Tracked in the CBOLT Dynamic Micro Model

Variable	*Description/Possible Values*
Birth Year	Year of birth; given simulation year, derive age
Immigration Year	Year of immigration
Sex	1=male, 2=female
Education	0=<14 year of education, 1=14+
Marital Status	For each age 0 to 100, 1=never married, 2=married, 3=divorced, 4=widowed
Need Spouse	Logical variable = true if in marriage market
Divorce Eligible	Indicator if divorced with marriage of 10+ years
Earnings	For each age 16 to 90, level of actual earnings
Percentage Earnings Uncovered	For each age 16 to 90, the fraction of person's earnings not covered by Social Security
Full Time Equivalent Earnings	For current age, level of earnings if the person worked full time with no unemployment
Lifetime Earnings	Average earnings through current year
Lifetime Earnings Quintile	Classifier for lifetime earnings
Labor Force Participation	For each age 16 to 90, 0 = not in labor force, 1 = part time with unemployment, 2 = part time, 3 = full time with unemployment, 4 = full time
Permanent Earnings Differential	Evolving heterogeneous earnings component; subject to "permanent" shock each year
Social Security Benefit	Level of current-year benefit
Social Security Benefit Type	For each age 16 to 100, 1=OAI, 2=DI, 3=OAI dual, 4=SI dual, 5(6)=OAI (div) aged spouse, 7(8)=SI (div) aged widow, 9(10)=SI (div) disabled widow, 11(12)=DI (div) aged spouse
IA Participation	Indicator for whether participating in IA
DI Earnings Decile	Earnings classifier for DI incidence equation
Spouse Identifier	Pointer variable to spouse's information

Figure 8.
Person-Level Transition Processes in the CBOLT Dynamic Micro Model

Transition Process	***Micro Determinants***	***Estimation/Implementation in Dynamic Model***
Mortality	Age, sex, cohort, disability insurance status	C No estimation; mortality rates in actuarial model determine population by age and sex in each year; ratio of micro sample to target population by age and sex determines mortality probability. C Actual mortality based on comparing derived probability to a random number; within age/sex groups people on DI more likely to die.
Fertility	None	C No estimation; fertility rates in actuarial module determine the number of new births. C New observations are added to the micro sample at a rate of 1:1000 births.
Immigration	None	C No estimation; immigration rates in actuarial module determine the number of new immigrants added to the micro sample at a rate of 1:1000. C Age/sex distribution of immigrants is based on the fixed actuarial distribution but with a random component because rounding to nearest 1,000 loses too much detail.
First Marriage	Age, sex, earnings, education, cohort	C Separate logit equations for probability of first marriage estimated by single year of age (17 through 60) using matched SIPP/SSA earnings data. C Actual marital events are based on comparing derived probability to a random number.
Divorce	Age, sex, education, earnings, cohort, marital history and current duration	C Separate logit equations for probability of divorce estimated by single year of age (17 through 70) using matched SIPP/SSA earnings data. C Actual divorce events are based on comparing average of derived probabilities for couple to a random number.
Widowing	None	C Happens automatically in model when linked spouse is de-allocated in mortality module.

Figure 8, Continued

Transition Process	*Micro Determinants*	*Estimation/Implementation in Dynamic Model*
Remarriage	Age, sex, education, earnings, cohort, duration since last marriage	C Separate logit equations for probability of remarriage estimated by single year of age (17 through 70) using matched SIPP/SSA earnings data. C Actual remarriage events are based on comparing derived probability to a random number.
Mate Matching	Age difference between bride and groom; bride's marriage number; education and earnings gaps between bride and groom	C Logistic equations for probability of husband/wife union estimated using matched SIPP/SSA earnings data set with marriages that occurred in 1994, 1995, or 1996. C Separate equations estimated for men's first and higher order-marriages. C Actual unions in CBOLT are based on comparing the probability of a particular couple getting married to a random number in a stochastic "marriage market."
Labor Force Participation	Age, sex, marital status, beneficiary status, lagged participation, cohort	C Separate logit equations for probability of working for ages 16 through 90 estimated using pooled March CPS data for 1966 through 2001. C Labor force participation events are based on comparing derived probability to a random number.
Part Time/ Full Time	Age, sex, marital status, lagged participation, lagged part time/full time, cohort	C Separate logit equations for probability of working full time (conditional on working) for ages 16 through 90 estimated using PSID data for 1968 through 1992. C Working full time event is based on comparing derived probability to a random number.
Unemployment	Age, sex, part versus full time	C Unemployment rates are set exogenously, translated from weighted average rates to probability of spells and allocated across part versus full time workers on the basis of CPS data. C Actual implementation is based on comparing the probability of spell to random number, assigning average spell to all unemployed people.

Figure 8, Continued

Transition Process	*Micro Determinants*	*Estimation/Implementation in Dynamic Model*
Earnings	Age, sex, education, cohort	C Separate equations for men and women estimated using CPS data for 1976 through 2001. C Independent variable is the log of real full-time equivalent earnings. C Actual implementation is based on decomposing person-specific error term into (1) a highly auto-correlated "permanent differential"that evolves through iid "permanent" shocks and (2) an uncorrelated "transitory" shock; thus, each person gets two random numbers each year that together (with demographics) determine earnings.
Taxable Payroll	None	C No estimation; taxable payroll is all earnings below taxable maximum parameter.
DI Claiming and Benefits	Age, sex, lagged relative earnings, eligibility status	C OACT incidence probabilities are adjusted to reflect information derived from historical earnings of DI claimants using CWHS. C Actual DI claiming is based on comparing adjusted probabilities to a random number. C Benefit amounts are based on an individual's earnings and DI benefit parameters.
OAI Claiming and Benefits	Age, sex, OAI benefits relative to earnings, eligibility status	C Conditional claiming probabilities by age and sex are computed using CWHS in 1997 and 1998 for eligible claimants. C Additional benefit replacement rate parameter in logistic equation is used to adjust conditional probability; that introduces an effect of changes in benefit replacement rates on future cohorts. C Actual OAI claiming are based on comparing adjusted conditional probabilities to a random number. C Benefit amounts are based on an individual's earnings and OAI benefit parameters; includes retirement earnings test and recomputations.

Figure 9.
Stochastic Inputs in CBOLT Monte Carlo Simulations

Variable	*Description of Stochastic Process*
Mortality Improvement	Separate AR-1 equations for each of 21 age and 2 sex groups estimated using data for 1900 through 1995. Model draws correlated errors across 42 groups using multivariate normal distribution.
Fertility	ARMA(4,1) equation for overall fertility rate estimated using data for 1917 through 1997.
Immigration	ARMA(4,1) equation for total immigration estimated using data for 1901 through1995.
Unemployment	VAR model with two lags each on unemployment, inflation, and real interest rate (static) or real interest gap (growth model) estimated using data for 1954 through 1999.
Inflation	VAR model with two lags each on unemployment, inflation, and real interest rate (static) or real interest gap (growth model) estimated using data for 1954 through 1999.
Real Wage Growth (Static Model)	Level of nominal wage is a function of three economic variables; equation estimated using data for 1954 through 1999; real wage is nominal wage minus inflation.
Total Factor Productivity (Growth Model)	A white noise process; standard deviation of innovations is computed using data for 1950 through 2000.
Real Interest Rate (Static Model)	VAR model with two lags each on unemployment, inflation, and real interest rate (static) or real interest gap (growth model) estimated using data for 1954 through 1999.

Figure 9, Continued

Variable	*Description of Stochastic Process*
Interest Rate Gap (Growth Model)	VAR model with two lags each on unemployment, inflation, and real interest rate (static) or real interest gap (growth model) estimated using data for 1954 through1999.
DI Incidence	AR-1 model for overall DI incidence rate estimated using data for 1975 through 1998.
DI Termination	AR-1 model for overall DI termination rate estimated using data for 1975 through 1998.
Equity Returns	Several options for equity returns, all based on data for Ibbotson large cap returns during the period from 1954 through 1999. First option is a white noise process; second includes inflation, unemployment, and interest rate (level or gap, depending on solution) estimated over the period from 1954 through 1999; third includes same economic variables and a log dividend price ratio, estimated over the 1954-1999 period.
Corporate Bond Returns	Level of bond returns is a function of inflation, unemployment, and interest rate (level or gap, depending on solution mode), estimated over the 1954-1999 period.