# PENSIM Overview

by

Martin Holmer, Asa Janney, Bob Cohen
Policy Simulation Group

for

U.S. Department of Labor
Employee Benefits Security Administration
Office of Policy and Research

September 2006

# Preface

This overview has been produced as part of the PENSIM development activities supported at the Policy Simulation Group by U.S. Department of Labor task-order contracts J-9-P-7-0047 and J-9-P-2-0031 from the Office of Policy and Research in the Employee Benefits Security Administration.

PENSIM is a dynamic microsimulation model focused on the analysis of government policy toward employer-sponsored pensions. This overview describes the current version of PENSIM. The overview describes the simulation methodology and logical structure of PENSIM, as well as results from calibration and validation activities. The overview also includes appendices that describe, in detail, the behavioral events and input parameters that are used in PENSIM. A review of the table of contents provides detailed information on the contents of this overview.

Policy analysis conducted with PENSIM is reported in separate series of documents.

# Contents

# 1 Introduction

PENSIM is a pension policy simulation model that has been developed to analyze lifetime coverage and adequacy issues related to employer-sponsored pension plans in the United States.

The design of PENSIM has been influenced by the policy analysis needs of the Office of Policy and Research (OPR) in the Employee Benefits Security Administration (EBSA) of the U.S. Department of Labor (DOL), and by the recommendations of a National Research Council (NRC) panel on retirement income modeling.

OPR plays a central role in the analysis of proposed changes in government policy toward employer-sponsored retirement pension and health insurance plans regulated under the Employee Retirement Income Security Act of 1974 (ERISA). A recurring policy issue with both retirement pensions and health insurance is that not all employees are covered throughout their working years. This fact raises questions in the pension area about the adequacy of retirement income provided by employer-sponsored pensions for some segments of the population. This issue is complex not only because the cross-sectional pattern of coverage during one year does not directly provide information about lifetime coverage, but also because the adequacy of retirement income depends on social security benefit and private retirement savings levels as well as on employer-sponsored pension benefits.

In the late 1990s, an NRC panel reviewed the state of research knowledge concerning retirement behavior (Hanushek and Maritato 1996) and the state of retirement income policy simulation models (Citro and Hanushek 1997). The panel recommended that a number of ongoing modeling efforts be continued, but argued that it was premature to undertake any "large-scale" development of a new dynamic microsimulation model focused on analysis of retirement income policy. Instead, the panel recommended the development of a series of "ad hoc, special purpose models, ... using spreadsheet tools" (Citro and Hanushek 1997, page 160). Other than stating that CBO has had success with this approach in other policy areas, the panel offered no analysis of whether "spreadsheet tools" would be adequate to represent the lifetime dynamics and employee-employer interactions that the panel rightfully identified as being essential to successful retirement income modeling. Ironically, the Congressional Budget Office ignored this advice and embarked on a successful large-scale modeling effort that has produced CBOLT (CBO 2004).

PENSIM development began in September 1997 and follows a middle way between the two extreme model development strategies considered by the NRC panel. It represents the interaction between employees and employers that determines lifetime pension coverage and adequacy without becoming a "large-scale" model. This is done by limiting the model's scope of policy analysis, utilizing innovative simulation methods, and developing the model incrementally.

The sole policy analysis focus of PENSIM is employer-sponsored pension plans. There is no modeling of social security benefits or private retirement savings in PENSIM, even though PENSIM life histories are exported to another model and used by that model to estimate social security benefits (Holmer GEMINI Guide).

As explained in this overview, PENSIM combines a number of standard Monte Carlo techniques into an innovative approach that permits simulation of a large sample of complete pension-related life histories that reflect both the collective and individual risks facing the members of a birth cohort. This choice of simulation methodology permits the execution of a PENSIM run to be completed relatively quickly on a personal computer. This overview presents the rationale for this methodological approach, describes in detail how it has been implemented in PENSIM, and compares PENSIM results with statistics computed from actual data in a series of validation tests.

A small team has developed PENSIM incrementally over a number of years. This approach to model development limits annual costs and has produced, at completion, a more cost-effective model than one developed quickly by a large team (Brooks 1995). In addition, pursuing model development as a leading rather than a lagging policy research activity is expected to identify high-priority data collection and behavioral analysis projects.

# 2 Thumbnail Sketch of PENSIM

PENSIM is a dynamic simulation model that produces life histories for a sample of individuals born in a particular year. PENSIM can produce a cohort sample for any birth year beginning with 1935 and the sample can be any size up to ten percent of the population cohort. The simulated life histories, all of which begin at birth and end at death, are meant to represent the population of the United States.

The life history for a sample individual includes a variety of life events: birth, the timing of schooling events, the timing of marriage and divorce events, the timing of child births for women, the timing of immigration and emigration, the timing of disability onset, and the timing of death. In addition, a simulated life history contains a complete record of jobs held by the individual, including each job's starting date, job characteristics, pension coverage and plan characteristics, and ending date. This life history information is rich enough to allow the estimation for each year in an individual's life of not only social security taxes and benefits, but also employer-sponsored pension benefits.

Section 2.1, beginning on page 5, sketches the simulation of the non-pension aspects of an individual's life history.

Section 2.2, beginning on page 12, sketches the simulation of the pension-related aspects of an individual's life history.

The remainder of this section provides an introduction to the simulation methodology used in PENSIM.

PENSIM is a dynamic microsimulation model, but it does not utilize the conventional simulation methodology developed in the 1960s (Orcutt et al. 1976). Instead, PENSIM uses an innovative new method of dynamic microsimulation pioneered at Statistics Canada (Wolfson 1995). This method can be though of as involving the statistical matching of various longitudinal data sets, each one of which contains survey data on a limited number of life events. The timing of life events observed in a data set is summarized by using the data to estimate a waiting-time model (often called a hazard-function model) using standard survival-analysis methods. Many waiting-time models, estimated using a variety of longitudinal data sets, are then incorporated into a single dynamic simulation model, and this model can be used to simulate a synthetic sample of complete life histories.

This method has been used to develop the LifePaths simulation model in Canada (LifePaths homepage 2005), which has been used in a variety of interesting policy analysis projects (Wolfson et al. 1998, for example).

This new dynamic microsimulation method has also been used to develop PENSIM, a process that began in September 1997 with support from DOL, but without interaction with Statistics Canada.

The use of this innovative dynamic microsimulation method implies that the internal logic of PENSIM differs radically from that of models using the conventional method, primarily because PENSIM employs continuous-time, discrete-event simulation techniques. Life events can occur at any age (for example, giving birth when 25.39276 years old or starting a job at age 34.19472), not just annually on an individual's birthday. And the simulation of a life history jumps irregular lengths of time from one life event to the next pending event. This discrete-event simulation technique allows the easy integration of events that have very different time scales into a single simulation model. Although still not common in socioeconomic simulation, the use of discrete-event simulation techniques has been widespread in business and operations research models for decades (Fishman 1978).

Another novel feature of PENSIM is that it is externally coupled with two other simulation models: SSASIM (Holmer SSASIM Guide) and GEMINI (Holmer GEMINI Guide). SSASIM supplies PENSIM with input, while GEMINI can be used to supplement PENSIM output.

SSASIM, which is a social security policy simulation model, produces simulated time series of macro-demographic and macro-economic variables that describe the environment within which the PENSIM cohort sample lives their lives. So, for example, if the environmental scenario generated by SSASIM calls for a high rate of growth in nominal wages, then the job earnings experienced by PENSIM individuals will be higher than if the SSASIM scenario calls for a low rate of wage growth. SSASIM also supplies PENSIM with various social security policy parameters, such as the maximum taxable earnings level, which are required to calculate pension benefits in cases where the pension plan is integrated with social security.

GEMINI has the ability to estimate annual social security taxes and benefits for each individual in a PENSIM sample. Therefore, it is possible to pass a PENSIM-generated cohort sample through GEMINI to supplement the PENSIM-estimated pension benefits with social security taxes and benefits. This capability produces estimates of both pension and social security income in each retirement year for each individual in the cohort sample simulated in PENSIM. While PENSIM and GEMINI do not estimate private retirement savings, the combination of pension and social security benefits provides an estimate of the bulk of retirement income for a large fraction of the cohort sample.

SSASIM and GEMINI are also externally coupled, with GEMINI producing aggregate demographic and economic statistics, as well as annual social security tax and benefits, from micro samples of each cohort born since 1935, for input to SSASIM when it is operating in the Overlapping Cohorts (OLC) mode. The objective of the SSASIM OLC mode is to build up aggregate financial estimates of social security program solvency from microsimulation samples. The coupling of PENSIM and GEMINI, and of GEMINI and SSASIM OLC, to achieve this objective also permits a type of cross-model validation in which the PENSIM-generated demographic and economic projections produced by SSASIM OLC can be compared with the demographic and economic projections found in the social security Trustees' Report (2005).

## 2.1 Simulation of Lives

The methods used in PENSIM to simulate non-pension life events are summarized in this section. The subsequent section describes the simulation of pension-related life events.

### 2.1.1 Birth

The number of simulated births — the size of a PENSIM cohort sample — is determined jointly by the assumed PENSIM sampling rate, the size of the native-born cohort, and the number of immigrants and emigrants estimated by SSASIM for this birth cohort. Naturally, all births occur at age 0.00000 and the assumed gender ratio of live births is the same as used in the OASDI Trustees Report projection of social security.

For details, see Appendix B, Section B.1, beginning on page 99.

### 2.1.2 Schooling

Each individual in the cohort sample is assumed to start high school at the same age and then progress through school according to several discrete-time waiting-time models estimated using longitudinal data from the early 1990s waves of the Survey of Income and Program Participation (SIPP). The sequence of school-related events is: start high school, finish high school, start college, finish college, and earn a graduate degree. These five events require four waiting-time models (because everyone is assumed to start high school at age 14.00000).

Everyone is assumed to finish high school sooner or later, but the result of finishing high school is determined by a binary logit model that predicts the probability of graduating rather than dropping out. Those who graduate from high school have a waiting time to starting college simulated. This hazard-function model can predict a very long waiting time, which implies the individual never starts college. When college is finished, a binary logit model predicts whether or not a four-year degree is earned. Only those with four-year college degrees have a waiting time to earning a graduate degree simulated.

For details, see Appendix B, Sections B.2–B.6, beginning on page 102.

### 2.1.3 Marriage

At age 12.00000, the individual schedules a first wedding event. The waiting time to the first wedding event is simulated using the gender-specific continuous-time log-linear hazard-function models estimated by RAND for the Social Security Administration (SSA) MINT project (Panis and Lillard 1999, pages 22–27). The waiting time between the end of a marriage and a subsequent wedding event is also simulated with this same hazard-function model because the number of prior marriages is a model covariate. Marriage events scheduled to occur after age sixty are censored.

At the marriage event, the age difference and educational level of the spouse are simulated using probabilities tabulated from almost thirty years of data from the Panel Study of Income Dynamics (PSID). Given these two attributes of the spouse, PENSIM simulates a complete life history for someone from the appropriate birth cohort and with that schooling. It is important to understand that this simulated spouse is not a member of the birth cohort sample being simulated by PENSIM — that is, there is no “marriage market” operating in PENSIM. If that simulated spouse is already married at the time of the wedding, then PENSIM simulates another potential spouse until one is found to be unmarried at the time of the wedding. (Unavailable spouses are not discarded, but saved for consideration at other weddings.) The first available spouse is wed to the sample individual. This bonding of the individuals involves removing the post-wedding marital and childbearing history of the spouse and recording the presence of any step-children that individuals may bring to the marriage.

For details, see Appendix B, Sections B.9–B.10, beginning on page 107.

### 2.1.4 Divorce

The waiting time between the wedding event and the divorce event is simulated using the gender-specific continuous-time log-linear hazard-function models estimated by RAND for the SSA MINT project (Panis and Lillard 1999, pages 28–30). The length of all marriages are simulated with the same two hazard-function models because the number of marriages is a model covariate. If the marriage is scheduled to end in widowhood before the simulated divorce, the divorce event is censored. Divorce events scheduled to occur after age sixty are censored.

At the divorce event, PENSIM makes the conventional assumption that children born during the marriage go with their mother. Step-children from the marriage go with their biological parent.

For details, see Appendix B, Section B.11, beginning on page 109.

### 2.1.5 Fertility

At age 13.00000, a first child-birth event is scheduled for a female individual. The waiting time to the first child-birth event is simulated using one of four continuous-time proportional hazard-function models for separate educational-attainment groups. These four models are estimated for PENSIM using thirty years of PSID data.

The waiting time between each subsequent child-birth event is simulated using the same hazard-function model because the proportional hazard includes age and parity as covariates, as well as dummy variables for the interaction of being married and the husband's educational attainment. The baseline hazard is a piecewise Gompertz hazard function with the number of years since the last child-birth event at each of the kink points varying across the four education groups.

Pending child-birth events are rescheduled in PENSIM when a woman's marital status changes. Child-birth events are censored after age fifty.

Most child-birth events produce a single child, but in PENSIM there is a small probability of the child-birth event producing twins. The probability of twins is assumed to rise for younger birth cohort because the historical trend toward higher ages at child birth. The incidence of twins per 1,000 live births has varied from 18.9 during 1980 to 24.6 during 1994, according to the *Morbidity and Mortality Weekly Report*, Volume 46, Number 6, February 14, 1997.

The constant terms in the proportional hazard-function models vary by birth cohort so that the actual total fertility rates experienced by, or projected for, that birth cohort are replicated in the PENSIM cohort sample.

For details, see Appendix B, Sections B.13–B.14, beginning on page 112.

### 2.1.6 Migration

The volume and age of net immigration is determined by input from SSASIM, which simulates net immigration by year and uses an age distribution of immigrants provided by the SSA Office of the Chief Actuary and used in the intermediate-cost projection in the social security Trustees' Report (2005).

The rate of native-born emigration by age and the rate of foreign-born emigration by year since immigration are both drawn from data used in the SSA Polisim microsimulation model (Polisim homepage 2005) as provided by the SSA Office of Policy.

The information about assumed net immigration and emigration is used in PENSIM to compute implied gross immigration by age and implied emigration by age for the birth cohort.

For details, see Appendix B, Sections B.7–B.8, beginning on page 106.

### 2.1.7 Employment

An individual's PENSIM employment history, which is characterized as a first job and a series of subsequent jobs, is simulated using hazard-function models estimated with SIPP and PSID longitudinal data. There are hazard-function models for the waiting time until the start of the first job, for the duration of the first job, for the waiting time between jobs (when the individual is not employed), and for the duration of subsequent jobs. All these models, except the first-job-start model, are continuous-time hazard functions, so the start and finish of jobs can occur at any (fractional) age.

For details on the simulation of job starts and finishes, see Appendix B, Sections B.17–B.20, beginning on page 114.

When an individual starts a job, six characteristics of that job are simulated using a recursive system of models estimated with SIPP data. The recursive nature of the models means that, in addition to individual attributes (age, gender, education) each simulated job characteristic is included as a covariate in the logically-subsequent job-characteristic models. This recursive structure allows the models to characterize the correlations between job characteristics that appear in the data. In their logical recursive order, the six job characteristics are: employer industry (ten categories), job unionization (two categories), job hours (two categories), initial job earnings (continuous), employer firm size (three categories), and employer pension sponsorship (two categories). Except for the log-linear regression of initial job earnings, these job characteristic models are multinomial logit models.

For details on the simulation of first job characteristics, see Appendix B, Section B.17.1, beginning on page 114. For details on the simulation of subsequent job characteristics, see Appendix B, Section B.19.1, beginning on page 121.

If a job is simulated to have an employer who sponsors a pension and if the individual will ever be pension eligible (often part-time employees will never be eligible for sponsored pensions), the characteristics of the sponsored pension(s) are imputed using a process described in Section 2.2 beginning on page 12.

### 2.1.8 Earnings

An earnings-adjustment event is scheduled to occur exactly one year after the job-start event and every year after that. At each earnings-adjustment event, the individual's job earnings are adjusted using a error-component model similar in structure to that used in the Congressional Budget Office's CBOLT model (Harris and Sabelhaus 2003). The parameters of the model have been calibrated so that PENSIM produces individual earnings histories that have the features of earnings histories observed in longitudinal data.

The nominal level of both initial and adjusted earnings on a job have been linked to the average wage index values simulated in SSASIM, so that the distribution of earnings simulated in PENSIM will shift up and down in nominal terms across scenarios that have higher and lower earnings levels.

For details, see Appendix B, Section B.21, beginning on page 125.

### 2.1.9 Disability

At age 25.00000, the individual schedules a functional-disability event. The waiting time to this event is simulated using a continuous-time log-linear hazard-function model estimated by RAND for the SSA MINT project (Panis and Lillard 1999, pages 31–32). Actually, two variants of the RAND model are used: one for older cohorts and one for younger cohorts. The difference between the two variants is in the magnitude of the piecewise-linear age slopes, with younger cohort experiencing a slightly faster rise in hazard rates at ages before 45 and a slower rise in hazard rates at age after 45.

At the functional-disability event, the waiting time to becoming a social security disabled-worker beneficiary is simulated using a piecewise-linear hazard-function model, where the annual hazard rate rises over calendar time in a way that has been calibrated so that PENSIM produces an incidence of disability by birth cohort that is similar to that in the social security Trustees' Report (2005). Becoming a disabled-worker beneficiary in PENSIM is not precisely the same as becoming a disabled-worker beneficiary in GEMINI because PENSIM does not represent all the details of the disability insured calculation.

Once disabled, individuals in PENSIM never recover. In the real world, a few disabled-worker beneficiaries do "recover" and move off the disability insurance program. PENSIM abstracts from this, but there is no evidence that this hinders the model's ability to simulate the overall features of disability in the population or the finances of the social security disability program.

In PENSIM, becoming disabled on a job causes the individual to quit the job and to be medically eligible for any disability benefits offered by employer-sponsored pension that cover that individual.

For details, see Appendix B, Sections B.15–B.16, beginning on page 112.

### 2.1.10 Retirement

Beginning at age 62.00000, the individual considers retiring, by which is meant withdrawing from employment (that is, quitting a job or quitting looking for a job) and claiming a social security retirement benefit. At each age between 62 and 68, the individual decides whether or not to defer retirement another year.

An individual always defers retirement at an age that is less than the social security early retirement age, which is passed to PENSIM from SSASIM. For an individual whose age is no less than the social security early retirement age, there is a probability of deferring retirement that is both age-specific and gender-specific. These probabilities produce the age pattern of social security retirement benefit claiming observed in recent data.

In the current version of PENSIM, the generosity of pensions on the current job or on past jobs does not affect the timing of retirement.

Defined-benefit pensions earned on prior jobs are claimed by disabled, widowed, and retired individuals at their earliest availability. Employed individuals are assumed to wait to claim defined-benefit pensions earned on prior jobs until there is no early retirement reduction in those pensions.

Defined-contribution and cash-account pension balances earned on prior jobs are transfered (with a certain probability) to a rollover account maintained by the individual. This rollover account is not accessed until the individual is older than the first non-penalty withdrawal age and is either disabled, widowed, or retired.

For details, see Appendix B, Section B.30, beginning on page 134.

### 2.1.11 Mortality

Just before each birthday (at age 0.99999 and every year after that), the individual is exposed to mortality risk. PENSIM includes several gender-specific period mortality tables for a number of years ranging from 1935 to 1997. SSASIM provides information on the rate of mortality improvement over of the course of the cohort's lifetime, which allows PENSIM to make annual adjustments to the male and female mortality tables in each year of a cohort individual's lifetime. The mortality probability to which a specific individual is exposed in a given year is drawn from the appropriate all-population mortality table for that year and then adjusted for the individual's education and disability status. In PENSIM the mortality rate rises to exactly one at age 124.99999.

For details, see Appendix B, Section B.35, beginning on page 137.

### 2.1.12 Validity of Simulation

Two different types of validation have been conducted on the simulated lives produced by PENSIM. First, the simulated distributions of certain life-history statistics have been compared with the actual distributions tabulated with historical data. And second, the simulated values of a number of projected aggregate social security statistics (derived from life histories simulated by PENSIM as described in Chapter 6) have been compared with projected statistics produced by other social security simulation models.

This first type of activity — comparing simulated estimates with known historical statistics — is rightfully viewed as validation. Strictly speaking, the second kind of activity is not validation because two future projections are being compared. It might be better to call this second activity something like cross-model comparison.

Over the years, a number of historical validation studies have been conducted on PENSIM output, especially with regards to disability, earnings, job tenure, and pension sponsorship. The results of these studies, which have been generally positive, are summarized in Chapter 8.

PENSIM has also been subjected to a wide range of cross-model comparisons. To date, PENSIM-derived SSASIM OLC-mode social security projections (see page 5) have been compared to social security projections produced by the Congressional Budget Office's CBOLT model (CBO 2004) and by the SSA Office of the Chief Actuary (Trustees' Report 2005). These cross-model comparisons are sensitive indicators of differences in the simulated lives of cohort individuals.

In these comparisons, PENSIM life histories (processed by SSASIM operating in the OLC mode) have been found to produce, over the years from 2000 to 2080, essentially the same annual population, taxable earnings, and disability benefits as projected by CBOLT and as shown in the Trustees Report. SSASIM OLC projections of retirement benefits are in line with those projected by CBOLT, but below those projected in the Trustees Report (Holmer 2006). Simulating current-law policy (that is, "scheduled benefits") using 2004 Trustees Report intermediate-cost demographic and economic assumptions, CBOLT produces a 75-year actuarial balance of –1.36 percent of taxable payroll (CBO 2004, page 31), SSASIM OLC mode (using PENSIM life histories) produces –1.35 percent, and the SSA Office of the Chief Actuary produces –1.89 percent (Trustees' Report 2004, page 13).

## 2.2 Simulation of Pensions

The methods used in PENSIM to simulate pension events are summarized in this section.

### 2.2.1 Plan Characteristics

At the start of each job held by an individual, PENSIM uses the simulated job characteristics to determine if the individual is covered by one or more pensions on this job. After assigning pension coverage to an individual, the detailed characteristics of each pension are simulated using a pension characteristics imputation model that has been estimated using 1996–98 plan data gathered from employers by the Bureau of Labor Statistics (Holmer and Janney 2003).

For details on pension assignment at the start of first jobs, see Appendix B, Sections B.17.2–B.17.3, beginning on page 115. For details on pension assignment at the start of subsequent jobs, see Appendix B, Sections B.19.2–B.19.3, beginning on page 122.

For more on the pension characteristics imputation model itself, see Appendix C, Section C.14, beginning on page 200.

### 2.2.2 Participation Behavior

An individual becomes eligible for a pension after meeting certain hours-worked, age, and job-tenure conditions specified in the pension plan. At the moment of eligibility in a conventional defined-contribution pension, an individual must decide whether or not to participate. In PENSIM, the defined-contribution eventual participation decision is based on probabilities that vary by age and relative earnings. If the individual does not participate in the plan, the eventual participation decision is revisited every ten years on this job.

In conventional defined-contribution plans, an individual who is simulated to eventually participate in the plan experiences a waiting time until the participation event. During the time between these two events, the individual does not participate in the plan. The participation event leads to an immediate active-participation event, which means that the individual makes active decisions about plan contributions and investments.

In defined-contribution plans with automatic-enrollment provisions, the eligibility and participation events occur at the same time, but there is a waiting time between the participation event and the active-participation event. During the time between these two events, the individual is a passive plan participant whose affairs are determined by the plan's default contribution rate and default investment.

For details, see Appendix B, Sections B.22–B.24, beginning on page 127.

### 2.2.3 Contribution Behavior

Active participants in defined-contribution plans have contributions simulated at the end of each calendar year and at the end of the job. The contribution event involves receiving returns on the prior year's investments, depositing contributions for this year in the plan account, and allocating the resulting assets among the different investment options available in the plan account. Some defined-contribution plans provide non-matching employer contributions, but most require employee contributions that may be matched in some way by the employer.

The contribution rate simulated for active participants is a function of age and relative earnings. In the current version of PENSIM, there is no variation around the mean contribution rate for each age-earnings group. These simulated contributions are subjected to a variety of time-varying government restrictions on earnings and contribution amounts. Of course, passive participants contribute at the plan's default contribution rate.

For details, see Appendix B, Section B.25, beginning on page 130.

### 2.2.4 Investment Behavior

The last step in the contribution event is allocating the total balance in the plan account across a set of investment options. Currently, PENSIM assumes that defined-contribution plans offer a very simple set of investment options: Treasury bonds and corporate equities (represented by the S&P 500 index). Company stock is also an investment option if the employer contribution is made in that form. PENSIM assumes that all defined-contribution plans require individuals to hold company-stock contributions until they are fifty-five years old, at which time they are assumed to continue holding the company stock. This lack of diversification assumption is consistent with data on defined-contribution asset allocation, but PENSIM has the capability of assuming that all company stock is sold at the diversification age.

Investment of defined-contribution plan assets by active participants is described by age- and gender-specific asset-allocation fractions. These fractions are assumed by the user of PENSIM rather than being empirically estimated as is the case for participation and contribution behavior. Of course, the plan balances of passive participants are assumed to be invested according to the plan's default asset-allocation rules.

For details, see Appendix B, Section B.25, beginning on page 130.

### 2.2.5 Accrual Calculations

The accrual logic for defined-contribution plan account balances has been described above. The accrual logic for conventional defined-benefit plans happens at the end of a job and the specifics of that logic depend on the plan type: terminal-earnings (i.e., final-average-pay), career-earnings, or dollar-amount. The accrual logic of cash-account defined-benefit plans is more like a defined-contribution plan except that contributions are made only by the employer at rates specified by the plan and the rate of return on the cash-account balance is specified by the plan. For all types of defined-benefit plans, the appropriate time-varying government restrictions on earnings and benefit amounts are applied.

At the end of a job the accrued benefit in both defined-benefit and defined-contribution plans depends in part on the individual's vesting status. In PENSIM, the vesting status at the end of the job is determined by the individual's job tenure and plan's vesting rules.

For details on the vesting event, see Appendix B, Section B.27, beginning on page 131.

### 2.2.6 Rollover Behavior

At the end of a job that produces a vested benefit in a defined-benefit plan, the individual places this benefit on a list of benefits to claim in retirement. The current version of PENSIM assumes no lump-sum distributions of vested benefits in non-cash-account defined-benefit plans.

At the end of each job that produces a vested balance in a defined-contribution or defined-benefit cash-account plan, the individual faces a decision about whether or not to transfer the balance into a rollover account maintained by the individual for retirement savings. Balances that are not transferred to the rollover account are assumed to be consumed by the individual and disappear from the simulation. The rollover probability is a function of the relative size of the balance.

For details on the rollover decision, see Appendix B, Section B.28, beginning on page 132.

The individual manages the rollover account at the end of each calendar year and at the time of a transfer. This rollover-management event involve receiving returns on the prior year's investments, handling any transfer, and allocating total assets among the different investment options available using the same asset-allocation fractions assumed for the the defined-contribution plan account.

For details concerning rollover account management, see Appendix B, Section B.29, beginning on page 133.

### 2.2.7 Withdrawal Behavior

Later in life, an individual faces a range of decisions regarding withdrawals of accrued pensions, which are represented by the rollover account, to which all saved defined-contribution balances have been transferred, and by the list of benefits earned under defined-benefit plans.

The timing of the initial rollover account withdrawal and the claiming of defined benefits is influenced by the interaction of government rules, plan rules, and the status of the individual.

In PENSIM, two types of withdrawals can be made out the individual's rollover account: an annual withdrawal to support the individual's retirement needs and a lump-sum withdrawal to purchase an annuity contract, the indexing and survivor attributes of which are determined by the interaction of parameters specified by the PENSIM user and the status of the individual. The PENSIM user specifies the fraction of the rollover account balance that is used to purchase an annuity at the initial withdrawal. Also, the PENSIM user can specify the size of annual withdrawals from any balance remaining after the annuity purchase. Any remaining balance in the rollover account is managed via the annual rollover-account-management event.

For details on the annuity-purchase event, see Appendix B, Section B.31, beginning on page 134.

Accrued defined-benefit pension benefits are contained on a list of vested benefits maintained by the individual. This list includes information about each benefit's early-retirement age, normal-retirement age, the benefit reduction rate applied when claimed before the normal-retirement age, and the terms of the joint-and-survivor benefit option.

For details on the pension-claim event, see Appendix B, Section B.32, beginning on page 135.

Once the initial rollover account withdrawal or the first defined benefit claim has been made, a pension-payment event occurs every year at the start of the year. At this event, the total amount of annuity payments, rollover account withdrawals, and defined benefit receipts are recorded as pension income for the individual. This pension-payment event occurs every year until the individual is dead.

For details on the pension-payment event, see Appendix B, Sections B.33, beginning on page 136.

### 2.2.8 Validity of Simulation

*The calibration of historical employer pension offerings in the years before 1996–98 and the validation of PENSIM estimates of pension income in retirement for the 1935 cohort are activities that are currently under way.*

For details on pension calibration, see Chapter 9, beginning on page 75.

For details on pension validation, see Chapter 10, beginning on page 77.

# 3 Policy Analysis with PENSIM

This chapter describes how PENSIM is used to analyze the effects of contemplated pension policy reforms. This description provides a context that makes the subsequent model design discussion easier to present and understand.

## 3.1 Nature of a Single Model Run

A PENSIM run involves the simulation of life histories for a large number of people in a birth cohort (and their spouses, who are not necessarily born in the same year and who are not considered members of the birth cohort sample). All simulated individuals experience demographic and economic events, the incidence and timing of which are determined by the stochastic realization (using Monte Carlo methods) of incidence and hazard functions estimated with recent cross-sectional and longitudinal data. The incidence and timing of events vary by age, gender, education, marital status, disability, employment status, etc. The types of life events that are modeled in PENSIM include demographic events (birth, migration, death), schooling events (leaving school at a certain age, receiving a certain educational credential), family events (marriage, divorce, childbirth), disability events, initial job placement, job mobility events (earnings increases while on a job, duration of a job, movement to a new job or out of the labor force), pension events (becoming eligible for plan participation, choosing to participate, becoming vested, choosing to rollover pension balances, choosing to buy an annuity, etc.), and retirement events.

This approach generates a wide variety of life histories for members of the birth cohort sample. The life histories vary for two major reasons. First, cohort members experience individual (idiosyncratic) risks that cause their life events to differ from others in the cohort. This is the stochastic element that traditional dynamic microsimulation models have always included. Individual risks simply generate the variation among cohort members as they age. Second, cohort members experience collective (systemic) risks if the demographic or economic environment within which they age is subject to uncertainty. In the case of defined-contribution pension plans, for example, the rate of return received on pre-retirement pension assets and the level of interest rates at retirement when converting plan assets into an annuity are two major collective risks.

Because one policy, or type of pension, may cushion individuals from collective risks better or worse than another policy, it is essential for PENSIM to represent explicitly these collective risks. This is done by specifying a PENSIM run to consist of a number of scenarios that represent variations in the environment (i.e., collective risks), where each scenario consists of a number of simulated life histories to represent individual variation within the cohort (i.e., individual risks). This joint sampling scheme produces a collection of life histories that represents the combined effects of collective and individual risks.

All the life histories simulated for each scenario in a PENSIM run are generated using the same government-policy, pension-offering, and employee-behavior assumptions. A variety of pension-relevant statistics are tabulated for the run. Analysis of a contemplated policy reform is accomplished by comparing the statistics from two runs whose input specifications are identical in all respects, except for policy parameters that characterize the reform and possibly some pension-offering or employee-behavior parameters that are directly affected by the proposed reform.

## 3.2 Types of Simulated Pension Statistics

Given the simulated life histories of cohort sample members, PENSIM can tabulate a variety of pension and employment statistics. Cross-sectional statistics (e.g., the fraction of individuals who are 40 years old and work for an employer that sponsors a pension plan) provide information on the distribution of pension-relevant attributes within the cohort sample at a specified age. Longitudinal statistics (e.g., the proportion of middle-aged men who change jobs over a nine-year period) provide information on the degree of continuity in people's life histories. Lifetime statistics (e.g., the fraction of individuals who work less than ten years of their career for employers that sponsor a pension plan) provide information on the distribution of pension-relevant attributes over the lifetime of cohort sample members.

The ability to tabulate both cross-sectional and longitudinal statistics, as well as lifetime statistics, is extremely useful. It allows PENSIM results to be validated through the direct comparison of age-specific statistics tabulated from actual cross-sectional survey data and between-age statistics tabulated with actual longitudinal survey data, and, at the same time, permits PENSIM results to be relevant, because most policy questions focus on lifetime results.

The current version of PENSIM represents several types of defined-benefit and defined-contribution pensions and the logic required to compute pension coverage and benefit adequacy for each of those types of pension plans. The model is, therefore, able to tabulate a full range of cross-sectional, longitudinal, and lifetime pension coverage statistics. The simulation of defined-contribution plans includes employee behavior related to participation, contributions, asset allocation, rollover and lump-sum distribution behavior when changing jobs, and retirement withdrawal and annuitization behavior. This means that PENSIM can produce pension benefit adequacy statistics (e.g., replacement rates or expected lifetime present values) similar to those routinely used to describe social security benefit adequacy.

Each of these cross-sectional, longitudinal, and lifetime pension statistics can be calculated for each individual member of the birth cohort sample in a particular scenario. These individual statistics can then be used to compute summary statistics (such as the mean, a specified percentile value, or the variance) for each scenario. The variability of a summary statistic across scenarios measures its sensitivity to the collective risks represented in the scenarios. Of course, if the PENSIM run specifies no collective risks, then the variation in the summary statistic across scenarios simply represents the sampling variability produced by individual risks. This approach of computing probability distributions for scenario summary statistics permits the quantification of a policy's ability to cushion individuals against collective risks.

The importance of being able to compute lifetime pension statistics is clear when one considers recent discussions concerning, for example, the effects of employers converting their traditional defined-benefit plans into cash-account plans. Estimates of the impact of this kind of plan conversion that use a cross-sectional sample to analyze current employees are biased because current employees under represent short-job-duration employees, who are likely to be advantaged over their lifetimes by such a plan conversion. Current employees over represent long-job-duration employees, who are most likely to be disadvantaged by a conversion to a cash-balance plan. This means that lifetime statistics for a cohort are needed to estimate correctly the fraction of all workers who are advantaged or disadvantaged by such a plan conversion. PENSIM has been used to study the effects of such conversions by the Government Accountability Office (GAO 2005).

## 3.3 Estimating a Policy Reform's Effects

Because a single PENSIM run produces the implications of one set of assumed government-policy, pension-offering, and employee-behavior parameters for a birth cohort sample, one model run cannot be used to estimate the implications of a reform in pension policy or the implications of an assumed change in pension offering or employee behavior. Policy reform analysis requires two model runs. The first run's input parameters typically represent current-law government policy, pension-offering and employee-behavior parameters consistent with current-law policy, and assumed collective and individual risks. The second run's input parameters are the same as the first run's except that they represent the reform's change in government policy as well as the reform's direct effects on pension offering and employee behavior. The impact of the reform, relative to current-law policy, is measured by the difference in pension-related statistics produced by the two runs. Notice that the reform could be in effect over the whole lifetime of the cohort individuals or could be assumed to be implemented in the midst of their work careers.

It is important to notice the difference between such a policy analysis simulation model and a dynamic simulation model that is developed to project the pension outcomes of a particular birth cohort under current-law policy. Such projections can yield interesting results, but they say nothing about how the lifetime pension experience of that cohort would change under a different pension policy regime, or whether the lifetime pension experience of other birth cohorts would be much different under the same pension policy regime.

The goal of the PENSIM development effort is to be able to estimate lifetime pension statistics for a birth cohort under current-law pension policy and under a variety of reform policies that affect employer pension offerings or employee pension behavior. The complete effect of a pension policy regime will be shown only by use of statistics that incorporate the full lifetime pension experience of individuals. PENSIM has the flexibility to simulate different kinds of cohorts, spanning the range of historical cohort experiences (with regards to mortality and employment patterns, for example). This flexibility allow us to determine how sensitive the policy effects are to historically changing cohort behavior.

The flexibility to represent a range of different historical birth cohorts has been found feasible and useful in the work conducted at Statistics Canada (Wolfson et al. 1997). PENSIM is implemented in a manner that permits the scaling of each behavioral parameter used as model input. Adjusting these input parameters will produce different kinds of mortality, schooling, family, and job mobility patterns. The parameters can be adjusted to produce a synthetic cohort whose lifetime experience resembles the experience of any particular historical cohort. See Chapter 7 for a discussion of our experience calibrating PENSIM input parameters to generate cohort samples that represent the cohorts born in 1935, 1955, 1970, and 1985.

## 3.4 Estimating a Pension Reform's Cost

Some readers may have already asked themselves the following question: if PENSIM focuses on simulating the lifetimes of a birth cohort rather than on simulating the whole population from year to year, how can the cost of a pension reform be estimated? The answer is that a single PENSIM run is not well suited to estimate a pension reform's annual budgetary cost to government or its annual compliance cost to employers. (We include in compliance costs both the additional administrative and funding costs associated with a pension reform.)

However, it is possible to simulate many birth cohorts (each one in a separate PENSIM run) and then to overlay the yearly results for each cohort so that they add up to population-wide totals for each calendar year. This method has been used successfully for years in the SSASIM Overlapping Cohorts (OLC) mode of operation to produce aggregate annual social security trust fund revenue and cost estimates, as described in Chapter 6.

But still, PENSIM does not generate government costs associated with employer-sponsored pensions, which would consist mainly of income tax expenditures on employee contributions to defined-contribution pension plans. While PENSIM can estimate aggregate contributions, it does not simulate the taxation of employee income, which is needed to estimate the tax expenditure.

The employer compliance cost can be estimated by PENSIM for all employers offering each of the several pension types. For details on the .cst output results file, see Chapter 4, Section 4.4, beginning on page 27.

But the compliance cost for individual employers cannot be estimated in PENSIM. In order to do this, a model would need to include linked samples of employers and employees. Such a model would permit the estimation of compliance costs for an employer by considering the particular pension offering and workforce characteristics of that employer. We are not aware of any simulation model that has these capabilities.

# 4 Logical Structure of PENSIM

The high-level structure of PENSIM is modular. The philosophy of modular structure rejects the idea of implementing a simulation model as a single computer program that provides a large number of very different kinds of services: providing access to input data, calculating simulation results, and summarizing or visualizing output results. Instead, modular philosophy calls for breaking up the model into a number of separate computer programs, each one of which specializes in providing one kind of service.

The main advantage of a modular model is the increase in productivity that flows from the specialization allowed by the division of labor among component programs. Data-handling tasks can be assigned to a relatively inexpensive, mass-marketed database management program that has superior data retrieval and ease-of-use features. Highly repetitive Monte Carlo simulation calculations can be performed efficiently by a compiled program that has been custom designed to implement the simulation logic. Statistical analysis and visualization of simulation results can be done with easy-to-use, relatively inexpensive, mass-marketed spreadsheets or statistical packages.

This modular approach to developing simulation models goes well beyond the recommendations of a National Research Council panel that reviewed the design of policy simulation models (Citro and Hanushek 1991, Cotton and Sadowsky 1991).

The modular structure of the PENSIM pension policy simulation model is represented in Figure 1. The model consists of two sets of disk files and three separate kinds of computer programs. The two sets of disk files are: the input database tables and the output results files for each model run. The three computer programs are: a user interface, a stochastic simulator, and output analyzers.

## 4.1 User Interface Program

The model's user interface program permits manipulation of the contents of the input database, which is organized as a relational database (see below). The user interface program also provides access to several database utility programs (not shown in Figure 1) that perform complex database operations, such as merging two separate databases into one. Because the user interface program also provides access the stochastic simulator, the output results files, several output analyzer programs, and different types of online help, it serves

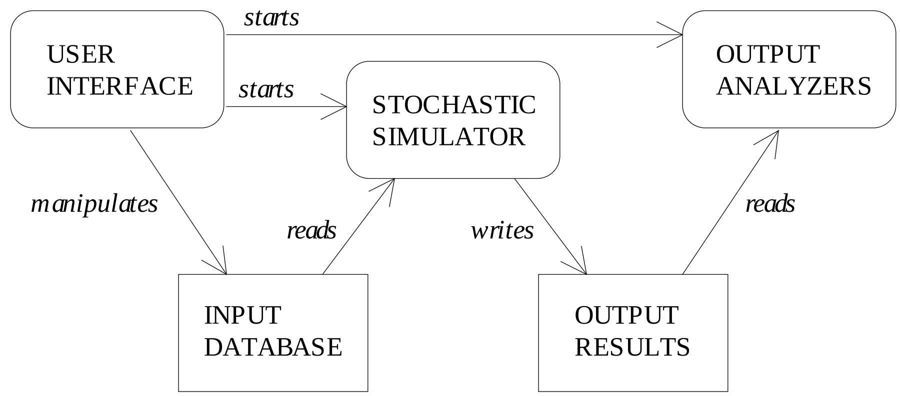


Figure 1: ***Modular Structure of the Model***. *This modular structure allows specialization among component programs and, if execution speed on a single computer chip is too slow, provides a development path to a multi-threaded architecture that permits distributed processing of model runs on multiple-chip computers.*

as the PENSIM working environment.

The user interface program is a free-standing computer program, which is developed using Microsoft's Visual FoxPro, but does not require a copy of FoxPro to run. It has a modern look and feel with pull-down menus and dialog boxes, multitasking, and close integration with standard office productivity applications. The structure of the menus is virtually identical to that of the SSASIM user interface program, which is described in detail in Holmer (SSASIM Guide, Section 3).

## 4.2 Input Database Tables

The model's input database is organized as a relational database, a collection of linked database tables. Relational databases facilitate the organization of complex data on the attributes of and relationships between several different kinds of objects in a way that is intuitive, efficient, and flexible. The advantages of relational principles for database design and implementation are widely appreciated in business, statistical, and scientific circles. For a technical survey of relational databases by one of the developers of the concept, see Date (1999).

The input database tables are stored in a FoxPro DBF format which can be easily imported into or exported from standard office productivity applications.

## 4.3 Stochastic Simulator Program

The model's stochastic simulator program reads a run's assumed policy and behavior parameters from the input database, performs the requested run, and writes the run's detailed simulation output and summary results to text files. Each model run consists of one or many Monte Carlo scenarios. The simulator processes one or more scheduled model runs automatically in a non-interactive manner. Ill-specified model runs generate detailed error messages in an error-log file and are aborted, but this does not prevent the processing of other scheduled runs. A complete transcript of a long series of runs can be saved to a run-log file rather than being shown on the screen.

The stochastic simulator is a compiled, computationally efficient computer program written in the C++ language. It is organized in a modular fashion using C++ classes that correspond to real-world entities whose interaction is being represented in the model. For a description of some of the C++ programming styles and idioms used in the simulator, see Coplien (1992). In particular, object-oriented programming methods and template classes are used to reduce the size and complexity of the simulator's source code.

## 4.4 Output Results Files

The model's output results are written to text files that are formatted to facilitate the files being easily processed by or imported into a wide variety of software for post-simulation output analysis. A set of output results files is produced by the stochastic simulator during the course of each model run, with each file named using the number of the run that created the output results. This enables those using the model to save the output results from many different runs and to compare results across runs long after the model runs were processed by the simulator.

The output results files are named runNNNNN.xxx, where NNNNN is the PENSIM run number (padded with leading zeros) and xxx is the three-letter extension that identifies each output results file.

Each simulation run produces the output results files called for in the specification of that run. The output results files that can be requested are as follows:

**.arc** contains revenue and cost statistics for the annuity provider so that the solvency of the annuity provider can be determined;

**.cov** contains lifetime pension coverage statistics for each individual in the cohort sample;

**.cst** contains the aggregate employer cost (expressed as a present value relative to the present value of eligible employee earnings) for each of seven types of pension plans;

**.leh** contains *binary* annual earnings for each individual in the cohort sample;

**.pen** contains pension benefits at each age for each individual in the cohort sample;

**.pro** contains end-of-job pension rollover statistics for each individual in the cohort sample;

**.xss** contains information gathered in the cross-sectional-survey event for each individual in the cohort sample.

The detailed contents of each of the above output results files can be accessed using the Results File Documentation item on the Output menu of the PENSIM user-interface program.

The annual pension benefits are also passed to GEMINI in the cohort (coh) file produced by PENSIM for GEMINI input. It is often more convenient to ask PENSIM to produce a cohort file and process it with GEMINI — even if you are not interested in the social security results produced by GEMINI — because the GEMINI .sum and .adq output files have certain advantages. For information on the GEMINI output results files and their post-simulation analysis, see Holmer (GEMINI Guide).

## 4.5 Output Analyzer Programs

The model's output analyzer program can be any computer program that is appropriate for the desired post-simulation analysis of the output results. The output results files are structured to be easily processed by or imported into a wide variety of software. The resulting flexibility in the selection of an analyzer program ensures that statistical analysis of the simulation output, visualization of the results, and preparation of presentation graphics can be conducted with software that is both well suited to the task and familiar to the person using the model.

Although several output analyzer programs have been developed to facilitate routine analysis of output results, it should be noted that many other options are available. Experience shows that one can never fully anticipate what post-simulation analysis will be required, and that the need for ad hoc analysis arises often. The approach adopted here — writing detailed output from each model run to a set of files, formatted in a way that allows them to be imported into or processed by a wide variety of software — permits a flexible combination of anticipated routine analysis (using one of the existing analyzer programs, for example) and unanticipated analysis (via custom-developed spreadsheets or tabulation programs, for example). Typically, ad hoc analysis of small summary results files involves importing them into a spreadsheet, while the larger detailed output results files, which make up the bulk of PENSIM output, are more easily tabulated using statistical software or short AWK programs (Aho, Kernighan, and Weinberger 1988). For more on the AWK programming language, which is available as part of the PENSIM installation, see Holmer (GEMINI Guide, Section 4).

# 5 PENSIM Stochastic Simulator

The stochastic simulator performs three types of tasks. One set of tasks involves simulating a sequence of events over the lifetime of a single individual. These tasks are referred to as *lifetime simulation tasks*. Another set of tasks involves using Monte Carlo methods (Hammersley and Handscomb 1964) to produce a sample of simulated individual lifetimes, which we call an individual sample. The variability of lifetime experiences in the individual sample represents the effects of the individual (or idiosyncratic) risks facing cohort members. These tasks are referred to as *individual-risk simulation tasks*. And a third set of tasks involves producing a number of scenario simulations in which Monte Carlo methods are used to produce a set of individual samples, the variability of which represents the effects of the collective (or systemic) risks facing cohort members. These tasks are referred to as *collective-risk simulation tasks*. A single PENSIM run requires the simulator to perform all three types of tasks.

This chapter discusses the procedural logic of the stochastic simulator and then its structure. The description of both is introductory, just sufficient to support the detailed discussion of PENSIM's behavioral events contained in Appendix B.

## 5.1 Logic of Stochastic Simulator

The procedural logic of the PENSIM stochastic simulator is discussed at three levels, corresponding to the three types of simulation tasks described above. Lifetime simulation tasks are described first, followed by the individual-risk simulation tasks that produce a single individual sample, and the collective-risk simulation tasks that produce individual samples, which form the cohort sample when combined together.

### 5.1.1 Logic of Lifetime Simulation Tasks

The timing of events during an individual's lifetime is simulated in PENSIM using the event scheduling method of discrete event simulation (Fishman 1978, for example). This method involves the use of a simulation monitor, a simulation clock, and an ordered list of pending simulation events. A simulation event consists of an age and the activities that occur at that age.

A lifetime simulation begins with the event list empty except for the birth event. The lifetime is simulated by the monitor repeating the following sequence of steps. First, the monitor retrieves the next pending event from the list. If there are no pending events, the monitor concludes the lifetime simulation. Next, the monitor advances the clock to the age at which the retrieved event is scheduled to occur. Then, the monitor executes that event's activities, which include changing one or more individual attributes and possibly scheduling future events in the individual's life. Finally, after the event's activities are completed, the monitor repeats this same sequence of steps.

Demographic events play an important role in the individual life simulation. The birth event's activities specify an individual's gender at birth, schedule the high-school-start event at age fourteen, and schedule a potential death event at year end. Mortality rates vary not only by age and gender, but also by disability and educational status. When an actual death event occurs, lifetime statistics are calculated for the individual and all pending events are removed from the list.

The basic sequence of events in PENSIM is being born, attending school and getting a first job, moving from one job to another over a work career, retiring, and collecting any earned pensions. Family, migration, disability, and death events can intervene at any age during a lifetime.

### 5.1.2 Logic of Individual-Risk Simulation Tasks

A sample of individual lifetimes is generated using Monte Carlo methods in order to represent the effects of individual risks. The resulting individual sample will, therefore, contain a wide variety of lifetime experiences, with some individuals living longer than others, some becoming disabled, some rich, others poor. Sample statistics are then computed from the individual statistics. A sample statistic may be the mean of an individual statistic (to measure the central tendency), or the fraction of individuals whose statistic exceeds an certain value (to measure the probability of a sampled individual having an extreme experience), or the variance of an individual statistic (to measure the variability of experience among sampled individuals).

Because the individual sample is generated with all the collective risk factors having the same value, one sample of individuals cannot represent the variability in lifetime experiences caused by collective risks. The tasks involved in generating many scenarios, in order to produce a collection of individual samples, is described next.

### 5.1.3 Logic of Collective-Risk Simulation Tasks

Collective risks are represented by a number of variables (such as equity returns, interest rates, and inflation rates) whose values differ across simulation scenarios and affect all members of the birth cohort. The application of Monte Carlo simulation methods at this level of the PENSIM simulation involves realizing a scenario projection of collective risk factors, conducting the individual-risk simulation task described above, and recording the values of the resulting sample statistics for that scenario. After repeating these steps for a specified number of scenarios, the values of the individual sample statistics can be summarized using measures of central tendency and dispersion that express the effects of collective risks on the members of the birth cohort.

It is important to note that older dynamic microsimulation models of pension policy do not represent such collective risks, but rather generate individual sample statistics for a single, pre-specified scenario as described by Ross (1991) and Citro and Hanushek (1997, pages 199–212). Given the growing importance of defined-contribution pension plans, uncertainties in the capital markets have important implications for pension statistics (e.g., market returns for plan asset growth and interest rates for annuity prices). No model seeking to analyze the implications of alternative pension policies can afford to ignore the variation in lifetime experiences caused by these collective risks.

## 5.2 Structure of Stochastic Simulator

The structural relationships among the real-world entities whose interactions are simulated over an individual's lifetime are represented in Figure 2. These real-world entities and relationships are mirrored in the modular (or class) structure of the C++ program that implements the logic of the PENSIM stochastic simulator. Standard object-oriented programming techniques are used in the simulator to represent these entities and the relationships between them.

The PENSIM stochastic simulator's individual simulation tasks involve the interactions of five major entities: employers, jobs, individuals, pensions,

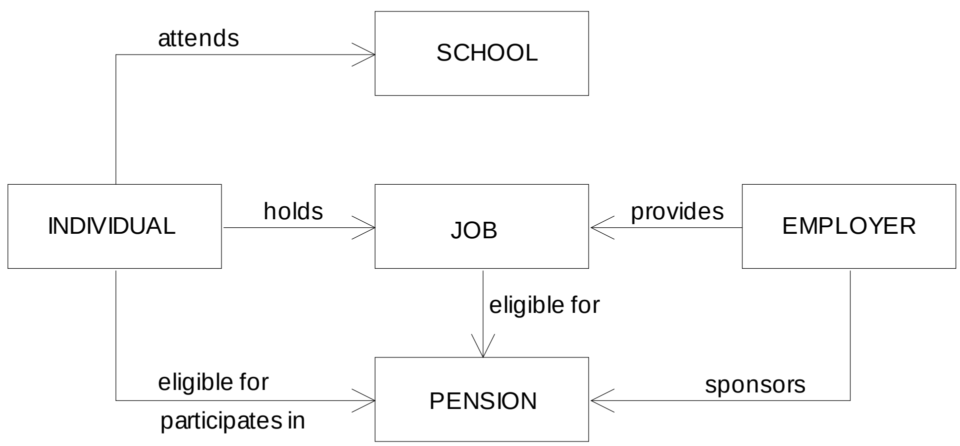


Figure 2: ***Structure of the PENSIM Stochastic Simulator's Lifetime Simulation***. *This entity-relationship diagram shows the major entities represented in the simulator's individual lifetime simulation as boxes and the relationships between the entities as arrows. The modeled attributes of each entity are described in the text.*

and schools. The attributes of these entities that are modeled in PENSIM are listed below:

- INDIVIDUAL: age, gender, schooling, family history,
  disability, job history, pension participation.
- SCHOOL: high school, type of college, graduate school.
- JOB: unionization, hours, earnings.
- EMPLOYER: size, industry, pension offerings.
- PENSION: type, rules (eligibility, vesting, benefits, etc.).

A cohort sample individual's family history is represented by two lists: (1) a list of spouses, each of whom has a complete simulated life history, and (2) a list of children.

A more detailed discussion of how the events simulated in PENSIM specify the attributes of these entities is provided in Appendix B.

# 6 PENSIM Links to SSASIM and GEMINI

The Policy Simulation Group maintains two other policy simulation models in addition to PENSIM. These other simulation models are used by PENSIM to provide cohort environment information required for its collective-risk simulation tasks and to provide its pension estimates with social security context. SSASIM is used to provide information about macroeconomic and macrodemographic scenarios (i.e., the variables that characterize collective, or systemic, risks) required to simulate pension accumulation and withdrawal. GEMINI is used to provide social security context for PENSIM pension results.

In this chapter, the SSASIM and GEMINI models are described briefly, and the nature of the links between these two models and PENSIM are outlined.

SSASIM is a cell-based actuarial model of social security that produces estimates of the aggregate financial condition of the social security program and estimates of benefits and money's worth statistics for a few exemplary individuals.

See Holmer (SSASIM Guide) for a more complete description of that model.

The SSASIM model can be run by other models in an engine mode that produces a binary data file containing annual time series for several macroeconomic and macrodemographic environment variables. These time series can be produced using VAR(2) or structural time-series models. The resulting environment file (named envNNNNN for SSASIM run NNNNN) can be read by other simulation models that require information on the cohort environment.

GEMINI is a dynamic microsimulation model that produces annual social security benefits for each individual life history (including all spouses) in sample of people born in a certain year. Notice that GEMINI needs information on individual life histories in order to calculate social security benefits. This information on individual life histories for the cohort sample is produced by PENSIM running in engine mode (typically assuming no employer-sponsored pensions in order to speed execution) and these life histories are written to a binary cohort file (named cohNNNNN for PENSIM run NNNNN). When PENSIM executes in full model mode, employer-sponsored pensions are simulated and the resulting cohort file contains pension information that is passed through to the standard GEMINI output files.

See Holmer (GEMINI Guide) for a more complete description of that model.

The ability of these three models to work together has allowed the creation of the SSASIM Overlapping Cohorts (OLC) mode of operation. When operating in OLC mode, SSASIM calls GEMINI repeatedly to produce a relatively small (typically one-in-a-thousand) sample for each birth cohort born from 1935 to the simulation end, which is usually 2080 or later. Those cohort samples are then used to compute annual aggregate social security benefits and taxes, which can be use to estimate standard social security solvency statistics. In other words, the SSASIM OLC mode operates like a dynamic microsimulation model of social security: trust-fund statistics are produced by aggregating individual estimates of benefits and taxes (rather than by using cell-based actuarial methods).

This microsimulation methodology has the advantage of using a common set of earnings histories to estimate both taxes and benefits, which enforces a logical consistency that is sometimes missing in actuarial methods.

In addition, the OLC mode of SSASIM operation has the attractive property of producing aggregate solvency estimates that are guaranteed to be logically consistent with distributional estimates tabulated from GEMINI samples. And, because the models can be specified to generate cohort samples of any size, there is flexibility to specify relatively small GEMINI samples for all the birth cohorts used to produce aggregate estimates and to specify relatively large GEMINI samples for the few birth cohorts used in distributional analysis.

The relatively fast execution speed of PENSIM and GEMINI imply that a high-end personal computer can execute a single-scenario OLC-mode run in about ten or fifteen minutes. Faster execution times are possible on multiple-CPU computers because the SSASIM OLC mode has been programmed to allow multi-threading of the GEMINI runs for each birth cohort (of which there are at least 150). So, for example, on a dual Intel CPU computer using Intel's hyper-threading technology, running four GEMINI runs simultaneously allows the execution of a single-scenario OLC-mode run in about five minutes. The new AMD Opteron chip is not only faster at single-threaded floating-point arithmetic than an Intel CPU, but has two complete CPUs on one chip. Using a computer with two Opteron 280 chips (containing four CPUs), a single-scenario SSASIM OLC-mode run takes about two and one-half minutes to execute when current-law social security policy is being simulated with no employer-sponsored pensions.

# 7 Calibration of PENSIM Lives

This chapter of the report describes the assumptions and methods used to produce samples of complete life histories for a several birth cohorts. To date, we have calibrated behavior for the 1935, 1955, 1970, and 1985 birth cohorts. As the need arises, other birth cohort can be added to this list using the methods described in this chapter.

The purpose of this calibration exercise is to adjust the life histories of those born in different decades so that they reflect the cohort experiences that have actually occurred in the past or that are predicted to occur in the future. In PENSIM, cohorts born in other years are assumed to have calibrations similar to those of the closest of these four calibrated birth cohorts.

This chapter is organized as follows. First, we identify key descriptive statistics that are the focus of the calibration work. Then we consult a variety of sources to specify target values for each statistic for each cohort. After setting these calibration targets, we describe which PENSIM input parameters are adjusted in order for the PENSIM sample to match the calibration targets. And finally, we present the results of the calibration exercise for each of these four birth cohorts.

## 7.1 Calibrated Behavior

We describe the range of behavior that is calibrated to be different across the birth cohorts.

### 7.1.1 Life Expectancy

The gender-specific period life expectancies at birth and cohort life expectancies both at birth and at age sixty-five are adjusted to reflect the longer life expectancies of younger cohorts.

### 7.1.2 Educational Attainment

The gender-specific educational attainment distribution is altered for each cohort to reflect the lower level of schooling in older cohorts.

### 7.1.3 Employment Pattern

Female employment rates are adjusted to reflect the substantial rise in labor force participation of women in younger cohorts. Male employment rates are adjusted slightly to reflect a small decline in participation among men in younger cohorts.

### 7.1.4 Earnings Level and Inequality

In PENSIM, the average level of earnings is automatically scaled each calendar year by the social security average wage index, which is simulated in SSASIM and read by PENSIM from the cohort environment file (see Chapter 6 for details). Therefore, inside PENSIM earnings are represented relative to average annual earnings.

The degree of earnings inequality has changes markedly across cohorts over the past half century. The gender gap in earnings has narrowed substantially while the education gap for men has widened substantially. In addition to calibrating overall earnings levels, we calibrate parameters that change the gender and education gaps in average earnings.

### 7.1.5 Marital Status

The all-gender marital status distribution at age sixty is adjusted to reflect the rising incidence of divorce and changes in marriage behavior among younger cohorts. No (re-)marriages or divorces are assumed to occur after age sixty.

### 7.1.6 Uncalibrated Behavior

Some behavior may change across the PENSIM runs for the four birth cohorts even though PENSIM input parameters for that behavior are not different. This is because some behavior depends on an individual's education, employment, or other behavior that differs across cohorts. One example of this uncalibrated, but yet changed, behavior is disability.

## 7.2 Calibration Targets

Next we present the target values of the key descriptive statistics identified above for each of the four birth cohorts that have been calibrated.

Table 1: ***Life Expectancy Calibration Targets***

| Gender – Life Expectancy Type | Life Expectancy in Years 1935 Cohort | 1955 Cohort | 1970 Cohort | 1985 Cohort |
|---|---|---|---|---|
| Female – period at birth | 63.2 | 72.8 | 74.9 | 78.2 |
| Female – cohort at birth | 74.4 | 79.1 | 81.0 | 82.9 |
| Female – cohort at age 65 | 19.6 | 20.6 | 21.4 | 22.2 |
| Male – period at birth | 59.4 | 66.7 | 67.2 | 71.1 |
| Male – cohort at birth | 67.7 | 72.9 | 75.5 | 77.9 |
| Male – cohort at age 65 | 16.2 | 17.5 | 18.3 | 19.1 |

#### 7.2.1 Life Expectancy

Calibration targets are specified for both period and cohort life expectancies at birth, and for cohort life expectancy at age sixty-five, for both women and men. The period life expectancies are known facts for the four cohorts under consideration, but the the four cohort life expectancies depend on assumptions regarding future rates of mortality decline. Period life expectancies at birth by gender are those tabulated by the Social Security Administration's Office of the Chief Actuary (Bell et al. 1992, Table 9, pages 94–95). Cohort life expectancies are those assumed by the same organization in their 2001 intermediate-cost projection (Trustees' Report 2001, page 78). Because the cohort life expectancies at birth are not shown for 1935 in the Trustees' Report, we estimated those values by multiplying the old 1935 cohort life expectancy estimate from Bell et al. (1992) by the ratio of the new and old estimates for the 1940 birth cohort, which is shown in the Trustees' Report. These calibration targets are shown in Table 1.

#### 7.2.2 Educational Attainment

The PENSIM run for the 1955 birth cohort, which uses SIPP-estimated education parameters, closely replicates the gender-specific educational attainment distributions for those aged 35–44 surveyed in the March 1997 Current Population Survey (CPS). We are not aware of any widely-recognized projections of ultimate educational attainment for cohorts born since 1955, so we assume that the estimated education parameters are unchanged into the

Table 2: ***Educational Attainment Calibration Targets***

| Gender – Educational Attainment | Percent with Attainment 1935 Cohort | 1955 Cohort | 1970 Cohort | 1985 Cohort |
|---|---|---|---|---|
| Female – h.s. dropout | 24 | 11 | 11 | 11 |
| Female – h.s. graduate | 45 | 32 | 32 | 32 |
| Female – college attendee | 17 | 31 | 31 | 31 |
| Female – college graduate | 9 | 16 | 16 | 16 |
| Female – graduate degree | 5 | 10 | 10 | 10 |
| Male – h.s. dropout | 28 | 12 | 12 | 12 |
| Male – h.s. graduate | 34 | 34 | 34 | 34 |
| Male – college attendee | 16 | 27 | 27 | 27 |
| Male – college graduate | 13 | 16 | 16 | 16 |
| Male – graduate degree | 9 | 11 | 11 | 11 |

future. In fact, the educational distributions in the PENSIM run for the 1955 birth cohort are similar to those reported for the 1956–60 birth cohorts included in the MINT project (Toder et al. 1999, page 34). On the other hand, there is evidence that educational attainment for the 1935 birth cohort is substantially different. We use as calibration targets for this older cohort, the MINT distributions for the 1931–35 cohorts (Toder et al. 1999, page 34). These calibration targets are shown in Table 2.

### 7.2.3 Employment Pattern

Employment pattern calibration targets are based on a review of employment to population ratios in each year from 1948–2000 for those aged 25–34, 35–44, 44–54, and the combined age group of 25–54. These four time series were obtained from the Selective Access feature of Bureau of Labor Statistics' "Labor Force Statistics from the Current Population Survey" web page ⟨http://stats.bls.gov/cpshome.htm⟩. Examination of these series shows a relatively steady rise in female employment rates over the 1948–2000 period with the magnitude and timing of the increases approximately the same across the three age subgroups. Employment rates for women approximately doubled over this time period. Over this same time period, male employment

Table 3: ***Employment Pattern Calibration Targets***

| | Employment Rate Index | | | |
|---|---|---|---|---|
| Gender | 1935 Cohort | 1955 Cohort | 1970 Cohort | 1985 Cohort |
| Female | 50 | 90 | 100 | 100 |
| Male | 105 | 98 | 100 | 100 |

rates were relatively steady (apart from business cycle effects) except for the relatively small decline that occurred during the 1970s. Given these patterns, we use the CPS employment-population ratios for the 25–34 age subgroup in 1965, 1985, and 2000 to calculate an employment rate index that has a value of 100 in 2000 for the 1970 birth cohort. We assume that the employment pattern of those born in 1985 is the same as for those born in 1970. The employment rate index values calculated from the CPS data are shown in Table 3.

### 7.2.4 Earnings Level and Inequality

Tabulation of March 1997 CPS data show that, among those with positive annual earnings during 1996, average annual earnings is 36.68 thousand dollars for those who are age fifty (1513 observations) and 37.18 thousand dollars for those who are age fifty-one (1149 observations). Assuming that people's birthdays are uniformly distributed across the calendar year, we estimate that average annual earnings during 1996 for those who are fifty years old in 1996 is 36.78 thousand dollars.

We rely on extensive tabulations by CBO staff of Panel Study of Income Dynamics data for calibration targets on the changing degree of earning inequality across gender and education groups (Harris and Sabelhaus 2003, Figures 13–16).

### 7.2.5 Marital Status

The all-gender distribution of marital status at age sixty for the four birth cohorts is obtained from unpublished projection results provided by the Social Security Administration's Office of the Chief Actuary. These results, which

Table 4: ***Marital Status Calibration Targets***

| All-Gender Marital Status | Percent with Status at Age Sixty 1935 Cohort | 1955 Cohort | 1970 Cohort | 1985 Cohort |
|---|---|---|---|---|
| Never married | 4.6 | 9.5 | 12.5 | 12.7 |
| Married | 75.1 | 69.9 | 69.5 | 69.6 |
| Divorced | 11.6 | 15.1 | 13.1 | 13.1 |
| (married ten years) | (7.6) | (8.8) | (7.5) | (7.5) |
| (married shorter) | (4.0) | (6.3) | (5.6) | (5.6) |
| Widowed | 8.7 | 5.5 | 4.9 | 4.6 |

are details of the 2001 intermediate-cost projection (Trustees' Report 2001), contain population estimates by age and marital status for each year from 1941 until 2081. These projection results combine divorcees whose previous marriage lasted ten years or more and divorcees whose prior marriage lasted less than ten years. The relative size of these two groups of divorcees can be estimated using results from the MINT project (Panis and Lillard 1999, page 43). Using the reported results and the fact that divorce rates in MINT are simulated to increase over time, we assume the fraction of divorcees whose prior marriage lasted less than ten years is 35 percent for the 1935 birth cohort (about the MINT value for those born in 1931–40), 42 percent for the 1955 birth cohort (about the MINT value for those born in 1951–60), 43 percent for the 1970 birth cohort, and 43 percent for the 1985 birth cohort. The resulting calibration target marital status distributions are shown in Table 4.

## 7.3 Calibration Methods

We describe the PENSIM input parameters that are adjusted so that the simulated sample exhibits descriptive statistics that match the calibration targets.

### 7.3.1 Life Expectancy

Mortality rates for the cohort's birth year are used as input for each of the four birth cohorts. These four mortality rate matrices (indexed by age and gender) are derived (in some cases by linear interpolation) from period life tables (Bell et al. 1992). The eight PENSIM input parameters that scale mortality rates by gender and education are adjusted so that the gender cohort life expectancies, given the aggregate mortality decline rates generated by SSASIM, match the calibration targets.

### 7.3.2 Educational Attainment

Simulated educational attainment is determined in PENSIM by the interaction of a number of hazard functions that predict the waiting time between starting and finishing high school and college, and a number of duration-specific probabilities of obtaining an high school or four-year college diploma, as described in Appendix C, Section C.2).

Our strategy for calibrating the educational attainment distributions for men and women is as follows: (1) Adjust proportionally PENSIM input parameters SCHHSF.diplomap_f and SCHHSF.diplomap_m for those finishing high school at age 18 and 19 so that the simulated fraction of high school dropouts matches the calibration targets. (2) Adjust proportionally the SCHCOS.hazardr_f and SCHCOS.hazardr_m parameters for those high school graduates who start college zero or one year after graduating so that the fraction of college attendees matches the calibration targets. (3) Adjust proportionally the SCHCOF.diplomap_f and SCHCOF.diplomap_m parameters for those finishing college three to five years after starting so that the fraction of four-year college graduates matches the calibration targets. And (4) adjust proportionally all the SCHGDF.hazardr_f and SCHGDF.hazardr_m parameters for college graduates so that the fraction of graduate degree holders matches the calibration targets.

### 7.3.3 Employment Pattern

The duration of non-employment between jobs is simulated in PENSIM using a non-employment hazard function, which is described in Appendix C, Section C.9). The constant term and female dummy coefficient in that hazard function, SJS_DUR.b_const and SJS_DUR.female, respectively, will be adjusted so the employment rate of men and women aged 25–54 matches the calibration targets. Also, the interaction terms between education and female are altered for only the 1935 birth cohort so that the female employment rates are lower than those for the 1955 cohort more for women with little schooling and less for women with more schooling.

In the cases of women born in 1935 and 1955, which are the only cases where the target employment rate index is more than five points different from 100, we adopt a modified method of calibration. Instead of lowering the employment rate by exclusively lengthening the time between jobs, we also shorten the time on jobs. We accomplish this reduction in subsequent job duration by adjusting the PENSIM input parameter SJF_DUR.female so that the median job tenure for women aged 45–54 matches approximately the 1983 CPS estimate of 6.3 years and the 2000 CPS estimate of 7.3 years for women aged 45–54 (BLS 2000a, Table 1) for the 1935 and 1955 birth cohorts, respectively. After matching the job duration calibration targets in these two cases, we adjust the length of time between jobs as described above so that the employment rate index calibration targets are matched.

### 7.3.4 Earnings Level and Inequality

The PENSIM input parameter that proportionally scales earnings for each individual in each year, AVGEARN.aer_sf, is adjusted so that average annual earnings at age fifty matches the calibration target. In addition, the earnings inequality parameters — ARVEARN.aer_m_loe, ARVEARN.aer_m_hie, ARVEARN.aer_f_loe, ARVEARN.aer_f_hie — are set based on the finding of Harris and Sabelhaus (2003, Figures 13–16).

### 7.3.5 Marital Status

Marital status at age sixty is simulated in PENSIM using an approach similar to that employed in the MINT project (Panis and Lillard 1999, pages 22–30). A log-linear hazard function generates the waiting time between age twelve and first marriage, or the waiting time between last divorce and remarriage. Another log-linear hazard function generates the waiting time between wedding and divorce. Our calibration strategy is to adjust the gender-specific coefficients of two covariates in the (re-)marriage hazard function: the constant term, whose coefficients are WEDDING.constant_f and WEDDING.constant_m, and the widowed dummy variable (WEDDING.widowed_f and WEDDING.widowed_m); and to adjust the gender-specific coefficients of three covariates in the divorce hazard function: the constant term (denoted by DIVORCE.constant_f and DIVORCE.constant_m), the 1–4 year portion of the piecewise linear marriage duration variable (DIVORCE.dur1_f and DI-

VORCE.dur1_m), and the 4–15 year portion of the same marriage duration variable (DIVORCE.dur4_f and DIVORCE.dur4_m). For each of the five covariates, the adjustment consists of adding an all-gender adjustment factor to the female and the male coefficients. In the case of the divorce duration coefficients we adjust the two of them in opposite directions so that the value of the duration variable is always the same after fifteen years, which is the end of the linear piece whose slope is defined by the dur_4 coefficient.

Marriages and divorces scheduled to occur after age sixty will not be simulated to occur. The adjustment of these coefficients should produce the marital status calibration targets.

Married couples are assumed to have no children. Because no children are simulated, these simulated life histories will not be suitable for analyzing social security benefits for young survivors or for dependents of disabled workers. None of these shortcomings will have an effect on the validity of social security benefits for those who age sixty and over, including survivors and disability benefits.

When sample individuals are simulated to marry, they are matched with a spouse who is not a member of the cohort sample. Spouses are assigned using a spouse age difference distribution and a spouse educational attainment distribution, both of which differ by the individual's gender. In addition, the spouse educational attainment distribution varies by the educational attainment of the sample individual, and the spouse age difference distribution varies by the age group of sample individual. For the purposes of spouse matching, educational attainment is measured at age thirty, rather than at the age when married. The spouse educational attainment and age difference distributions are tabulated from thirty years of Panel Study of Income Dynamics (PSID) data.

## 7.4 Calibration Results

This section of the chapter documents the PENSIM input parameter values required to produce the birth cohort samples that matched the calibration targets described above. After presenting those input values in a single tables, we present for each cohort the generated sample's values for each calibration statistic, comparing that value to the calibration target for that birth cohort. The calibrated value of earnings level is not shown in the tables because it is always matched exactly.

### 7.4.1 PENSIM Input Parameters

The values of the adjusted PENSIM input parameters for each birth cohort are shown in Table 5. All other PENSIM input parameters (except for the cohort-specific starting life tables and spouse age-difference and spouse education-difference matching probabilities) are the same across the PENSIM runs that represent the four birth cohorts. These common parameters are documented in Appendix C. The following paragraphs provide more details on the content of Table 5.

The parameter differences in the FUNCDIS table have been specified to calibrate the incidence of social security disability incidence to levels similar to those observed and predicted by the Social Security Administration's Office of the Chief Actuary.

In the SCHxxx tables that contain education attainment parameters, a multiplicative scaling factor (MF), which is applied to estimated parameters, is shown rather than the actual value of the parameters.

In the WEDDING and DIVORCE tables, an additive scaling factor (AF), which is applied to the PENSIM marriage and divorce parameters derived from Panis and Lillard (1999), is shown rather than the actual value of the hazard functions' parameters.

The parameter differences in the CBIRTH and COVCOEF tables have been calibrated to produce total fertility rates equal to those observed or predicted for each of the four birth cohorts.

Table 5: ***Calibrated PENSIM Input Parameter Values by Cohort***. *Parameter values in italics have* ***not*** *been adjusted in the calibration process. See text for details on the nature of the input parameter adjustments.*

| | Birth Cohort | | | |
|---|---|---|---|---|
| PENSIM Input Parameter | 1935 | 1955 | 1970 | 1985 |
| FUNCDIS.age_slope2 | 0.22 | 0.22 | 0.27 | 0.27 |
| DEATH.sf_hsa_f | 1.05 | 1.02 | 1.01 | 1.01 |
| DEATH.sf_hsa_m | 1.02 | 1.03 | 1.01 | 1.01 |
| DEATH.sf_hsg_f | 1.03 | 0.99 | 0.97 | 0.96 |
| DEATH.sf_hsg_m | 0.96 | 1.00 | 0.96 | 0.96 |
| DEATH.sf_coa_f | 1.00 | 0.97 | 0.94 | 0.93 |
| DEATH.sf_coa_m | 0.92 | 0.94 | 0.94 | 0.93 |
| DEATH.sf_cog_f | 0.96 | 0.91 | 0.88 | 0.88 |

*table continued on next page*

Table 5: ***Calibrated PENSIM Input Parameter Values by Cohort (continued)***

| | Birth Cohort | | | |
|---|---|---|---|---|
| PENSIM Input Parameter | 1935 | 1955 | 1970 | 1985 |
| DEATH.sf_cog_m | 0.84 | 0.86 | 0.88 | 0.88 |
| SCHHSF.diplomap_f MF | 0.81 | *1.00* | *1.00* | *1.00* |
| SCHHSF.diplomap_m MF | 0.76 | *1.00* | *1.00* | *1.00* |
| SCHCOS.hazardr_f MF | 0.35 | *1.00* | *1.00* | *1.00* |
| SCHCOS.hazardr_m MF | 0.73 | *1.00* | *1.00* | *1.00* |
| SCHCOF.diplomap_f MF | *1.00* | *1.00* | *1.00* | *1.00* |
| SCHCOF.diplomap_m MF | 1.22 | *1.00* | *1.00* | *1.00* |
| SCHGDF.hazardr_f MF | *1.00* | *1.00* | *1.00* | *1.00* |
| SCHGDF.hazardr_m MF | *1.00* | *1.00* | *1.00* | *1.00* |
| WEDDING.constant_f/_m AF | +0.470 | +0.205 | +0.078 | +0.076 |
| WEDDING.widowed_f/_m AF | –0.800 | +0.376 | +0.200 | –2.100 |
| DIVORCE.constant_f/_m AF | –0.280 | –0.190 | –0.340 | –0.340 |
| DIVORCE.dur1_f/_m AF | +0.176 | +0.154 | +0.143 | +0.154 |
| DIVORCE.dur4_f/_m AF | –0.048 | –0.042 | –0.039 | –0.042 |
| CBIRTH.twins_prob | 0.012 | 0.017 | 0.021 | 0.025 |
| COVCOEF.constant | 0.34 | –0.02 | –0.04 | –0.04 |
| SJS_DUR.b_const | 0.39 | –0.58 | –0.3714 | –0.41 |
| SJS_DUR.female | –2.95 | –0.95 | –1.1405 | –1.08 |
| SJS_DUR.hsa*female | –1.30 | *0.01016* | *0.01016* | *0.01016* |
| SJS_DUR.coa*female | 1.10 | *0.15307* | *0.15307* | *0.15307* |
| SJS_DUR.cog*female | 1.30 | *0.25095* | *0.25095* | *0.25095* |
| SJF_DUR.female | 0.849 | 0.548 | *0.06229* | *0.06229* |
| AVGEARN.aer_sf | 1.85 | 1.85 | 1.85 | 1.85 |
| AVGEARN.aer_m_loe | 1.00 | 0.95 | 0.85 | 0.85 |
| AVGEARN.aer_m_hie | 1.00 | 1.05 | 1.09 | 1.09 |
| AVGEARN.aer_f_loe | 1.00 | 1.20 | 1.25 | 1.25 |
| AVGEARN.aer_f_hie | 1.00 | 1.20 | 1.25 | 1.25 |

### 7.4.2 Results for 1935 Cohort

We used the PENSIM input parameters described at the beginning of this section to generate a sample of 100,000 life histories for the 1935 birth cohort. This sample has been tabulated to produce statistics regarding calibrated behavior. The sample value for each of these calibration statistics is compared with the calibration target in Table 6 on the facing page. There is a close match between the sample statistics and the target values, which indicates that the cohort sample closely represents the actual birth cohort along the calibrated dimensions.

#### 7.4.3 Results for 1955 Cohort

We used the PENSIM input parameters described at the beginning of this section to generate a sample of 100,000 life histories for the 1955 birth cohort. This sample has been tabulated to produce statistics regarding calibrated behavior. The sample value for each of these calibration statistics is compared with the calibration target in Table 7 on page 50. There is a close match between the sample statistics and the target values, which indicates that the cohort sample closely represents the actual birth cohort along the calibrated dimensions.

#### 7.4.4 Results for 1970 Cohort

We used the PENSIM input parameters described at the beginning of this section to generate a sample of 100,000 life histories for the 1970 birth cohort. This sample has been tabulated to produce statistics regarding calibrated behavior. The sample value for each of these calibration statistics is compared with the calibration target in Table 8 on page 51. There is a close match between the sample statistics and the target values, which indicates that the cohort sample closely represents the actual birth cohort along the calibrated dimensions.

#### 7.4.5 Results for 1985 Cohort

We used the PENSIM input parameters described at the beginning of this section to generate a sample of 100,000 life histories for the 1985 birth cohort. This sample has been tabulated to produce statistics regarding calibrated behavior. The sample value for each of these calibration statistics is compared with the calibration target in Table 9 on page 52. There is a close match between the sample statistics and the target values, which indicates that the cohort sample closely represents the actual birth cohort along the calibrated dimensions.

Table 6: ***Results from PENSIM Sample for 1935 Birth Cohort***. *PENSIM sample generated using input parameters as described above. Calibration target values as described above.*

| Cohort Statistic | Sample Value | Target Value |
|---|---|---|
| *Life Expectancy in years:* | | |
| Female cohort life expectancy at birth | 73.8 | 74.4 |
| Female cohort life expectancy at age 65 | 19.7 | 19.6 |
| Male cohort life expectancy at birth | 67.7 | 67.7 |
| Male cohort life expectancy at age 65 | 16.6 | 16.2 |
| *Educational Attainment Distribution in percent:* | | |
| Female high school dropout | 23.6 | 24 |
| Female high school graduate | 44.9 | 45 |
| Female college attendee | 17.0 | 17 |
| Female (4-year) college graduate | 9.0 | 9 |
| Female graduate degree holder | 5.6 | 5 |
| Male high school dropout | 28.2 | 28 |
| Male high school graduate | 33.7 | 34 |
| Male college attendee | 16.0 | 16 |
| Male (four-year) college graduate | 13.6 | 13 |
| Male graduate degree holder | 8.5 | 9 |
| *Employment Rate (ages 25–54) in percent:* | | |
| Female | 35.7 | 35.8 |
| Male | 90.6 | 91.0 |
| *Marital Status at Age 60 Distribution in percent:* | | |
| All-gender never married | 4.6 | 4.6 |
| All-gender currently married | 73.9 | 75.1 |
| All-gender divorced (ever married 10+ years) | 8.0 | 7.6 |
| All-gender divorced (never married 10 years) | 4.7 | 4.0 |
| All-gender widowed (from prior marriage) | 8.8 | 8.7 |

Table 7: ***Results from PENSIM Sample for 1955 Birth Cohort***. *PENSIM sample generated using input parameters as described above. Calibration target values as described above.*

| Cohort Statistic | Sample Value | Target Value |
|---|---|---|
| *Life Expectancy in years:* | | |
| Female cohort life expectancy at birth | 78.8 | 79.1 |
| Female cohort life expectancy at age 65 | 20.8 | 20.6 |
| Male cohort life expectancy at birth | 72.6 | 72.9 |
| Male cohort life expectancy at age 65 | 17.8 | 17.5 |
| *Educational Attainment Distribution in percent:* | | |
| Female high school dropout | 11.0 | 11 |
| Female high school graduate | 31.7 | 32 |
| Female college attendee | 31.0 | 31 |
| Female (4-year) college graduate | 16.2 | 16 |
| Female graduate degree holder | 10.1 | 10 |
| Male high school dropout | 12.3 | 12 |
| Male high school graduate | 33.5 | 34 |
| Male college attendee | 27.0 | 27 |
| Male (four-year) college graduate | 16.5 | 16 |
| Male graduate degree holder | 10.6 | 11 |
| *Employment Rate (ages 25–54) in percent:* | | |
| Female | 64.9 | 64.5 |
| Male | 84.9 | 85.0 |
| *Marital Status at Age 60 Distribution in percent:* | | |
| All-gender never married | 9.6 | 9.5 |
| All-gender currently married | 68.8 | 69.9 |
| All-gender divorced (ever married 10+ years) | 9.0 | 8.8 |
| All-gender divorced (never married 10 years) | 6.9 | 6.3 |
| All-gender widowed (from prior marriage) | 5.7 | 5.5 |

Table 8: ***Results from PENSIM Sample for 1970 Birth Cohort***. *PENSIM sample generated using input parameters as described above. Calibration target values as described above.*

| Cohort Statistic | Sample Value | Target Value |
|---|---|---|
| *Life Expectancy in years:* | | |
| Female cohort life expectancy at birth | 80.6 | 81.0 |
| Female cohort life expectancy at age 65 | 21.5 | 21.4 |
| Male cohort life expectancy at birth | 75.1 | 75.5 |
| Male cohort life expectancy at age 65 | 18.5 | 18.3 |
| *Educational Attainment Distribution in percent:* | | |
| Female high school dropout | 11.0 | 11 |
| Female high school graduate | 31.7 | 32 |
| Female college attendee | 30.9 | 31 |
| Female (4-year) college graduate | 16.3 | 16 |
| Female graduate degree holder | 10.1 | 10 |
| Male high school dropout | 12.5 | 12 |
| Male high school graduate | 33.4 | 34 |
| Male college attendee | 27.1 | 27 |
| Male (four-year) college graduate | 16.4 | 16 |
| Male graduate degree holder | 10.7 | 11 |
| *Employment Rate (ages 25–54) in percent:* | | |
| Female | 71.1 | 71.7 |
| Male | 86.1 | 86.7 |
| *Marital Status at Age 60 Distribution in percent:* | | |
| All-gender never married | 12.6 | 12.5 |
| All-gender currently married | 68.2 | 69.5 |
| All-gender divorced (ever married 10+ years) | 7.8 | 7.5 |
| All-gender divorced (never married 10 years) | 6.2 | 5.6 |
| All-gender widowed (from prior marriage) | 5.2 | 4.9 |

Table 9: ***Results from PENSIM Sample for 1985 Birth Cohort****. PENSIM sample generated using input parameters as described above. Calibration target values as described above.*

| Cohort Statistic | Sample Value | Target Value |
|---|---|---|
| *Life Expectancy in years:* | | |
| Female cohort life expectancy at birth | 82.4 | 82.9 |
| Female cohort life expectancy at age 65 | 22.3 | 22.2 |
| Male cohort life expectancy at birth | 77.3 | 77.9 |
| Male cohort life expectancy at age 65 | 19.3 | 19.1 |
| *Educational Attainment Distribution in percent:* | | |
| Female high school dropout | 11.0 | 11 |
| Female high school graduate | 31.7 | 32 |
| Female college attendee | 30.9 | 31 |
| Female (4-year) college graduate | 16.3 | 16 |
| Female graduate degree holder | 10.1 | 10 |
| Male high school dropout | 12.4 | 12 |
| Male high school graduate | 33.4 | 34 |
| Male college attendee | 27.2 | 27 |
| Male (four-year) college graduate | 16.4 | 16 |
| Male graduate degree holder | 10.6 | 11 |
| *Employment Rate (ages 25–54) in percent:* | | |
| Female | 71.4 | 71.7 |
| Male | 86.4 | 86.7 |
| *Marital Status at Age 60 Distribution in percent:* | | |
| All-gender never married | 12.4 | 12.7 |
| All-gender currently married | 68.6 | 69.6 |
| All-gender divorced (ever married 10+ years) | 7.8 | 7.5 |
| All-gender divorced (never married 10 years) | 6.1 | 5.6 |
| All-gender widowed (from prior marriage) | 5.1 | 4.6 |

# 8 Validation of PENSIM Lives

This chapter of the overview presents results from several validation tests. In each test, statistics calculated with simulated data from a PENSIM run are compared with conceptually equivalent statistics calculated with actual data. In one case (disability), actual data are not readily available, and therefore, statistics from another simulation model are used in the comparison. In this case, the test is best described as a cross-model comparison test rather than an historical validation test.

The validation tests presented in this chapter supplement the cross-model comparisons of aggregate social security projections described on page 12. In those SSASIM OLC-mode comparisons, PENSIM life histories for cohorts born in 1935 and following years have been found to produce aggregate population, taxable earnings, and disability benefits projections over the 2005–2080 period that closely match projections from the Congressional Budget Office and the SSA Office of the Chief Actuary. OLC-mode projections of retirement benefits are in line with CBO's projection, but lower than the Actuary's (Holmer 2006). The close agreement between the CBO projection and the OLC-mode projection would be extremely unlikely if the underlying simulated life histories differ in significant ways. This is because aggregate social security projections are determined not only by population and earnings totals, but also by life expectancy and the distribution of earnings. And furthermore, OLC-mode estimates of the aggregate social security solvency effects of a wide variety of simple reform provisions have been shown to be quite close to the estimated effects produced by the SSA Office of the Chief Actuary. This means that PENSIM-generate life histories have successfully undergone an extensive range of cross-model comparison tests as part of the comparison of SSASIM OLC-mode estimates with estimates produced by other social security models.

The validation tests presented in this chapter focus on disability histories, earnings histories, and job histories. Before presenting the results of the tests, we discuss some methodological issues involved in the validation process.

## 8.1 Validation Methods

In this part of the chapter, we describe the PENSIM runs that generate the simulated data, and then address some methodological issues that will arise when we conduct the validation tests with those data.

### 8.1.1 PENSIM Validation Runs

All the simulated results presented in this chapter are derived from the output of PENSIM runs that simulate one macroeconomic and macrodemographic scenario that is similar to the intermediate-cost projection in one of the recent annual reports of the trustees of the social security program. The validation runs, therefore, have no collective (or systemic) risk factors that vary across scenarios.

In some cases, PENSIM is used to generate a longitudinal sample of a single birth cohort. In other cases, PENSIM is used to generate a cross-sectional sample for a given calendar year by combining information gathered at different ages for each of the cohorts alive in that year.

The longitudinal PENSIM validation statistics presented in this chapter are calculated from output generated for one of the birth cohorts described in the calibration of lives discussion contained in Chapter 7. The longitudinal samples are two percent samples that contain nearly 100,000 cohort individuals.

The cross-sectional PENSIM validation statistics presented in this chapter are calculated from output generated for all the birth cohorts whose members are alive in the cross-sectional sample's calendar year. Each one of these overlapping cohort samples is a one-tenth percent (one-in-a-thousand) sample that contain roughly 4,000 individual per cohort, with the actual size varying with the historical size of the birth cohort. This means, for example, that if we generate a sample of those age 25–64 in 2004, the sample will contain individuals from forty different birth cohorts (1940–1979) and will total about 160,000 individuals.

### 8.1.2 Tabulating Actual Longitudinal Statistics

In some cases, actual longitudinal survey data are not used to compute earnings statistics at each age for a single birth cohort. Instead, age-specific statistics are computed from eight March CPS surveys with information on annual earnings for the years 1991–1998. We use these CPS samples to track six different birth cohorts over an eight-year period. The cohort born in 1969 was 22–29 years of age in the 1991–1998 period, those born in 1962 were 29–36 in that same period, those born in 1955 were 36–43, the cohort born in 1948 was 43–50 years old, those born in 1941 were 50–57, and the cohort born in 1934 was 57–64 years old during the 1991–1998 period.

This method of using repeated cross-sectional surveys to track the members of a birth cohort is a standard practice in the research literature (Deaton and Paxson 1994, for example). This method does not produce genuine longitudinal data, but it is useful in producing aggregate statistics for a birth cohort as the cohort individuals age.

### 8.1.3 Comparing Simulated and Actual Statistics

When comparing the value of a statistic that is computed using simulated data with the value of the same statistic computed using actual data, how can we tell whether they are close enough to say the model "passed" the validation test? As in the grading of any test, different standards can be applied.

The strictest grading standard would be to require the simulated and actual statistics to differ by no more than a statistically insignificant amount. In a number of situations in this chapter, we use a chi-square goodness-of-fit test to impose this high grading standard. If the simulated and actual statistics are found to be insignificantly different in this statistical sense, few would deny that the validation test has been passed.

But it is common for experienced researchers to adopt somewhat less strict grading standards, primarily because it is felt that the simulated statistics are close enough to the actual statistics (even though there is a statistically significant difference) to give the model a passing grade on the validation test. This puts us in a subjective realm where knowledge about whether the observed difference is likely to have a meaningful effect on the simulation model's results comes into play. Part of this judgment turns on the fact that some validation statistics are more important than others in gauging the realism of a model's simulated data.

The practical problems of comparing simulated and actual statistics and of choosing sensible grading standards for a validation test are illustrated in the following example taken from the MINT project (Panis and Lillard 1999, page 27):

> *Goodness of Fit of Marriage Transition Models*
>
> Our hazard models of getting married and divorced ... are based on experiences of the 1990 and 1991 SIPP respondents. We applied these estimates to 1992 and 1993 SIPP respondents to asses the goodness of fit. Starting all respondents at age 12 (when

no one is married yet), we project marital transitions until the last [SIPP] interview date. Table 2.6 shows actual marital status and projected marital status for these 1992 and 1993 respondents.

As is clear from the table, projected and actual marital status are very close. The discrepancies may be due to stochasticity (because of duration draws in the projection method) or to a mild self-selection. The projection ... assumes that all respondents survive through the last [SIPP] survey. In reality, SIPP respondents are the survivors of their birth cohorts, and thus somewhat self selected.

Table 2.6: Actual and Projected Marital Status [abbreviated]

| | Actual status (%) | Projected status (%) |
|---|---|---|
| Never married | 11.3 | 10.4 |
| Married | 73.7 | 71.4 |
| Widowed | 3.6 | 4.9 |
| Divorced | 11.4 | 13.3 |

We agree completely with the authors' judgment that the MINT simulated marital histories pass this validation test. We present this example to illustrate two points. First, it is typical for a validation test to be complicated by some kind of non-comparability in the simulated and actual data that are being compared. And second, the strictest grading standards can misleadingly suggest a model has failed a validation test. In this case, if we actually conduct a chi-square goodness-of-fit test using the table percents (which are based on a total sample of 38,094 individuals), we find that $\chi^2(3)$=345.8 (p=0.0000), which strongly indicates that there is a statistically significant difference between the simulated (i.e., projected) and actual distributions of marital status.

Bearing these methodological issues in mind, we begin the validation tests. In the next section, we present results from a cross-model comparison test on disability incidence. In subsequent sections of the chapter, we present results from validation tests on earnings histories and on job histories.

## 8.2 Validating Disability Histories

Figure 3 shows by age and gender the prevalence of disabled-worker benefit receipt for a longitudinal sample born of individuals born in 1955. The

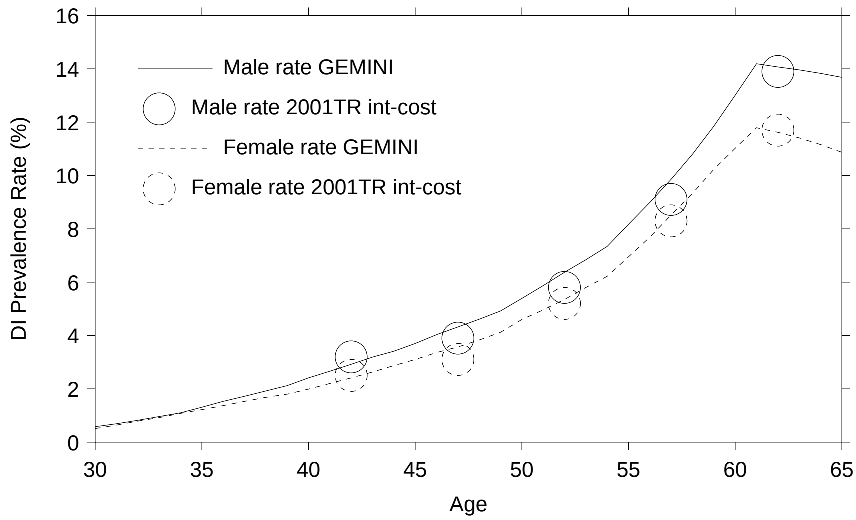


Figure 3: ***Disabled-Worker Beneficiary Prevalence Rates by Age and Gender.*** *Simulated rates for individuals born in 1955 (shown as lines) are from a GEMINI (10/3/03 version) run that uses intermediate-cost assumptions from the 2001 Trustees Report. Comparison rates (shown as circles) are from projections based on SSA Office of the Chief Actuary's intermediate-cost projection assumptions, as presented in the MINT3 Report, Table 2-4 (Toder, et al., "Final Report: Modeling Income in the Near Term: Revised Projections for Retirement Income Through 2020 for the 1931–1960 Birth Cohorts," Washington, DC: Urban Institute, June 2002).*

simulated results are based on disability status simulated by PENSIM and detailed disability-insured calculations in GEMINI. PENSIM does impose a crude disability-insured test, but the GEMINI test is more accurate. Of the 12,285 people simulated to become disabled in PENSIM, only 92 (that is, 0.75 percent) fail the GEMINI disability-insured test. This means the disability-insurance prevalence rates from the two models are almost the same.

Comparing the percent of living cohort individuals simulated to have received (currently or in the past) social security disabled-worker benefits at each age with projections made by the SSA Office of the Chief Actuary show a close correspondence. The simulated age profile of disability exhibits an expected pattern: incidence rising sharply when individuals reach their late forties and their fifties.

## 8.3 Validating Earnings Histories

Now we turn to validating economic statistics. We begin in this section of the chapter by testing the validity of a range of statistics computed from earnings histories. Earnings histories contain no information about how many jobs are held over a work career, and therefore, by themselves are not very useful for simulating pension-related behavior. In the section after this, we examine simulated job histories and focus on their pension-related characteristics.

We examine several aspects of earnings histories using longitudinal samples: a measure of labor force participation, the changing degree of earnings inequality as a birth cohort ages, the aggregate age-earnings profile, the variability in individual earnings histories around that aggregate age-earnings profile, and the correlation between earnings earlier and latter in life. And finally, we examine the distribution of earnings in a cross-sectional sample.

### 8.3.1 Zero Cohort Earnings Incidence by Age

One measure of labor force participation is whether the individual has positive (rather than zero) annual earnings. The percent of individuals in a longitudinal sample who are simulated to have zero annual earnings is shown for each age from 22 through 64 in Figure 4. The PENSIM and CPS estimates of the age-specific incidence of zero annual earnings are reasonably close.

### 8.3.2 Cohort Earnings Inequality by Age

Moving beyond labor force participation, we now examine the distribution of earnings across the members of a birth cohort, and how the degree of inequality in the earnings distribution changes as the cohort ages. There is substantial evidence from widely differing countries that the degree of earnings inequality rises as a birth cohort ages (Deaton and Paxson 1994). It is important for any simulation model attempting to simulate earnings histories to produce this growing inequality phenomenon. Here we examine the age trend in the cross-sectional distribution of earnings, considering at each age only those with positive annual earnings that year.

Using the annual earnings histories for individuals from a longitudinal sample of a single birth cohort, we can measure age-specific earnings inequality by computing a Gini coefficient for the cohort sample at each age.

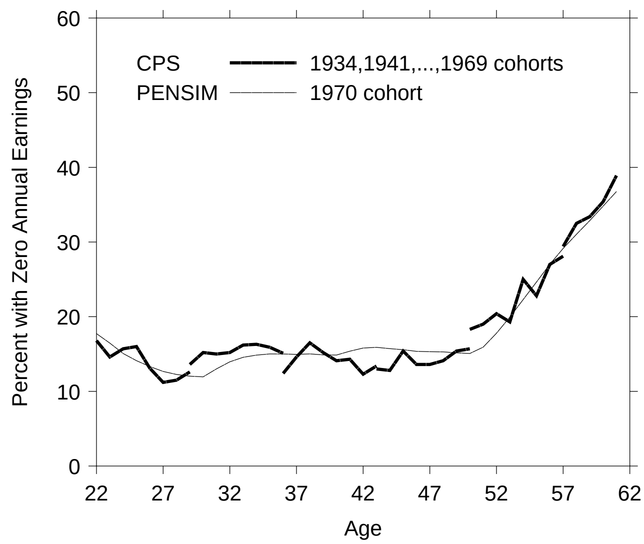


Figure 4: ***Percent of Individuals with Zero Annual Earnings by Age.*** *Sources: PENSIM values for individuals born in 1970 are from a PENSIM (1/20/06 version) run that uses 2005 Trustees Report intermediate-cost assumptions; CPS values are from authors' tabulation of CPS March supplement data on annual earnings for 1991–1998 as described in the text.*

The age trend of this measure of simulated earnings inequality is shown in by the thin lines in Figure 5 for two different birth cohorts.

The simulated degree of earnings inequality is compared with that observed in CPS data for the years between 1991 and 1998. These CPS surveys are used to track six different birth cohorts over the eight years from 1991 to 1998. The actual degree of earnings inequality in these CPS data is shown in Figure 5 by the thick lines.

The agreement between the simulated and actual degree of earnings inequality is approximate. Both the PENSIM and CPS Gini coefficients show a tendency for cohort earnings inequality to rise as members of the cohort age, as has been documented with other data in both the U.S. and other countries (Deaton and Paxson 1994).

The secular rise in earnings inequality during the 1991–1998 period is apparent in Figure 5. PENSIM does not simulate a secular shift in earning inequality for any one cohort, but does assume that younger cohorts will experience more earnings inequality than older cohorts.

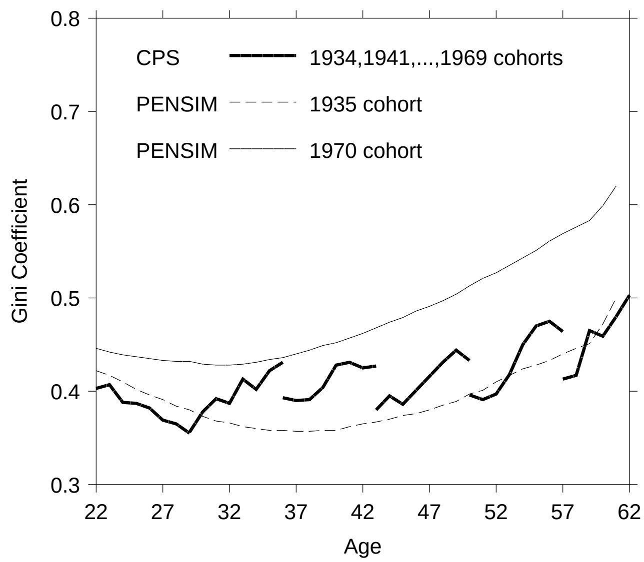


Figure 5: ***Inequality of Annual Positive Earnings by Age.*** *Gini coefficients calculated using only positive earnings. Sources: PENSIM values for individuals born in 1935 and 1970 are from a PENSIM (1/20/06 version) run that uses 2005 Trustees Report intermediate-cost assumptions; CPS values are from authors' tabulation of CPS March supplement data on annual earnings for 1991–1998 as described in the text.*

### 8.3.3 Average Relative Cohort Earnings by Age

In this section we construct statistics that allow us to present an aggregate age-earnings profile. Such a profile shows how earnings averaged over all those with positive earnings varies by age. But the aggregate age-earnings profile says nothing about how much individual age-earnings profiles vary around this average. We look at the variety of individual age-earnings profiles in the section after this.

We use the simulated annual earnings histories for individuals in a longitudinal sample of a single cohort to calculate at each age the ratio of average earnings at that age and average earnings at all ages. Normally such an aggregate age-earnings profile is calculated using cross-sectional data. In order for the all-age average earnings level computed from the simulated earnings histories to be comparable to a cross-sectional all-age average, we divide average earnings at each age by the SSA average wage index for that year. This

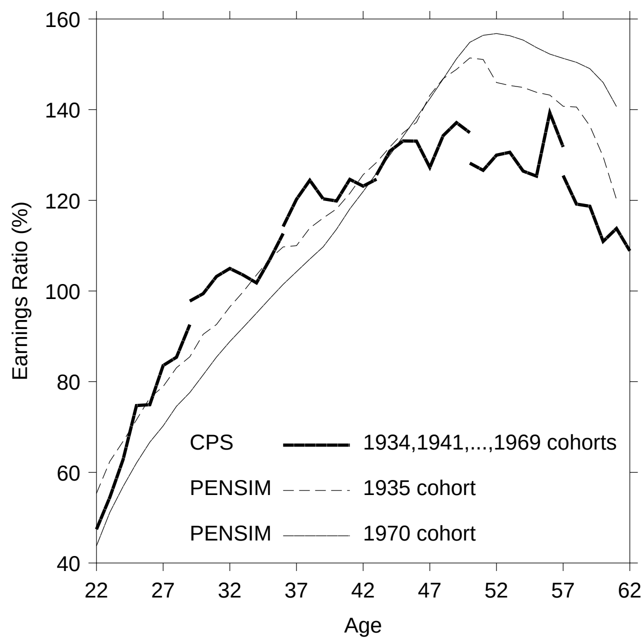


Figure 6: ***Ratio of Age-Specific Average Relative Earnings to All-Age Mean of Average Relative Earnings.*** *Relative earnings are average cohort earnings divided by the economy-wide SSA average wage index. The thin lines show for two cohorts the ratio calculated assuming the all-age mean is 93% of average earnings for those 20–64 years old, which was the case in the CPS March supplement data for all six years from 1993 to 1998. The thick line shows the relative earnings level tabulated from CPS data. Sources: PENSIM values for individuals born in 1935 and 1970 are from a PENSIM (1/20/06 version) run that uses 2005 Trustees Report intermediate-cost assumptions; CPS values from authors' tabulation of CPS March supplement data on annual earnings for 1991–1998 as described in the text.*

produces average relative earnings at each age. The simulated age-earnings profile is constructed by calculating at each age the ratio of average relative earnings and an all-age mean of average relative earnings for the cohort.

We compare the age-earnings profile calculated from PENSIM earnings histories with eight-year profiles for six birth cohorts calculated from CPS data for the years 1991 through 1998 in Figure 6. The simulated and actual aggregate age-earnings profiles are roughly similar, with both exhibiting a sharp rise in relative average earnings for individuals in their twenties and thirties. This rise levels off for individuals in their late forties, and turns into a decline for those in their fifties and early sixties.

#### 8.3.4 Variation in Cohort Earnings Histories

Individual age-earnings profiles generated by PENSIM vary widely around the aggregate age-earnings profile shown in Figure 6. Some individuals start life earning above average and end up below average later in life. Others have just the opposite kind of relative earnings history. To show the variety of earnings histories generated by PENSIM, we classify the simulated earnings histories into nine level-trend categories using a classification scheme similar to that used in the MINT project (Toder et al. 1999, Chapter 8). These nine categories are defined by the *level* of the lifetime average of relative earnings (low, average, and high) and by the *trend* in relative earnings over a lifetime (declining, steady, and rising).

The classification scheme used here is identical to that used in the MINT project except for one difference. The scheme used here divides an individual's annual earnings by the cohort-wide average earnings for that year (rather than economy-wide average earnings). We do this to avoid confounding the lifetime earnings patterns of a single birth cohort with historical changes in the relative standing of different birth cohorts. Our classification scheme includes those with zero earnings. In fact, a major cause of irregular earnings histories is disability (and other causes of labor force withdrawal).

Other than this different concept of relative earnings, the classification scheme employed here is identical to that used in the MINT project when both positive and zero earnings are considered. The lifetime level of an individual's relative earnings is calculated as the mean of relative earnings over the thirty years from age 32 through age 61, and is denoted by $L$. Ten-year averages are also calculated for each individual with the first-decade average denoted by $A$ and the third-decade average denoted by $C$. Individuals are then classified into tertiles defined by the level of $L$. Also, individuals are assigned a relative earnings trend (either declining, steady, or rising) depending on their value of $(C-A)/(C+A)$. If that value is less than $-\frac{1}{9}$ the trend is said to be declining; if it is greater than $+\frac{1}{9}$ the trend is said to be rising; otherwise the trend is said to be steady.

The fraction of earnings histories simulated by the PENSIM validation run that fall into the nine categories defined by the interaction of the level tertiles and the trend assignments is shown in Figure 7. The figure also shows average relative earnings at each age for the individuals in each of these nine categories. While it is not possible to compare these results with those based on actual earnings histories for the 1931–35 birth cohorts (Toder et al. 1999, page 278) because of differences in the definition of relative earnings, differences in the degree of earnings truncation, differences in the accuracy of the tertile classification, and differences in the historical experiences of the actual and simulated cohorts, it is clear that the PENSIM earnings histories show the same kind of wide variety in their level-trend character as seen in actual earnings histories.

This categorization of individual relative earnings profiles reveals that large groups of people are experiencing substantial changes in their relative earnings over their lifetimes. Consider, for example, the 18.1 percent of the birth cohort who make up the average-declining group. On average this group goes from earning about 20 percent above average in their late thirties to earning about 60 percent below average in their late fifties. Or, consider the 11.5 percent of the birth cohort who make up the high-rising group. On average that group goes from earning no more than about 150 percent of the average during their twenties to earning over three times the average in their fifties. It appears as if only part of this sharp rise in relative earnings is caused by declining labor force participation during the fifties.

### 8.3.5 Covariance of Cohort Earnings Histories

While the above qualitative classification of earnings histories shows that relative earnings vary considerably over a lifetime for most individuals, it is important to measure quantitatively the degree of fluctuations in earnings. There are many ways to measure the degree of change in earnings over a lifetime. In order to facilitate the validation effort, we focus on calculating statistics that are comparable to statistics available in the research literature.

The first statistic we compute is the variance in the year-to-year change

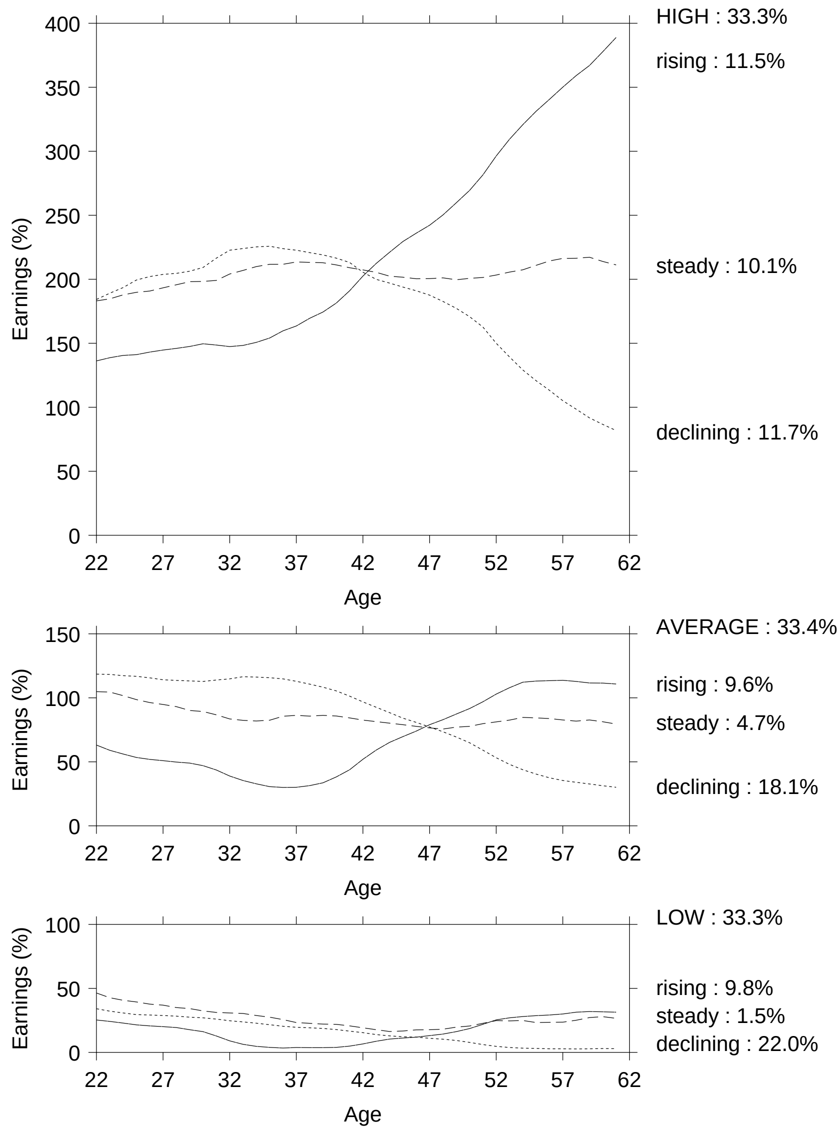


Figure 7: ***Level-Trend Categorization of Relative Earnings Over Lifetime in Simulated Life Histories from PENSIM.*** *Annual earnings, which can be zero, are expressed as percent of cohort-wide average earnings for that age. Source: earnings for individuals born in 1935 are from a PENSIM (1/20/06 version) run that uses 2005 Trustees Report intermediate-cost assumptions.*

Table 10: ***Variance of One-Year Change in Log Annual Earnings and Correlation of Full-Time Earnings at Different Ages.*** *PENSIM variance calculated for both men and women who had earnings in every year of a randomly selected eleven-year window, with the one-year change selected at random from this window. PENSIM correlation coefficients calculated using log annual earnings (not log hourly earnings) for both men and women who were employed full-time at both the ages used in the calculation.*

| Longitudinal Earnings Statistic | PENSIM value | Other value [stdev] |
|---|---|---|
| Variance of one-year change in log annual earnings | 0.182 | 0.172 [0.028] all ten<br>0.194 [0.018] last five |
| Correlation coefficient between full-time earnings at ages 37 and 42 | 0.89 | 0.77 for 1953 cohort<br>0.82 for 1943 cohort |
| Correlation coefficient between full-time earnings at ages 37 and 52 | 0.74 | 0.68 for 1943 cohort |

*Sources: PENSIM values for individuals born in 1935 are from a PENSIM (1/20/06 version) run that uses 2005 Trustees Report intermediate-cost assumptions; other values: variance of one-year change in log earnings is from Abowd and Card (1989, Table IV) using PSID data on prime-age men from 1969–1979 who had positive earnings and headed a household in each of those years, and correlations of log hourly earnings at different ages are from Dickens (2000, Figure 1) using British longitudinal data for 1975–1995 on prime-age men who work full-time.*

in the natural logarithm of annual earnings. This statistic has been reported for U.S. men during the 1970s (Abowd and Card 1989).

The second kind of statistic we compute is a correlation coefficient between full-time earnings at two different ages. We compute a correlation coefficient between earnings at ages 37 and 42, and another correlation coefficient between earnings at ages 37 and 52. These two measures of correlation quantify the degree of change in earnings over a five and a fifteen year period during midlife. These two statistics have been reported for U.K. men during the 1975–1995 period (Dickens 2000).

Pairs of values for each of these three statistics, one computed with PENSIM earnings histories and the other from the research literature, are shown in Table 10. The U.S. pair of statistics differ by a statistically insignificant amount, but the PENSIM correlation coefficients are somewhat above those for U.K. workers. It is unclear whether these differences are statistically significant, and if so, whether they accurately reflect differences in the two countries. The results of these tests suggest that the longitudinal properties of the earnings histories generated by PENSIM are not unrealistic.

### 8.3.6 Cross-Sectional Distribution of Earnings

Another way to check the earnings histories generated by PENSIM is to use all the cohorts simulated by PENSIM for a SSASIM OLC-mode run (see page 12) to create a cross-sectional sample. The simulated cross-sectional sample for 2004 contains individuals age 25–64 from forty cohorts (born 1940–1979) who are simulated from birth. This sample is used to compute several earnings distribution statistics that can be compared to equivalent statistics tabulated from March 2005 CPS data. Table 11 shows the differences between the simulated and actual distribution of earnings, as well as the distribution of education and gender among earners, and mean earnings for each education and gender group.

The total number of earners and the mean earnings for all earners are quite close in the simulated and actual cross-section samples.

The comparison of the distribution of earnings shows that PENSIM simulates relatively more low earners, relatively fewer middle earners, and the same number of high earners as observed in CPS data. While the size of these groups is somewhat different the group means are about the except for the highest group. It is unclear how much of the group size differences are caused by modest differences in the continuous distribution being amplified

Table 11: ***2004 Cross-Sectional Earnings Statistics.*** *Statistics include only those individuals who are in the 25–64 age range and have earnings. Number of earners is expressed in millions and mean earnings in thousands of dollars.*

| Category of Earner | PENSIM number | PENSIM mean | CPS number | CPS mean |
|---|---|---|---|---|
| All earners | 125 | 41 | 124 | 42 |
| 0–20 earnings range | 36% | 11 | 27% | 10 |
| 20–30 earnings range | 16% | 25 | 18% | 24 |
| 30–40 earnings range | 12% | 35 | 16% | 34 |
| 40–60 earnings range | 16% | 49 | 20% | 47 |
| 60–100 earnings range | 13% | 76 | 13% | 73 |
| 100+ earnings range | 7% | 162 | 6% | 176 |
| All earners | 125 | 41 | 124 | 42 |
| high school dropout | 12% | 24 | 10% | 22 |
| high school graduate | 32% | 34 | 30% | 31 |
| some college | 29% | 42 | 28% | 37 |
| four-year college degree | 17% | 55 | 21% | 55 |
| advanced graduate degree | 10% | 53 | 11% | 79 |
| All earners | 125 | 41 | 124 | 42 |
| female | 45% | 37 | 47% | 32 |
| male | 55% | 43 | 53% | 51 |

*Sources: PENSIM values for individuals age 25–64 in 2004 with positive earnings are from PENSIM (1/20/06) run that generates one-in-a-thousand samples for each birth cohort whose individuals are in that age range during 2004; CPS values are tabulated from raw data from the 2005 CPS March supplement.*

by the categorization.

The PENSIM distribution of earners across educational groups is roughly the same as seen in CPS, and mean earnings in each educational group is roughly the same except for those with advanced degrees.

The PENSIM gender distribution of earners is slightly different from the CPS data, with the fraction of women being somewhat lower. Also, the male-female earnings gap is smaller in the simulation than in the CPS data.

From a high level, the simulated cross-sectional earnings distribution looks approximately like the actual CPS distribution. Looking more closely, there are differences. It is unclear whether or not these differences are large enough to have a significant effect on the simulation results produced by PENSIM. Whether or not this is the case depends, in part, on what PENSIM results are being considered. And, of course, without other models being subjected to the same validation test, we have no idea whether the simulations of other models are more valid.

## 8.4 Validating Job Histories

The calibration and validation results presented so far suggest that PENSIM produces reasonably realistic demographic histories and annual earnings histories. The aspects of life histories discussed so far include all the information that is required to estimate social security benefits using the SSASIM OLC mode of operation (see page 12).

But this information is not sufficient to estimate employer-sponsored pension benefits. What is needed for pension analysis is a job history, not just an earnings history. The remainder of this chapter presents validation results on two key aspects of job histories: the duration of each job and whether the employer on each job sponsors a pension. These two aspects of job histories, in addition to the detailed characteristics of the pension(s) sponsored by each employer in the job history, are the kind of information necessary for the pension accumulation and withdrawal calculations performed by PENSIM.

### 8.4.1 Cross-Sectional Distribution of Job Tenure

We begin our examination of job histories by calculating job tenure statistics.

PENSIM uses hazard functions to generate the completed duration of a job, given the characteristics of the job and the attributes of the individual holding that job (for details see Appendix B, Section B.17.6 and Section B.19.6). It would be very difficult to compare the simulated distribution of completed job duration with actual survey data because an extremely long longitudinal sample would be required. Here we use a simulated cross-sectional sample to collect job tenure (i.e., time since starting the current job) and compare the calculated mean job tenure for different age-gender groups with equivalent statistics calculated from CPS data. This is a tricky exercise because job tenure in CPS data can change over the course of the business cycle (as more new workers are hired during an expansion and shorter tenured workers are laid off during a contraction) and over time because of structural changes in job-holding behavior. We compare simulated mean job tenure from a cross-sectional sample for 2000 with mean job tenure from February 2000 CPS data.

Simulated and actual mean job tenure are shown in Table 12 for two age groups: employees age 25–64 and employees age 55-64.

The simulated and actual number of employees in these two age groups are quite similar. The simulated mean job tenure is somewhat higher than the actual mean job tenure: 8.8 years versus 7.9 years for the broader age group and 14.2 versus 13.3 for the older age group.

Comparing the simulated and actual distributions of job tenure, we see that PENSIM simulates fewer employees with short job tenure, more with medium job tenure, and about the same number of employees with long (20+ years) job tenure. It is unclear whether the larger number of employees with less than five years job tenure is attributable to the rapid employment growth during the 1990s boom. Also, PENSIM generates a larger gender gap in mean job tenure than observed in the CPS data.

There are two aspects of the job-tenure distribution are more important to pension simulation than other aspects: the number with short job tenure and the number with long job tenure. Short (completed) job tenure often means little or no pension vesting. Long job tenure means the employee would benefit substantially from a terminal-earnings (that is, final-pay) defined-benefit pension, if the employer offers such a plan.

Both the simulated and actual data show a substantial number of short (less than five years) job tenures. These are incomplete job durations, so we have no idea from the statistics in Table 12 how many in this group will end up with short completed job durations, and therefore, run the risk of less than full vesting. The longitudinal samples simulated in PENSIM take all this into account in simulating pension accumulation.

One conclusion that can be drawn from this validation test is that there

Table 12: ***2000 Cross-Sectional Job Tenure Statistics.*** *Statistics include only those individuals who are in the 25–64 age range and are employed. Number of employed is expressed in millions and mean (incomplete) job tenure in years.*

| Category of Employee | PENSIM number | PENSIM mean | CPS number | CPS mean |
|---|---|---|---|---|
| All employees age 25–64 | 111.0 | 8.8 | 110.6 | 7.9 |
| 0–5 years job tenure | 39% | 2.3 | 48% | 1.8 |
| 5–10 years job tenure | 26% | 7.3 | 20% | 6.5 |
| 10–20 years job tenure | 25% | 14.1 | 21% | 13.2 |
| 20–30 years job tenure | 8% | 24.0 | 9% | 23.1 |
| 30+ years job tenure | 2% | 34.3 | 2% | 33.0 |
| All employees age 25–64 | 111.0 | 8.8 | 110.6 | 7.9 |
| female | 43% | 7.5 | 47% | 7.1 |
| male | 57% | 9.8 | 53% | 8.5 |
| All employees age 55–64 | 13.4 | 14.2 | 13.7 | 13.3 |
| 0–5 years job tenure | 19% | 2.4 | 28% | 1.9 |
| 5–10 years job tenure | 22% | 7.5 | 16% | 6.6 |
| 10–20 years job tenure | 34% | 14.5 | 26% | 13.4 |
| 20–30 years job tenure | 16% | 24.4 | 18% | 23.5 |
| 30+ years job tenure | 9% | 35.6 | 12% | 34.3 |
| All employees age 55–64 | 13.4 | 14.2 | 13.7 | 13.3 |
| female | 45% | 11.4 | 46% | 12.1 |
| male | 55% | 16.6 | 54% | 14.3 |

*Sources: PENSIM values for employed individuals age 25–64 in 2000 are from PENSIM (1/20/06) run that generates one-in-a-thousand samples for each birth cohort whose individuals are in that age range during 2000; CPS values are tabulated using raw data from the February 2000 CPS, as documented in "Employee Tenure in 2000" a Bureau of Labor Statistics press release 00-245 dated August 29, 2000.*

is no evidence that PENSIM overestimates the prevalence of short jobs. This finding is relevant in assessing the validity of PENSIM simulations that show a sizable minority of cohort individuals accumulating no pension rights over their whole work career.

Another conclusion that can be drawn from this validation test is that there is no evidence that PENSIM overestimates the prevalence of long jobs. PENSIM finds 25 percent of employees age 55-64 in 2000 have held their current job for twenty or more years, but this is less than the 30 percent observed in the CPS data. This finding is relevant in assessing, for example, the validity of PENSIM simulations that show many would be worse off being covered by a typical cash-balance rather than typical final-pay pension plan (GAO 2005).

### 8.4.2 Cross-Sectional Distribution of Pension Sponsorship

Another aspect of job histories is whether or not an individual's employer sponsors a pension. If the answer is no, then obviously no pension accumulation will happen on that job. The less than universal sponsorship of pensions by employers makes this a critical aspect of job histories.

Both an individual's attributes and the job's characteristics affect the likelihood that the employer of the individual holding that job sponsors a pension. Here we compare numerous statistics on pension sponsorship calculated using a cross-sectional sample for 2003 generated by PENSIM and cross-sectional SIPP data gathered during January-April 2003 as part topical module 7 for the 2001 panel. The SIPP tabulations are taken from a recent Employee Benefit Research Institute publication (Copeland 2006).

Simulated and actual pension sponsorship rates are compared for all employees age 16+ and for subgroups defined by individual attributes (age, gender, education) and job characteristics (firm size, earnings, unionization, industry sector, job tenure). The relative size of these various subgroups, as well as each subgroup's pension sponsorship rate, are compared in Table 13 and Table 14.

The first thing to note is that PENSIM estimates the number of employees age 16+ in 2003 to be about 135 million, while the SIPP estimate is 126 million. This is puzzling because the PENSIM employment estimates for 2000 (see Table 12) and for 2004 (see Table 11) are in close agreement with the CPS estimates for those years. Tabulation of the CPS March 2005 supplement data for 2004 produce an estimate of 146 million employees age

Table 13: ***2003 Cross-Sectional Pension Sponsorship Statistics — Part 1.*** *Statistics include all employed individuals who are at least 16 years old in 2003. Number of employed is expressed in millions and pension sponsorship rate in percent.*

| Category of Employee | PENSIM number | PENSIM rate | SIPP number | SIPP rate |
|---|---|---|---|---|
| All employees age 16+ | 135 | 69 | 126 | 67 |
| 0–1 years job tenure | 13% | 57 | 18% | 54 |
| 1–5 years job tenure | 33% | 63 | 35% | 64 |
| 5–10 years job tenure | 23% | 71 | 19% | 71 |
| 10–15 years job tenure | 14% | 76 | 11% | 76 |
| 15+ years job tenure | 17% | 84 | 17% | 80 |
| All employees age 16+ | 135 | 69 | 126 | 67 |
| age 16–20 | 4% | 52 | 6% | 39 |
| age 21–30 | 22% | 67 | 21% | 61 |
| age 31–40 | 24% | 72 | 24% | 70 |
| age 41–50 | 26% | 71 | 26% | 73 |
| age 51–60 | 20% | 70 | 17% | 75 |
| age 61+ | 4% | 68 | 6% | 60 |
| All employees age 16+ | 135 | 69 | 126 | 67 |
| private firm size 1–24 | 22% | 33 | 22% | 31 |
| private firm size 25–99 | 9% | 58 | 11% | 57 |
| private firm size 100+ | 52% | 79 | 50% | 79 |
| public sector | 17% | 93 | 17% | 87 |
| All employees age 16+ | 135 | 69 | 126 | 67 |
| female | 44% | 70 | 48% | 68 |
| male | 56% | 69 | 52% | 67 |

*Sources: PENSIM values for employed individuals age 16+ in 2003 are from PENSIM (1/20/06) run that generates one-in-a-thousand samples for each birth cohort whose individuals are in that age range during 2003; SIPP values are from tabulations by Craig Copeland reported in EBRI Issue Brief 289 (January 2006), Figure 1 on pages 6–7.*

Table 14: ***2003 Cross-Sectional Pension Sponsorship Statistics — Part 2.*** *Statistics include all employed individuals who are at least 16 years old in 2003. Number of employed is expressed in millions and pension sponsorship rate in percent. Annual earnings categories are expressed in thousands of dollars.*

| Category of Employee | PENSIM number | PENSIM rate | SIPP number | SIPP rate |
|---|---|---|---|---|
| All employees age 16+ | 135 | 69 | 126 | 67 |
| earnings 0–5 | 7% | 49 | 8% | 43 |
| earnings 5–10 | 12% | 58 | 9% | 47 |
| earnings 10–15 | 11% | 65 | 12% | 50 |
| earnings 15–20 | 10% | 68 | 12% | 61 |
| earnings 20–25 | 8% | 69 | 11% | 68 |
| earnings 25–30 | 8% | 71 | 10% | 74 |
| earnings 30–50 | 21% | 74 | 22% | 81 |
| earnings 50+ | 23% | 79 | 16% | 85 |
| All employees age 16+ | 135 | 69 | 126 | 67 |
| unionized job | 17% | 89 | 13% | 91 |
| non-unionized job | 83% | 65 | 87% | 64 |
| All employees age 16+ | 135 | 69 | 126 | 67 |
| private-sector job | 83% | 64 | 83% | 63 |
| public-sector job | 17% | 93 | 17% | 87 |
| All employees age 16+ | 135 | 69 | 126 | 67 |
| high school dropout | 14% | 51 | 11% | 42 |
| high school graduate | 34% | 65 | 28% | 63 |
| some college | 30% | 74 | 32% | 69 |
| four-year college degree | 16% | 80 | 19% | 78 |
| advanced graduate degree | 7% | 80 | 10% | 83 |

*Sources: PENSIM values for employed individuals age 16+ in 2003 are from PENSIM (1/20/06) run that generates one-in-a-thousand samples for each birth cohort whose individuals are in that age range during 2003; SIPP values are from tabulations by Craig Copeland reported in EBRI Issue Brief 289 (January 2006), Figure 1 on pages 6–7.*

16–64. This would suggest that the SIPP number of employees is underestimated. Because the total employment estimates differ by so much, the size of the various subgroups are given in percentage terms.

The overall pension sponsorship rate is simulated to be 69 percent, which is somewhat higher than the 67 percent observed in the SIPP data. In both simulated and actual data, the pension sponsorship rate rises with job tenure, age (except for the 61+ subgroup), private-sector firm size, earnings, and education. In addition, both simulated and actual pension sponsorship rates are much higher in public-sector jobs relative to private-sector jobs and in unionized jobs relative to non-unionized jobs. And in both simulated and actual data, men and women have about the same pension sponsorship rate.

A close examination of these comparisons shows some differences in the size of the subgroup and their pension sponsorship rates. However, the overall conclusion is that PENSIM simulates distributions of individual attributes and job characteristics that are not unrealistic, and furthermore, the the model produces reasonable correlations between these attributes and characteristics, on the one hand, and the rate of pension sponsorship, on the other hand.

# 9 Calibration of PENSIM Pensions

*The calibration of historical employer pension-offering rates has not yet been completed. Until this task is completed, PENSIM assumes that the 1996-98 employer pension-offering probabilities (Holmer and Janney 2003) extend back in time. This temporary state of affairs clearly causes distortions in the simulated pension experience of individuals born before about 1970. Perhaps the main distortion is an overestimate of the prevalence of defined-contribution pensions for these older cohorts. The results of the historical calibration task will be reported in this chapter when the task has been completed.*

# 10 Validation of PENSIM Pensions

This chapter of the overview presents results from several validation tests. In each test, statistics calculated with simulated data from a PENSIM run are compared with conceptually equivalent statistics calculated with actual data. The validation tests focus on aspects of pension accumulation and pension withdrawal (which results in pension income during retirement). When considering the results of these validation tests, the issues discussed in the validation methods section of the prior chapter should be kept in mind.

## 10.1 Validating Pension Accumulation Results

This section of the chapter presents results from validation tests of statistics related to the accumulation of pensions during an individual's work career.

### 10.1.1 Pension Participation

In PENSIM, pension participation means that an individual who is eligible for a plan actually participates in that plan. For defined-benefit plans, all eligibles participate by definition. For defined-contribution plans with a non-matching employer contribution and no minimum employee contribution (typically money-purchase or profit-sharing plans), PENSIM assumes that all eligibles participate. For the remaining (most) defined-contribution plans, PENSIM simulates the participation process as described on page 13. Here we compare simulated participation rates for private-sector defined-contribution plans in 2004 to participation rates reported by large plan administrators. Participation rates calculated from government surveys are not comparable because the survey cannot determine which individuals are eligible for a plan.

In PENSIM, participation in a defined-contribution plan means that the individual has a plan account. This means that a PENSIM participation rate is the fraction of defined-contribution eligibles who have established a plan account. On the other hand, plan administrators (Fidelity 2005, Vanguard 2005) typically define the participation rate to be the fraction of eligibles who make a contribution to the plan that year. We can translate the contribution-oriented participation rates into PENSIM participation rates by using the estimate that eight or nine percent of those with a plan account make no contribution in a year (Holden and VanDerhei 2005, footnote 21).

Table 15: ***2004 Cross-Sectional Participation Rate in Private-Sector Defined-Contribution Plans.*** *Estimated rates are expressed in percentage terms.*

| Category of Employee | PENSIM rate | Fidelity rate | Vanguard rate |
|---|---|---|---|
| All plan eligibles | 66 | 65 | 69 |

*Sources: the PENSIM value for individuals age 16–69 who are eligible for a private-sector defined-contribution plan in 2004 is from PENSIM (7/24/06) run that generates one-in-a-thousand samples for each birth cohort whose individuals are in that age range during 2004; the Fidelity value of 59 percent (*Building Futures, Volume VI*) and the Vanguard value of 63 percent (*How America Saves 2005*) are both divided by 0.91 to adjust for an assumed nine percent who make no contribution as described in the text.*

Simulated and actual private-sector defined-contribution participation rates for 2004 are presented in Table 15.

The simulated and actual aggregate participation rates are roughly the same. The variation in experience between the individuals whose accounts are administered by Fidelity and Vanguard is larger than the simulated-actual differences. Although not shown in the table, the age and earnings variation in participation rates are similar in the simulated and actual data. It is not possible to make exact comparisons of participation rates for these subgroups because the fraction in each subgroup who make no contributions is unknown, and is likely to vary substantially across subgroups.

### 10.1.2 Other Aspects of Pension Accumulation

*Validation tests of other aspects of pension accumulation will be added in the future.*

## 10.2 Validating Pension Withdrawal Results

*The validation of simulated pension income in retirement has not yet been completed because employer pension-offering rates in the years before 1996–*

*98 have not yet been calibrated to historical data (see Chapter 9). Once the calibration has been completed, the validation will compare pension income in retirement simulated by PENSIM for individuals born in the late 1930s with recent historical data on pension income in retirement for individuals born in those same years. The results of this historical validation task will be reported here when the task has been completed.*

# A Appendix: PENSIM Simulation Methods

*Note to the reader: the text in this appendix was written early in the PENSIM development process (probably no later than 2000) in response to a number of questions raised at a PENSIM review conference organized by the Department of Labor.*

Perhaps the most distinctive feature of PENSIM is the type of dynamic microsimulation methodology it uses to generate a sample of pension-related life histories. This appendix of the overview presents the rationale for the choice of this methodology. We begin by recounting the benefit-cost analysis that produced the choice of simulation methodology used in PENSIM. The remaining discussion touches on several other methodological issues that arise given that basic choice.

## A.1 Benefit-Cost Analysis of Model Design

There is no one kind of simulation model that is optimal for all policy analysis. Policy simulation models differ in their methodology and structure so that they can specialize in estimating the reform effects of certain kinds of policies. Our thinking about what kind of model is best suited to analyze government policy concerning employer-sponsored pensions begins by identifying exactly what sort of analytical capabilities are required. Given these policy analysis goals, we identify two dynamic microsimulation methodologies to assess. The assessment involves weighing the strengths and weaknesses of each methodology given the goals.

### A.1.1 Policy Analysis Goals

Given the fact that employer-sponsored pension coverage accumulates over a work career that may involve jobs with numerous employers, some of whom offer no pension and others of whom offer different kinds of pensions, it is clear that a successful pension policy simulation model must simulate an individual's *complete job history*. Such a job history must include not only the duration of and earnings on each job, but also the characteristics of any pension offered on a job and the individual's decisions regarding each offered pension. Notice that a lifetime earnings history, which is extremely useful for social security policy analysis, is essentially useless in this context because it provides no information about the duration of the jobs or the pension offerings on those jobs. Conversely, a lifetime job history implies a complete earnings history.

In addition, a successful model must simulate a *pension income history* that describes the level of pension income received by the individual during each year of retirement. Because different kinds of pensions provide income over retirement years in different patterns, any comprehensive set of pension outcome measures must be based on such a pension income history. For example, a pension that provides a retirement annuity is likely to produce a pension income history that is very different from a similar sized pension that provides a single lump sum at retirement. Even if the expected present values of the two retirement income streams are equal, there is likely to be a difference between the income variability (over retirement years and different macroeconomic scenarios) of the pension income histories generated by the two pensions. In order to represent the different levels of risk inherent in these two pay-out schemes, a complete pension income history needs to be simulated and used in the calculation of pension policy outcome measures for the individual.

Because an individual's pension accumulation occurs over the course of a work career and pension withdrawal occurs over all retirement years, pension outcome statistics are intrinsically lifetime statistics that cannot be calculated until the end of an individual's life.

The lifetime nature of pension outcome statistics puts them in the same class as life expectancy statistics. Period life expectancy at birth can be thought of as a statistic calculated from the results of a dynamic microsimulation of the death age of an individual under the assumption of no change in current age-specific mortality rates. In such a simulation, Monte Carlo methods are used to randomly realize the individual's age at death given the age-specific mortality rates. The mean death age, averaged over all the Monte Carlo replications or scenarios, is the period life expectancy at birth, which summarizes the lifetime effects of current age-specific mortality rates on longevity. Notice that the life expectancy statistic cannot be calculated unless complete lifetimes are simulated. The same is true of pension outcome statistics.

An important implication of these first two goals is the need to simulate an individual's lifetime history — that is, a complete job history plus a pension income history — in order to compute outcome measures. Simulation of pensions is, therefore, very different from that of income taxes or health insurance where sensible outcome measures can be calculated based on the annual experience of an individual. Unlike the pension situation, such annual outcome statistics can be calculated for an individual of any age using a static microsimulation model.

It is also clear that we need a *representative sample* of these lifetime histories to represent variation in pension experience. Reforms in pension policy will have different effects on individuals that have different kinds of lifetime histories. Use of a representative sample permits analysis of the distributional details of a reform's impact.

Taken together, these three goals imply that we need a dynamic microsimulation model that generates a representative sample of lifetime histories. Neither a dynamic aggregate model nor a static microsimulation model will support the required analysis.

We identify one additional high-level goal for pension policy analysis. The model must be able to characterize not only individual, or idiosyncratic, risks (e.g., that an individual may die only a year after retiring) that have traditionally been the focus of dynamic microsimulation models, but also collective, or systemic, risks (e.g., that the rate of return on the S&P 500 index may be less than –30% in a given year) that are experienced by everyone in the simulation sample. The rise in importance of defined-contribution pension plans over the past two decades makes it imperative to represent *asset return risk*, which is collective to a large extent. If asset return risk is ignored in the simulation, there will be no way to assess the validity of claims that the declining importance of defined-benefit plans and the increasing importance of defined-contribution plans is raising the uncertainty in an individual's pension income to undesirable levels. The same issue has arisen in social security policy analysis with the recent discussion of introducing defined-contribution self-managed accounts into the program.

Collective asset return risk should not be confused with individual portfolio risk. The latter stems from differences across sample individuals in the asset composition of their defined-contribution pension portfolios. The former derives from the fact that future asset returns are uncertain in the same way for everyone, and hence, each sample individual views defined-contribution pension accumulation as an uncertain process. In order to effectively represent this kind of collective risk, the normal process of simulating lifetime histories for all sample individuals needs to replicated a large number of times with each replication using a time series of asset returns that is drawn randomly from an assumed distribution. In this way, the probability distribution of a pension outcome statistic (e.g., the percent of sample lifetimes with pension income below some low threshold level) can be estimated. This distribution would be useful in determining if different kinds of pensions produce significantly different outcome distributions, indicating their differing abilities in cushioning individuals against the risk of uncertain asset returns. We will address below the implications of the substantial increase in computation time caused by using this Monte Carlo representation of collective risk in the simulation model.

This concludes our discussion of the major goals we have identified for our development of a pension policy simulation model. Models with other policy focuses might have different goals. These four high-level goals are the key issues considered in the following benefit-cost analysis.

### A.1.2 Simulation Methodology Options

We are aware of only two dynamic microsimulation methodologies that can be considered in the benefit-cost analysis.

The first option is the conventional methodology (Orcutt et al. 1976) that begins with a population sample from a survey and advances the individuals in that sample one year at a time. Typically, annual transition probabilities are estimated to support this sort of annual advance simulation. The advance of the whole sample one year at a time allows population totals to be calculated each year, thereby creating the option for the model to be designed so that those totals affect the simulated behavior of individuals in that year through some kind of market equilibration scheme. When forced to describe this method in a phrase, we awkwardly call it the population-advance-one-year-at-a-time methodology (or sometimes, more simply, the conventional methodology).

A number of retirement-related models have been developed over the years using this conventional methodology (Ross 1991), and recently several new efforts have adopted this methodology.

The second option has been developed at Statistics Canada (Wolfson 1995, Wolfson et al. 1997) as part the LifePaths model development effort. This methodology involves simulating individuals in a birth cohort from the beginning of life to the end of life. Each individual's life is simulated by moving from one life event to the next life event (implying variable advances in time) until the individual dies. Hazard functions estimated with various longitudinal data sets are used to simulate the waiting times between events. This kind of discrete-event simulation is standard in areas like operations research (Fishman 1978, among many others), but is considered novel in other areas like economics. There is no use of a population sample from a survey as the starting point of the simulation because simulated behavior begins at birth, generating a synthetic sample of those born in a particular year. When forced to describe this method in a phrase, we awkwardly call it the cohort-lifetime-one-event-at-a-time methodology (or sometimes, more simply, the alternative methodology).

### A.1.3 Weighing Options Given Goals

In a benefit-cost analysis, the preferred option is the one with the largest net benefit (i.e., benefits minus costs). We need to estimate the benefits by assessing the degree of success in reaching our four high-level goals when using the two methodologies, as well as the costs involved in using the two methodologies.

Cost has been an important issue from the beginning of this project. The objective of creating a pension policy simulation model that could perform some limited policy analysis at early stages of development and could be developed at a modest total cost has been a critical issue for two reasons. First, a large-scale effort was ruled out because of the recommendations of a National Academy of Sciences panel to avoid a large-scale model development effort and instead develop "spreadsheet tools" (Citro and Hanushek 1997, page 160). A second reason was that the federal government has allocated the funding agency only a modest sized policy research budget.

There are really just three key issues in the benefit-cost analysis. We frame these issues in terms of questions about the advantages or disadvantages of the conventional population-advance-one-year-at-a-time methodology in comparison to the alternative cohort-lifetime-one-event-at-a-time methodology. We adopt this framework because most retirement modeling projects have adopted the conventional methodology.

The three questions are as follows. First, is the use of a starting population sample in the conventional methodology an advantage or disadvantage in the context of employer-sponsored pension policy analysis? Second, is the use of annual transition probabilities in the context of advancing one year at a time an advantage or disadvantage in simulating complete job histories? And third, is the use of advance-one-year-at-a-time for all individual behavior so consuming of computer time that it makes the simulation of collective risks (like uncertain asset returns) computationally impractical or too costly?

Consider the *first question* in the benefit-cost analysis: Is the use of a starting population sample in the conventional methodology an advantage or disadvantage in the context of employer-sponsored pension policy analysis? There is no doubt that the use of a population sample in the conventional methodology is a strong advantage when the simulation horizon is short and when meaningful outcome statistics can be calculated for all members of the population whether they be young or old. This advantage is rooted in the ability of the sample to represent the variety of individuals in the population and the complex correlations between individual attributes. But as the simulation horizon lengthens, the population sample gradually becomes a synthetic sample whose features are determined by the behavioral equations embedded in the model rather than by the nature of the starting population sample. And if lifetime (rather than annual) outcome statistics are required, the population sample contains, at any moment of time, mostly individuals for whom outcome statistics cannot be calculated because their lifetimes are not yet complete. So, the major advantage of using a population sample to start the simulation disappears when long term simulation of complete lifetimes is required.

In addition, the use of a population sample raises the question of whether the survey data include the prior job history (including its pension aspects) for each sample individual who has already started a work career at the time of the survey. We are aware of no surveys that include such a job history prior to the survey date. The Health and Retirement Study (HRS) data do include information on up to three prior jobs that lasted at least five years and involved some kind of pension coverage, which is a major improvement over earlier surveys. Such data, while still somewhat incomplete, are useful in projecting pension income for the surveyed cohorts in retirement, but not very useful for estimating the effects of policy reforms that change employers' pension offerings.

It seems to us that the only way to proceed with the conventional methodology would be to impute this prior job history. But when considering how this might be done, we conclude that the only reasonable way to do the imputation would be to simulate the job history from the first job up to the job held at the time of the survey. And because the simulation model needs to simulate such a job history from the survey date until retirement, there seems no advantage to using a population sample as the starting point for an employer-sponsored pension policy simulation model.

In fact, we conclude that the use of a population survey would add little or no benefit to a pension policy simulation model, but would substantially increase its development cost. The key difference in modeling philosophies between these two methodologies is a different view about the most effective way to incorporate the implications of survey data into a dynamic microsimulation model. The population-advance-one-year-at-a-time methodology emphasizes using survey data as the starting sample for the simulation, while the cohort-lifetime-one-event-at-a-time methodology emphasizes using numerous longitudinal surveys to estimate event-timing equations, which when incorporated in the model simulate complete life histories beginning at birth. While some of this difference in view may be rooted in differences in research style, there are also substantive differences. Clearly, the conventional methodology is likely to be the better choice for short-term (or static) simulations of annual (rather than lifetime) outcome statistics, but it turns out that pension simulation does not fall into this category.

Now we consider the *second question* in the benefit-cost analysis: Is the use of annual transition probabilities in the context of advancing one year at a time an advantage or disadvantage in simulating complete job histories? At the start of the project, we felt that estimating hazard functions for job duration and non-employment duration (i.e., the time between jobs) would produce more realistic job histories at a lower development cost than would the conventional approach used in other dynamic microsimulation models.

Since conducting this benefit-cost analysis in 1997, work on the Social Security Administration's MINT project has underscored the importance of non-employment spells in shaping earnings histories and the difficulty of simulating them in an advance-one-year-at-a-time context (Toder et al. 1999, Chapter 8 and Chapter 2, respectively). Our original assessment was an educated guess, but our subsequent experience has reinforced our original view: after a considerable initial effort, the MINT simulated earnings histories were flawed (Holmer 2000), while, after a relatively modest effort, the PENSIM simulated earnings histories are fairly realistic (see Chapter 8 below). We are not arguing that it is impossible to simulate realistic earnings histories using the advance-one-year-at-a-time methodology, only that it appears easier, and therefore, less costly to do so with estimated hazard functions and the lifetime-one-event-at-a-time methodology.

And finally, we consider the *third question* in the benefit-cost analysis: Is the use of advance-one-year-at-a-time for all individual behavior so consuming of computer time that it makes the simulation of collective risks computationally impractical or too costly? Here we make the judgment that the cost constraints of the project rule out developing a model that would require special computer hardware or software to run in a practical amount of time. We have experience developing a Monte Carlo simulation model that uses distributed processing techniques to execute parts of a portfolio-management simulation run on dozens of networked Unix workstations, a technique that was developed to avoid the costs of acquiring a supercomputer (Holmer 1994). This experience shows that dynamic microsimulation models developed with either methodology can be made computationally feasible when they incorporate collective risks, but that there are substantial costs when the computational threshold of a standard personal computer is exceeded. This observation implies that the conventional methodology is disadvantaged because its annual simulation of *all* behavior is computationally more intense than the alternative method of jumping through an individual's life one event at a time. In addition, as noted above, the conventional methodology suffers the computational burden of simulating partial lifetimes for a relatively large fraction of the population, whose pension outcome statistics cannot be calculated.

After pondering these three questions, we conclude that for the purpose of developing an employer-sponsored pension policy simulation model on a tight budget, the cohort-lifetime-one-event-at-a-time methodology is more promising than the population-advance-one-year-at-a-time methodology. At the same time, we realize that others might disagree because of differences in research style or differences in opinion about the magnitude of the benefits and costs being weighed. While it could be argued that we have not developed PENSIM far enough to say definitively that our benefit-cost analysis was correct, our experience so far has been supportive of the choice of the alternative methodology.

The rest of this chapter discusses a number of issues that arise given our choice of the alternative simulation methodology.

## A.2 Simulating Both Collective and Individual Risks

As mentioned above, it is essential for retirement-related dynamic simulation models to simulate both collective (or systemic) risks and individual (or idiosyncratic) risks. Here we describe briefly how we plan to do that in PENSIM.

Individual risks (including mortality risk, disability risk, and earnings risk, as well as various pension-related risks) are represented by using standard Monte Carlo simulation methods to simulate a sample of lifetime histories. Each simulated lifetime is somewhat different because a different stream of random numbers is used to realize the assumed risk probabilities during each lifetime. The degree of variability in lifetime outcomes represents the combined effect of the individual risks. This is the standard approach in both the conventional population-advance-one-year-at-a-time and the alternative cohort-lifetime-one-event-at-a-time methodologies.

Collective risks (e.g., asset return risk) expose all individuals to the same uncertainty. This means that a different sample of lifetimes must be simulated for each Monte Carlo realization of the collective-risk probabilities. This implies that if we want to represent collective risks with one thousand Monte Carlo replications or scenarios (to ensure that rare events like the 1929 collapse in equity prices are represented among the scenarios) and want to represent individual risks with one thousand lifetimes lived out under each collective-risk scenario, then we need to simulate one million lifetimes. The large number of simulated lifetime histories required to represent the joint effects of both collective and individual risks is why computational cost weighs heavily in our benefit-cost analysis of the advantages and disadvantages of the two simulation methodologies.

The current version of PENSIM [*remember this was written early in the development process, probably around the year 2000*] generates life histories for a cohort sample of 100,000 individuals, and all their spouses, in less than three minutes on a standard personal computer (i.e., a 1.0 GHz Pentium 3 running Windows 2000). This execution speed implies that less than half an hour is required to simulate one million sample life histories. While the current version of PENSIM does not incorporate all planned pension calculations, it is appears unlikely that adding the planned pension calculations will cause PENSIM execution time to increase by more than computer speed is expected to increase during the next year or two. It seems likely, therefore, that even after all planned pension calculations have been implemented, PENSIM will be able to simulate a million sample life histories in less than thirty minutes on a single inexpensive personal computer.

A computationally efficient method of jointly realizing collective and individual risks provides us with the ability to calculate an average experience (by computing the mean of realized experiences) as well as the likelihood of atypical experiences (by tabulating the fraction of realized experiences that are atypical). Knowing the likelihood and nature of atypical experiences is essential if we want to assess how individuals rank pension policies that generate different probability distributions of lifetime experiences. For an extended discussion of issues involved in using Monte Carlo simulation results to assess how individuals rank retirement-related policies that produce uncertain life outcomes, see Holmer (SSASIM Guide, Section 1.5: Risk Adjusting Policy Performance Measures).

## A.3 Synthetic Cohort as Longitudinal Data Matching

Another way to think about the alternative cohort-lifetime-one-event-at-a-time methodology is that it is a form of statistical data matching for longitudinal data. Statistical matching of cross-sectional data has a long history in static microsimulation modeling (Cohen 1991), and the alternative dynamic microsimulation methodology can be viewed as an extension of that work into the realm of longitudinal data. In fact, the realization that no single longitudinal survey would be broad enough to gather information on all relevant behavior or be long enough to cover full lifetimes has been a major motivation in the development of the alternative methodology at Statistics Canada (Wolfson 1995).

It is important to note that the longitudinal data matching character of the alternative simulation methodology is distinct from its simulation of a synthetic cohort sample from birth. The technique of estimating hazard functions with various special-purpose longitudinal data sets, and then combining these hazard functions in a single model, has been used in the Social Security Administration's MINT project to simulate the remaining demographic life events of an actual population sample (Panis and Lillard 1999).

If there were a population sample that included all retirement-related life events that occurred before the survey was conducted, then a hybrid methodology, which combines that comprehensive survey sample with lifetime-one-event-at-a-time methods, would be worth considering. As mentioned above, the fatal shortcoming of such a hybrid methodology is that there are always significant elements of the life history that are not present in the actual population sample, requiring some kind of statistical imputation of the missing data. And once we start contemplating this sort of activity, we find ourselves contemplating the synthesis of significant parts of an individual's life history using longitudinal data matching methods to impute the missing data. Our approach is different only in degree; we think it is simpler conceptually, and less costly, to dispense with the actual population sample, and undertake a complete (rather than a partial) synthesis of lifetime histories.

The issues outlined in the previous paragraph are not hypothetical. The Social Security Administration's MINT project has come face-to-face with the issues raised by the fact that the MINT population sample (an exact match of early 1990s SIPP samples and administrative earnings histories) does not contain complete information about job histories before the survey. This means that if the MINT project is to simulate pension coverage over a lifetime, MINT researchers will have to impute pension coverage on past jobs as well as on jobs held after the survey date. After considering a number of options, MINT researchers obtained a sample of PENSIM simulated job histories to assist them in their task of simulating past and future job histories.

## A.4 Synthetic vs Actual Cohort Data

The first question most people ask when we say PENSIM generates synthetic lifetime histories for cohort members from birth is "How do you know whether the simulated life histories are realistic?" The answer is we undertake a series of validation tests to establish the realism of the synthetic cohort data. Each validation test compares statistics computed from the simulated data with the same statistics calculated using actual survey data. These statistics can be cross-sectional (e.g., the fraction of those in their forties who have zero annual earnings) or longitudinal (e.g., the correlation coefficient between annual earnings at age 37 and age 52). We present results from a series of PENSIM validation tests in Chapter 8.

It is important to realize that other retirement-related models that use the conventional population-advance-one-year-at-a-time methodology also need to undertake such validation tests to establish the realism of their simulated life histories. This is true because, as we discussed above, the actual population samples used in such models routinely require statistical imputation of missing data for years before the survey date and complete simulation of life histories for years following the survey date. This means that the complete life histories generated from actual population samples often contain substantial amounts of synthetic data, and therefore, should be subjected to validation tests to determine their realism. Using the conventional methodology to simulate life histories that are a mixture of actual and synthetic data is no guarantee of realism (Holmer 2000).

In Chapter 7 we show how a modest number of PENSIM input parameters can be calibrated to produce samples that resemble different historical birth cohorts, ranging from those born in 1935 to those born in 1985. Each cohort sample exhibits appropriately different statistics describing the experience of sample individuals in the areas of mortality, education, marital status, and employment. The capability of generating historically realistic samples of individuals born in a wide variety of years is a major advantage of the alternative cohort-lifetime-one-event-at-a-time methodology.

The conventional methodology is limited in this regard by the age of people sampled in the starting population survey. For example, the MINT project can simulate the lifetimes of those born between about 1930 and 1960, but not those born in subsequent years. Of course, a model using a starting population sample could overcome this disadvantage by simulating the birth of children to sample individuals, and then simulating those children's complete life histories. But if this approach is used to represent younger cohorts, then a completely synthetic sample of individuals is being generated, and whatever advantages the original population sample offered are lost.

The ability to generate realistic samples of younger birth cohorts is critical for pension policy analysis. This is because pension reforms are usually structured to leave those close to retirement unaffected, and are assessed on the basis of how they will affect individuals who experience the reform during a substantial part or all of their work careers. This means that to be useful a pension policy simulation model must be able to simulate realistic samples of individuals who are currently young: either part way through their work careers or at the start of their work careers.

## A.5 Implicit vs Explicit Variation in Attributes

We have been asked on different occasions why an individual attribute or job/employer characteristic is included or excluded from PENSIM. Given our tight development budget, we have conducted informal benefit-cost analysis on the inclusion of many different variables. In all cases the choice is between making the variable explicit or leaving it implicit as part of the unobserved heterogeneity that we have represented in a number of our estimated hazard functions.

Age, gender, education, marital status, and disability are explicit individual attributes in PENSIM. Our decision to leave race and ethnicity implicit is based on our view that discrimination is not a major problem in pension policy. Pension experiences vary substantially across the population, but this variation appears to be rooted in differences in income levels and job histories.

Explicit job characteristics include part-time/full-time status, unionization status, and earnings level. Explicit employer characteristics include industry, firm size, and the nature of pension offerings.

Making industry an explicit employer characteristic is important for a simulation model focused on pensions. The number and kind of pensions offered on a job depends, in an important way, on that job's industry, even after considering a number of other job characteristics.

In addition, making firm size an explicit employer characteristic is essential for a pension-related simulation model. Many researchers have observed that pension offering rates vary widely between smaller firms and larger firms. This observation has generated a substantial policy discussion that focuses on devising reforms (usually involving new tax expenditures) that seek to induce more small employers to offer pensions to their employees, and therefore, narrow the gap in offering rates.

The implication for model design is that the benefits of making both industry and firm size explicit characteristics of employers are quite high. Our choice of longitudinal data for estimating job/employer characteristic equations and job duration equations, and our choice of cross-sectional data to estimate the probability of employers offering different kinds of pensions, have been strongly influenced by our decision to make both industry and firm size explicit employer characteristics.

## A.6 Reduced-Form vs Structural Models of Behavior

Another model design issue is how to characterize individual behavior. Given our tight development budget, we have decided not to represent behavior with structural choice models, which represent decisions, like retirement, as the outcome of an explicit utility optimization under uncertainty (i.e., dynamic programming) problem (Rust 1994, Rust and Phelan 1997, for example). We decided at the beginning of the project in 1997 not to attempt this sort of approach, despite its theoretical appeal, because of its computation cost, which would severely limit the scope of behavior that could be included in the model.

Instead, reduced-form survival-analysis models, in which hazard functions are estimated with longitudinal survey data, are used in PENSIM. This representation is generally compatible with the structural choice models, leaving open the possibility of selectively incorporating those kinds of behavioral models in future versions of PENSIM. One possible use of structural choice models of behavior is private saving (Gourinchas and Parker 1999, Benítez-Silva 2000, for example), if that aspect of retirement-related behavior were to be added to PENSIM.

It is important to note that using reduced-form models does not necessarily imply that we are ignoring behavioral incentive effects. We have, for example, included pension offering by the current employer as a covariate in the job duration hazard function, and intend to include measures of an individual's pension situation in other behavioral equations where appropriate.

In the years since 1997, developments in the field of structural models and in pension research suggest that our original decision was sensible. Even if we were beginning the model development project now, the use of reduced-form models of behavior would be the preferred strategy. The computation cost of estimating structural models is still a serious problem that limits the scope of behavioral models that can be estimated. And in the years since 1997, pension research has become one of the most active areas of behavioral economics research, producing a range of findings that suggest the hyper-rational individual decision process typically assumed in structural models may not be a very accurate characterization of how individuals make pension-related decisions. This is a serious problem for those who argue that structural models are better than reduced-form models of behavior, as evidenced in the following quote:

> One can push this line of argument [that a structural model is better than a reduced-form model] only so far, since its validity depends on the assumption that agents really are rational expected-utility maximizers and the structural model is correctly specified. If we admit that a tightly parameterized structural model is at best an abstract and approximate representation of reality, there is no reason why a structural model necessarily yields more accurate forecasts than reduced-form models. (Rust 1994, p. 3086)

With computational costs still limiting the ability of researchers to correctly specify structural models and with mounting evidence from behavior research suggesting that individuals do not always make pension decisions like they are rational expected-utility maximizers, using reduced-form models of behavior in a large-scale simulation model such as PENSIM remains the prudent choice. While this is our own opinion, it must be widely shared because we know of no large-scale policy simulation projects underway that use structural models of behavior as their core building blocks.

It strikes us that a better use of structural models of behavior is to use them to test competing theories of how individuals make decisions. Instead of simply assuming individuals are expected-utility maximizers, it would be more scientific to test whether expected-utility theory does a better or worse job in explaining individual behavior than an alternative theory of choice under uncertainty, such as prospect theory. When this sort of testing was done in the context of health insurance choice, expected-utility theory was found to have more difficulty in predicting choices than a variant of prospect theory (Marquis and Holmer 1996).

## A.7 Simulating Couples and Marital Behavior

PENSIM simulates an individual's changing marital status and the personal attributes and job history of each spouse. This means that PENSIM can be used to analyze a couple's pension income during retirement. Here we discuss briefly the general approach used to simulate marital status and spouses.

When PENSIM was reviewed at the end of 1999, several reviewers pointed out the desirability of simulating not just individuals, but couples. And one reviewer suggested that when that task was undertaken, our reliance on the cohort-lifetime-one-event-at-a-time methodology would cause difficulties because it would not be possible to model marriages and remarriages as an annual marriage market (i.e., a matching process between unmarried individuals in the simulation sample). We have not pursued that approach because, even if the marriage market concept is thought to be more desirable on some theoretical grounds, we are concerned by the magnitude of its computational burden and the plausibility of its marital matching assignments when the number of unmarried sample members happens to be small.

The current version of PENSIM simulates the timing of marriage, divorce, and remarriage, using the same estimated hazard functions used in the Social Security Administration's MINT project (Panis and Lillard 1999). Incorporating these MINT results directly into PENSIM is possible because the Panis-Lillard work uses the one-event-at-a-time methodology, rather than the one-year-at-a-time methodology, to simulate the timing of weddings and divorces. And also in the spirit of the Panis-Lillard MINT work, the age and education attributes of a sample individual's spouse are represented in PENSIM by an extra-sample individual, whose simulated jobs and pensions contribute to the sample individual's pension experience, but who is not included in the sample for purposes of calculating outcome statistics.

In addition to simulating marital status, we have added the capability of simulating childbirths (see Appendix B, Section B.14, for details). Our initial experience with this approach to simulating family dynamics has been positive. The use of hazard functions to simulate the timing of weddings, childbirths, and divorces, and the use of extra-sample individuals to represent spouses, has been cost-effective to implement and is capable of generating a wide range of historical behavior (see Chapter 7).

# B Appendix: PENSIM Behavioral Events

The events included in PENSIM represent both individual and employer behavior that is relevant to pension policy. Even though PENSIM does not represent employers as explicit entities whose lifetime is simulated, it does include a number of events that represent employer behavior related to jobs and pensions. The joint effect of these employer events and individual events in determining a person's simulated lifetime experience is PENSIM's representation of the real-world interactions between individual and employer decisions that determine pension coverage and adequacy.

The events included in PENSIM are discussed roughly in the order that they occur during an individual's life. The activities associated with each event include updating individual attributes, possibly scheduling the timing of other events in the future, and possibly canceling previously scheduled events that are no longer logically possible. There is no mention below of the fact that every event's activities include recording the age at which that event occurred in the individual's lifetime and possibly canceling pending events that are no longer logically possible.

This appendix does not include detailed parameter estimation results for each behavioral event. Those statistical estimation results are presented in Appendix C. Section and page references will be supplied here in Appendix B to allow the reader to jump to the relevant part of Appendix C on PENSIM input parameters. This appendix has been designed so that reviewing the model's behavioral parameters is optional in the sense that a reader who does not look at the input parameters in Appendix C will not be hindered in reading or understanding the material presented here in Appendix B.

## B.1 Birth Event

The birth event begins a simulated individual life at age zero. The event specifies the individual's gender and schedules the timing of the high-school-start event, the disability-scheduling event, the retirement event, the cross-sectional-survey event, and the first of annual possible death events.

### B.1.1 Specify Gender

The proportion of male and female live births in a cohort is determined by an input parameter, the value of which is derived from 1992 gender-specific data on the number of infants less that one-year old and first-year mortality rates from the Social Security Administration's Office of the Chief Actuary. The parameter value used in all PENSIM runs implies that about 48.76 percent of newborns are female

#### B.1.2 Possibly Schedule Immigration Event

If a cohort individual is assumed to be foreign-born, an immigration event is scheduled at the appropriate age. The number of people in the cohort sample who immigrate at each age is calculated using information passed from SSASIM to PENSIM via the environment (env) file.

#### B.1.3 Possibly Schedule Emigration Event

If a cohort individual is assumed to emigrate (either after being native-born or after being foreign-born and immigrating), an emigration event is scheduled at the appropriate age. The number of people in the cohort sample who emigrate at each age is calculated using information passed from SSASIM to PENSIM via the environment (env) file in conjunction with emigration assumptions drawn from Polisim.

#### B.1.4 Schedule High-School-Start Event

The high-school-start event is scheduled at age fourteen for everyone in the birth cohort in all PENSIM runs.

#### B.1.5 Schedule Disability-Scheduling Event

In all PENSIM runs, the disability-scheduling event is scheduled at age thirty, at which time the disability event is scheduled using information about the individual's educational attainment.

#### B.1.6 Possibly Schedule Cash-Account-Conversion Event

If a conversion from private-sector terminal-earnings plans to cash-account plans is being simulated, a conversion event is scheduled at the beginning of the year in which the cohort individual is the age specified for the cash-account conversion. Note that cash-account the the plan type used in the

Bureau of Labor Statistics' National Compensation Survey to describe what others call cash-balance or hybrid plans.

### B.1.7 Schedule Retirement Event

The retirement event is scheduled to occur at gender-specific ages specified in the input database. In all the PENSIM runs executed to date, the retirement age for both genders is set to sixty-five, but this can easily be changed to determine the sensitivity of simulation results to the assumed retirement age.

This simple specification of retirement behavior will be made more complex in future versions of PENSIM. One possibility would be to characterize retirement decision-making as a sequence of annual calculations, beginning at a certain age, that determine the relative advantages of continuing to work at a job versus quitting that job and accepting a pension. This sort of retirement decision algorithm is in the spirit of the dynamic programming models of retirement behavior pioneered by Rust (1989).

### B.1.8 Schedule Cross-Sectional-Survey Event

The timing of the cross-sectional-survey event is generated stochastically by randomly selecting a survey age from an age range specified in the input database. Only one survey event is scheduled during the lifetime of each cohort individual. By sampling members of the birth cohort at different ages, PENSIM can produce statistics that are comparable with those generated from cross-sectional surveys (e.g., the Current Population Survey).

Rather than attempting to summarize the results of the surveys, a file that contains the individual survey responses for each individual in the cohort can be included among the PENSIM output results files. Then after the PENSIM run is finished, that file can be tabulated in any desired manner using whatever statistical analysis software is convenient.

### B.1.9 Schedule Death Event

Potential death events are scheduled just before each birthday. The mortality probabilities that are used to determine whether or not the potential event becomes real are based on the 1992 age- and gender-specific annual mortality rates supplied by the Social Security Administration's Office of the Chief Actuary. These mortality rates are for single-year age categories and extend up to 125 years of age. PENSIM has the capability of scaling these mortality rates by gender-education groups (in order to represent class variation in mortality rates within a birth cohort, or to approximate cohorts born in other years) and the capability of reducing the mortality rates each year (in order to calculate cohort life expectancies rather than period life expectancies). In addition, there is a capability of specifying a disability multiple, which raises the mortality rates of those who are receiving Disability Insurance benefits.

Each annual potential death event calculates the appropriate mortality rate for the individual, based on the individual's gender, age, educational attainment, and disability status, and then uses that mortality rate to determine stochastically whether or not death actually occurs.

The basic mortality rates are presented in Appendix C, Section C.1, beginning on page 140.

## B.2 High-School-Start Event

The activities associated with high-school-start event, which occurs at fourteen in all PENSIM runs, include scheduling the high-school-finish event and possibly scheduling the first-job-start event.

### B.2.1 Schedule High-School-Finish Event

The integer age at which the individual is scheduled to finish high school is generated stochastically using gender-specific non-parametric hazard functions. These hazard functions are estimated using data from the 1991 and 1992 Survey of Income and Program Participation (SIPP) panels on individuals who were no more than forty years old when they underwent the education and training topical module questioning. The resulting estimation samples were in excess of fourteen thousand people for each gender.

The parameters of these non-parametric hazard functions are presented in Appendix C, Section C.2, beginning on page 141.

### B.2.2 Possibly Schedule First-Job-Start Event

The waiting time between the high-school-start event and the possible start of the individual's first job is generated stochastically using gender-specific non-parametric hazard functions estimated with the same data used to estimate the high-school-finish hazard functions. These SIPP data define the first job as one that involved either part-time or full-time work continuously over at least six months of time. The first-job-start event does not always happen while the individual is attending high school. If the first-job-start waiting time is longer than the high-school-finish waiting time, then it is assumed that the individual does not start a first job before finishing high school. Conversely, if the first-job-start waiting time is shorter than the high-school-finish waiting time, then it is assumed that the individual does start a first job while attending high school and then goes on to finish high school at the previously scheduled age.

The parameters of these non-parametric hazard functions are presented in Appendix C, Section C.6, beginning on page 164.

## B.3 High-School-Finish Event

The high-school-finish event involves specifying whether or not the individual graduated and possibly scheduling college-start and first-job-start events.

### B.3.1 Specify Whether Graduated

The same SIPP data used to estimate the waiting-time distribution of the high-school-finish event are used to estimate the age- and gender-specific fraction of those finishing high school with a diploma rather than dropping out. In other words, graduating and dropping out are considered competing risks.

These age-specific graduation probabilities are presented in Appendix C, Section C.2, beginning on page 141.

### B.3.2 Possibly Schedule College-Start Event

For those who graduate from high school, the waiting time until starting college (zero or more whole years) is generated randomly using gender-specific non-parametric hazard functions that are estimated using a subset of these same SIPP data on education histories. In order to limit the complexity of the model, those whose waiting times imply a college-start age of more than forty are assumed not to start college.

The parameters of these non-parametric hazard functions are presented in Appendix C, Section C.2, beginning on page 141.

### B.3.3 Possibly Schedule First-Job-Start Event

For those who have finished high school and have not started their first job, the waiting time between the high-school-finish event and the start of the individual's first job is generated in the same manner as the time between the finish of one job and the start of the next job. This waiting time distribution is discussed in Appendix C, Section C.9, beginning on page 180.

## B.4 College-Start Event

The activities associated with the college-start event include scheduling the college-finish event and possibly scheduling the first-job-start event.

### B.4.1 Schedule College-Finish Event

The waiting time between starting and finishing college is generated stochastically using gender-specific non-parametric hazard functions. These hazard functions are estimated using a college-entering subset of data from the 1991 and 1992 SIPP panels on individuals who were no more than forty years old when the education and training topical modules were administered.

The non-parametric hazard functions, as all of those used in the schooling and first-job event analysis, are defined using duration data measured in whole numbers of years. The parameters of these non-parametric hazard functions are presented in Appendix C, Section C.2, beginning on page 141.

### B.4.2 Possibly Schedule First-Job-Start Event

For those who start college without having held a first job, the whole number of years between the college-start event and the possible start of the individual's first job is generated randomly using gender-specific non-parametric hazard functions. These hazard functions are estimated with an appropriate subset of the SIPP education data. It is possible that an individual does not start a first job while attending college. If the first-job-start waiting time is longer than the college-finish waiting time, then it is assumed that the individual does not start a first job before finishing college. If, on the other hand, the first-job-start waiting time is shorter than the college-finish waiting time, then it is assumed that the individual starts a first job before finishing college, and then goes on to finish college at the previously scheduled age.

The parameters of these non-parametric hazard functions are presented in Appendix C, Section C.6, beginning on page 164.

## B.5 College-Finish Event

The college-finish event involves specifying whether or not an individual graduates with a four-year degree and possibly scheduling graduate-degree-finish and first-job-start events.

### B.5.1 Specify Whether Graduates with Four-Year Degree

The same SIPP data used to estimate the waiting-time distribution of the college-finish event are used to estimate the age- and gender-specific fraction of those finishing college with a four-year degree rather than a two-year degree or dropping out. In other words, graduating with a four-year degree and all other college outcomes are considered competing risks.

These age-specific four-year-college-degree probabilities are presented in Appendix C, Section C.2, beginning on page 141.

### B.5.2 Possibly Schedule Graduate-Degree-Finish Event

For those who finish college with a four-year degree, the whole number of years between the college-finish event and the possible completion of a graduate degree is generated randomly using gender-specific non-parametric hazard functions. These hazard functions are estimated with an appropriate subset of the SIPP education data. In order to limit the complexity of the model, those whose waiting times imply a graduate-degree-finish age of more than forty are assumed not to earn a graduate degree.

The parameters of these non-parametric hazard functions are presented in Appendix C, Section C.2, beginning on page 141.

### B.5.3 Possibly Schedule First-Job-Start Event

For those who finish college without having held a first job, the waiting time between the college-finish event and the start of the individual's first job is generated in the same manner as the time between the finish of one job and the start of the next job. This waiting time distribution is described in Appendix C, Section C.9, beginning on page 180.

## B.6 Graduate-Degree-Finish Event

Completion of a graduate degree changes the individual's highest attained level of schooling, but causes no other events to be scheduled in PENSIM's simplified lifetime logic.

The higher level of schooling has the possibility of influencing the characteristics of the individual's future jobs (for example, getting a higher paying or more secure job), but, as described below, the limited size of the SIPP samples used to estimate job characteristic and duration equations often forced the combination of graduate-degree holders with four-year-college-degree holders. Given this frequent consolidation of the two levels of schooling, the graduate degree has little explicit effect in the current version of PENSIM.

## B.7 Immigration Event

The activities at the immigration event involve specifying the individual's immigration age and possibly scheduling a first-job-start event.

### B.7.1 Specify Immigration Age

The cohort individual's immigration age is specified. Note that native-born individuals are assumed to have an immigration age of zero, while foreign-born immigrants have an immigration age ranging from one to eighty-four (with the gender-specific age profile of immigrants being the same as those assumed by the Office of the Chief Actuary in the Social Security Administration).

### B.7.2 Possibly Schedule First-Job-Start Event

If an immigrant is not disabled or retired and has started school, a first-job-start event is scheduled if one is has not already been schedule. The hazard-function model that is used to simulate the waiting time until the first-job-start event is described in Appendix C, Section C.6, beginning on page 164.

## B.8 Emigration Event

The activities at the emigration event involve specifying the individual's emigration age, schedules an immediate job-finish event if employed, and schedules immediate pension-claim events for any vested defined benefits that are not being received.

### B.8.1 Specify Immigration Age

The cohort individual's emigration age is specified. Note that individuals who do not emigrate are assumed to have an infinite emigration age.

### B.8.2 Possibly Schedule Job-Finish Event

If the emigrant is employed, either a first-job-finish or a subsequent-job-finish event is scheduled to occur immediately. The job-finish event schedules all appropriate pension events.

### B.8.3 Possibly Schedule Pension-Claim Event

Even if the emigrant is not employed, the individual may have earned vested defined benefits on prior jobs. If so, a pension-claim event is scheduled to occur immediately.

## B.9 Wedding-Scheduling Event

The wedding-scheduling event, which occurs when an individual is twelve years old, involve specifying the waiting time until the individual's first wedding.

### B.9.1 Schedule Wedding Event

The hazard-function model that is used to simulate the waiting time until the next wedding event is described in Appendix C, Section C.3, beginning on page 146.

## B.10 Wedding Event

The activities of the wedding event involve determining the attributes of the individual's spouse, rescheduling the next childbirth event, and scheduling the next divorce or widowed event.

### B.10.1 Specify Marital Status and Spouse Attributes

At the marriage event, the age difference and educational level of the spouse are simulated using probabilities tabulated from almost thirty years of data from the Panel Study of Income Dynamics (PSID). Given these two attributes of the spouse, PENSIM simulates a complete life history for someone from that birth cohort and with that schooling.

It is important to understand that this simulated spouse is not a member of the birth cohort sample being simulated by PENSIM. Put a different way, there is no "marriage market" operating in PENSIM. The spouses are extra individuals simulated to permit the tabulation of family statistics for the cohort sample members, who are the sole focus of the simulated results.

If that simulated spouse is already married at the time of the wedding, then PENSIM simulates another potential spouse until one is found to be unmarried at the time of the wedding. Unavailable spouses are not discarded, but saved for consideration at other weddings. This caching of unused potential spouses greatly accelerates the process of simulating spouses for the cohort sample members.

The first available spouse is wed to the sample individual. This bonding of the individuals involves removing the post-wedding marital and childbearing history of the spouse and recording the presence of any step-children that individuals may bring to the marriage.

The parameters of the spouse characteristics models used in PENSIM are presented in Appendix C, Section C.3, beginning on page 146.

### B.10.2 Schedule Childbirth Event

Using the woman's new marital status, the education level of her new husband, and updated age, the next birth event is scheduled, unless it is predicted to occur after menopause.

This scheduling replaces any birth already scheduled when the woman was single. This needs to be done because the woman's childbirth hazard function changes with marital status; it is higher for a married woman. In effect, a change in marital status is a risk that competes with childbirth.

The parameters of the education-specific proportional hazard functions used in PENSIM are presented in Appendix C, Section C.4, beginning on page 156.

### B.10.3 Schedule Divorce or Widowed Event

At the wedding, the waiting time until the divorce event is simulated using the gender-specific continuous-time log-linear hazard-function models estimated by RAND for the SSA MINT project (Panis and Lillard 1999, pages 28–30). The length of all marriages are simulated with the same two hazard-function models because the number of marriages is included as a model covariate.

The hazard-function model that is used to simulate the waiting time until the next divorce is described in Appendix C, Section C.3, beginning on page 146.

Also, at the wedding, the new spouse's life history has been completely simulated so that the spouse's death age is known. If the waiting time until the spouse's death age is less than the waiting time until divorce, then a widowed event (rather than a divorce event) is scheduled.

## B.11 Divorce Event

At the divorce event, PENSIM makes the conventional assumption that children born during the marriage go with their mother. Step-children from the marriage go with their biological parent. Also, the next childbirth event is rescheduled. And finally, the waiting time to the next wedding event is determined.

### B.11.1 Specify Marital Status

At the divorce event, the marital status of the individual is changed from married to divorced.

### B.11.2 Schedule Childbirth Event

Using the woman's new marital status and updated age, the next birth event is scheduled, unless it is predicted to occur after menopause. Again, when a woman's marital status changes, the model drops any scheduled birth event and reschedules with the woman's updated childbirth hazard function.

The parameters of the education-specific proportional hazard functions used in PENSIM are presented in Appendix C, Section C.4, beginning on page 156.

### B.11.3 Schedule Wedding Event

At the divorce event, the waiting time until the next wedding event is simulated using the gender-specific log-linear hazard-function models described in Appendix C, Section C.3, beginning on page 146.

## B.12 Widowed Event

At the widowed event, individual marital status is changed, pensions are possibly inherited, and a number of life events are scheduled or rescheduled.

### B.12.1 Specify Marital Status

At the widowed event, the marital status of the individual is changed from married to widowed.

### B.12.2 Specify Pensions Received or Inherited

At the widowed event, the surviving individual may be entitled to receive survivor's benefits from a defined-benefit pension or to receive a survivor's annuity payment. In addition, any rollover account balance owned by the deceased spouse is inherited by the surviving spouse.

### B.12.3 Schedule Wedding Event

At the divorce event, the waiting time until the next wedding event is simulated using the gender-specific log-linear hazard-function models described in Appendix C, Section C.3, beginning on page 146.

### B.12.4 Reschedule Childbirth Event

Using the woman's new marital status and updated age, the next birth event is scheduled, unless it is predicted to occur after menopause. Again, when a woman's marital status changes, the model drops any scheduled birth event and reschedules with the woman's updated childbirth hazard function.

The parameters of the education-specific proportional hazard functions used in PENSIM are presented in Appendix C, Section C.4, beginning on page 156.

## B.13 Childbirth-Scheduling Event

The childbirth-scheduling event, which occurs at age thirteen, schedules the first childbirth event using a hazard function that predicts the time between births (or between age thirteen in the case of the first birth).

Our fertility hazard function estimation shows that a woman's education level at age thirty is an important factor in the determining the timing between births. In order to know the value of this variable at earlier ages, PENSIM projects educational attainment ahead to age thirty at the start of each simulated life, and then returns to simulating life events in chronological order. There is no circularity involved here; the number of children born to a woman is not a factor in determining her educational attainment.

### B.13.1 Schedule Childbirth Event

A woman's age-thirty education level — high-school dropout, high-school graduate, some college, or graduated from college — determines which of four hazard functions used to schedule the next childbirth. Other factors that enter the hazard functions are marital status, husband's education level, birth cohort of the mother, age of the mother, and birth order (first child, second child, etc.).

Every time a childbirth event is scheduled, it competes with menopause. If the hazard function predicts that the next child would be born after age fifty, that birth is not scheduled and no more childbirths are scheduled for that woman.

The hazard functions are estimated with data merged together from several PSID files using the Cox regression routine in SPSS. The parameters of the education-specific proportional hazard functions used in PENSIM are presented in Appendix C, Section C.4, beginning on page 156, along with a description of their derivation using Cox regression.

## B.14 Childbirth Event

When PENSIM simulates a childbirth, the woman's birth-order number is incremented, and the model tries to schedule another childbirth. The model also determines at this point whether the event was a multiple birth. The only multiple birth the model recognizes is twins; the incidence of higher multiples is so low that it can be ignored or represented by making the incidence rate of twins slightly higher.

We take our data on the incidence of twin births from the Centers for Disease Control and Prevention (CDCP 1997), which show that the twin birth rate (i.e., the number of twin births to total live births) increased 30 percent, from 18.9 to 24.6 per 1000 live births, during the period 1980 to 1994. The rate of twin births rises with mother's cohort birth year, so we have PENSIM mimic that behavior.

### B.14.1 Schedule Childbirth Event

Using the woman's updated age and birth order, the model schedules the next birth event, unless it is predicted to occur after menopause.

The parameters of the education-specific proportional hazard functions used in PENSIM are presented in Appendix C, Section C.4, beginning on page 156.

## B.15 Disability-Scheduling Event

The disability-scheduling event, which occurs when an individual is thirty years old, involves the determination of whether or not an individual will become disabled sometime later in life. In PENSIM disabled means receiving social security Disability Insurance (DI). Whether or not this occurs sometime later in life is simulated using two sets of results from the MINT project and a simple algorithm to determine whether or not an individual is disability insured.

### B.15.1 Schedule Disability Event

The disability capabilities in PENSIM rely extensively on the disability work conducted for the Social Security Administration's MINT project. In order to become a DI beneficiary in PENSIM, an individual must become functionally disabled and then experience DI onset. The timing of the occurrence of functional disability is simulated using a continuous-time log-linear hazard function similar to that estimated for the MINT project (Panis and Lillard 1999, pages 31–32). We have used those estimated parameters, making only a small change in the constant term to represent all ethnic groups (because PENSIM does not simulate an individual's ethnicity or race).

The annual probability of DI onset is simulated using a constant hazard function in which the constant hazard rate has been calibrated so that PENSIM produces a DI onset incidence that is similar to that shown in the social security Trustees Report. We have implemented a crude disability insured test that checks that the individual has worked half of the time in the ten years prior to being simulated to become a DI beneficiary. We find, not surprisingly, that such a test causes very little reduction in the number of men simulated to become DI beneficiaries, but does result in a noticeable reduction in the number of women who are simulated to become DI beneficiaries.

The parameters of the functional disability hazard function and DI-onset probit model used in PENSIM are presented in Appendix C, Section C.5, beginning on page 162.

## B.16 Disability Event

The activities of the disability event involve the ending of the individual's current job, if the individual is still employed at the time the the disability event. In the current version of PENSIM, there is no recovery from disability, so a disability-recovery event is not scheduled.

### B.16.1 Specify Disability Status

At the disability event, the age at which the individual became disabled is recorded.

### B.16.2 Possibly Schedule Job-Finish Event

If the individual is employed at the time of the disability event, a job-finish event is scheduled to occur immediately. All pension events are schedule by the job-finish event.

## B.17 First-Job-Start Event

The first-job-start event specifies several characteristics of that job and of the employer who offers the job, and schedules the first-job-finish event, the first earnings-adjustment event, and possibly one or more pension-eligibility events.

### B.17.1 Specify Job and Employer Characteristics

The first job's characteristics are specified through the sequential stochastic realization of six characteristic equations that together form a recursive model of job and employer characteristics. Each of the six characteristics is estimated using 1992 SIPP data — a merger of data on the 1991 panel (wave 7) and the 1992 panel (wave 4) — to be a function of individual attributes measured at the start of the first job and other realized values of the first job's characteristics.

*Recursive Job-Characteristic Equations.* The structure of this recursive system of job and employer characteristics equations is as follows:

1. Employer Industry (I) depends on Individual Attributes (X)
2. Job Unionized (U) depends on X, I
3. Job Part-time (H) depends on X, I, U
4. Job Earnings (E) depends on X, I, U, H
5. Employer Size (S) depends on X, I, U, H, E
6. Employer Sponsors Pension (P) depends on X, I, U, H, E, S

The individual attributes (X) include gender, age, and educational attainment at the start of the first job. Education is a categorical variable with four values: attended high school, graduated from high school, attended college, and graduated from college with a four-year degree (including earned a graduate degree).

The employer's industry is represented by a one-digit SIPP industry code that classifies the employer in one of ten SIPP industry categories.

The job's unionization status is a binary variable that is defined as unionized if the individual holding the job is a union member or the job is covered by a union contract or collective-bargaining agreement, and non-unionized in all other situations.

The job's usual hours is used to create a binary variable that categorizes the job as part-time when usual weekly hours are less than 35 and full-time when they are 35 or more.

The job's usual monthly earnings is transformed into the natural logarithm of earnings, and therefore, is a continuous variable.

The employer's firm size is a discrete variable with three categories: 1–24, 25–99, and 100+ employees.

And finally, the employer's pension sponsorship status is represented by a binary variable that indicates sponsorship if the job's employer sponsors any kind of pension plan (even if the person holding the job is not eligible or does not participate in the plan), and non-sponsorship if the employer does not offer a pension plan to any employees.

Of the six equations estimated with 1992 SIPP data on first jobs, five have categorical dependent variables and one has a continuous dependent variable. The five equations with categorical dependent variables are specified as multinomial logit models (Maddala 1983, pages 34–37), while the log-earnings equation is specified as a linear regression model.

The six equations are estimated in the order they appear above, with the order being determined by the sample size available to estimate each equation. Given the limited number of times that questions about employer size and pension sponsorship are asked over the course of SIPP interviews, the number of observations on first jobs that have data on size and sponsorship is sharply lower than the number with data on industry, unionization, hours, and earnings.

The estimated parameters of these six equations are presented in Appendix C, Section C.7, beginning on page 167.

### B.17.2 Possibly Specify Pension Plan Offering

If an employer is a pension sponsor, the employer can have one of three kinds of pension offerings: sponsoring only a defined-benefit (DB) pension plan, sponsoring only a defined-contribution (DC) pension plan, or sponsoring both a DB and DC pension plan.

The probabilities of a sponsoring employer having one of these three kinds of offerings are estimated as part of PENSIM's pension characteristics imputation model (Holmer and Janney 2003). See Appendix C, Section C.14, beginning on page 200, for more details.

### B.17.3 Possibly Specify Pension Plan Characteristics

The probabilities of a sponsored pension having a particular set of plan characteristics are estimated as part of PENSIM's pension characteristics imputation model (Holmer and Janney 2003). See Appendix C, Section C.14, beginning on page 200, for more details.

### B.17.4 Possibly Schedule Pension-Eligibility Event(s)

If the individual's employer is a pension sponsor and the individual will eventually be eligible for a pension, then a pension-eligibility event is scheduled. Separate events are scheduled for the DB plan and the DC plan, if both are offered.

The waiting time between the start of the first job and a pension-eligibility event depends on the individual's age and the pension's eligibility rules. These rules involve a minimum age, a maximum waiting period, how many times per year newly eligible individuals are enrolled in the plan, and whether part-time workers are eligible. The characteristics of the pension plan are specified in the PENSIM input database. Given these pension eligibility parameters and the individual's age and part-time status on the job, the exact waiting time until the eligibility event can be calculated (assuming the individual will ever be eligible). This means that there is no stochastic element in the eligibility event's simulated waiting time.

### B.17.5 Schedule Earnings-Adjustment Event

The first-job-start event specifies the characteristics of the job, including its starting earnings level. This level is adjusted periodically by the activities of the earnings-adjustment event. In PENSIM, on-the-job earnings adjustments occur once a year on the anniversary of the individual's job start.

### B.17.6 Schedule First-Job-Finish Event

The first-job-start event specifies the characteristics of the new job and employer, including the employer's pension offering. It also determines the timing of the first-job-finish event.

The waiting time between the first-job-start and first-job-finish events is estimated using survival analysis techniques on data from the 1992 SIPP data. The estimated model is a proportional hazard model with a number of individual attributes and first-job characteristics as covariates ($x$), with a Weibull baseline hazard function (with shape parameter $\alpha$), and with multiplicative lognormal error terms to represent individual-specific and spell-specific unmeasured heterogeneity (Lancaster 1990). This model implies the following general form for the hazard function:

$$h(\alpha, \beta, \sigma_i^2, \sigma_s^2; x, t) = exp(x\beta + \epsilon_i + \epsilon_s)\alpha t^{\alpha-1}$$

where $t$ denotes the waiting time and each $\epsilon$ error term is normally distributed with mean equal to minus one-half of its variance ($\sigma^2$), which implies that $exp(\epsilon)$ has a lognormal distribution with mean one.

*Simulated Maximum Likelihood.* The use of up to two lognormal error terms (rather than one gamma error term) has been accomplished by using simulated (rather than analytic) maximum likelihood techniques (Stern 1997). This simulation-based, maximum-likelihood estimation method integrates out the effects on the log-likelihood function of the lognormal error terms that represent unmeasured heterogeneity across individuals and spells. As applied here, the method utilizes 100 random draws from the lognormal distributions with mean equal to one. Or, R = 100, using the terminology of Stern (1997, page 2006ff.). The variance of each lognormal distribution is estimated with sample data along with the other parameters of the model. Monte Carlo studies conducted as part of this task indicate that 100 draws is well in excess of the number needed to eliminate sampling inefficiency in the simulation-based estimation procedure. These Monte Carlo studies show that this simulated maximum-likelihood procedure is consistently able to estimate accurately the known parameters of a range of models similar to the ones used in PENSIM.

Several versions of this first-job duration model are estimated. All of these versions suppress the individual-specific error term because there is only one first-job spell for each individual. These versions of the basic model range from one in which no covariates are used to a version that uses all the covariates (i.e., the gender, age, and schooling attributes as well as the six job characteristics). In all versions, the estimation sample has been restricted to first-job spells that ended during the three-year SIPP survey period. The starting date of all of these spells is known because of the nature of the questions, which means that none of the first-job spells are left censored.

There are 2,909 first-job spells available for the no-covariates version of the model and the mean completed duration (because none of the spells are right censored) of those spells is 2.28 years. After eliminating spells for which there are incomplete data on individual attributes and job characteristics, the sample for the all-covariates version contains just 214 spells, the mean completed duration of which is 9.67 years. Most of the no-covariates sample is lost when employer size and pension sponsorship are added to the list of covariates. This is not surprising given that SIPP asks about employer size only in the initial employment and training history topical module and in the pension topical module, not every month like the other non-pension job characteristics. And employer pension sponsorship is queried only once during the three-year interview period, as part of the pension topical module.

*Two-Stage Estimation.* Clearly, the large no-covariate sample provides the most accurate measure of the typical duration of a first job, but suffers from an inability to determine how spell duration varies across different individuals and jobs. So, while the all-covariates sample permits the estimation of how employer size, pension sponsorship, and other covariates affect first-job duration, this subsample of spells is clearly biased toward longer than average first-job spells, which is to be expected given the difference in query frequency between the size and sponsorship variables, on the one hand, and the other variables, on the other hand.

How can this dilemma be resolved? The best that can be done with these SIPP data is to appeal to one of the key implications of the proportional hazard model that is being assumed here. One of the features of such a model is that its functional form implies that the parameters associated with the covariates can be estimated with just rank data on spell duration (Lancaster 1990, Chapter 9: Limited Information Inference). Rank data contain no information about the length of spells (which is called order data), only information about the relative length of spells (e.g., which spell was longest and which spell was shortest). These rank data, of course, do not permit estimation of constant term's parameter, which is part of the baseline hazard function.

Given this feature of the proportional hazard model, the estimation strategy pursued here involves two stages. In the first stage, the all-covariates version of the model is estimated with the small sample, biased toward longer spells. The parameters estimated in the first stage are then used in PENSIM to generate the mean completed duration of first jobs, which is well in excess of the 2.28 years indicated by the large, unbiased no-covariates sample. The second stage of the estimation involves calibrating the value of the parameter of the proportional hazard model's constant term that is used in PENSIM so that it produces a mean first-job duration statistic of 2.28 years.

A similar dilemma will be faced again in the estimation of the duration of subsequent-job spells and the duration of non-employment spells. These issues will be discussed in more detail below.

The estimated parameters of several versions of the proportional hazard model of first-job duration are presented in Appendix C, Section C.8, beginning on page 177. That section also includes the results of the second calibration stage of the hazard model estimation.

## B.18 First-Job-Finish Event

The first-job-finish event specifies the individual's age at which the first job ends, handles appropriate pension calculations, schedules appropriate pension events, and schedules the event at which the individual will start a subsequent job.

### B.18.1 Specify Job-Finish Age

At the first-job-finish event, the age of the individual is recorded in a list of jobs held by the individual.

### B.18.2 Specify Vested Defined Pension Benefits

At the first-job-finish event, any vested pension benefits in a defined-benefit plan are calculated and placed on a list of unaccepted pension claims.

### B.18.3 Possibly Schedule Pension-Contribution Event

If the individual is participating in a defined-contribution pension, schedule a pension-contribution event to occur immediately. This event, which normally occurs at the end of each calendar year, will handle partial year contributions because a job can end at any time during the year.

### B.18.4 Possibly Schedule Pension-Claim Event

If the individual has any unaccepted vested benefits from defined-benefit pension plans, a pension-claim event is scheduled to occur immediately.

### B.18.5 Schedule Subsequent-Job-Start Event

The waiting time between jobs consists of time during which the individual is either unemployed or out of the labor force. The duration of these non-employment spells is represented with a hazard function. The same hazard function is used to realize the waiting time between first-job-finish and subsequent-job-start events as is used below to generate the waiting time between subsequent-job-finish and subsequent-job-start events. This hazard function is estimated with a SIPP sample — a merger of data on the 1991 panel (wave 7) and the 1992 panel (wave 4) — of between-job spells that contains multiple spells for nearly a third of the individuals.

The hazard function is specified as a proportional hazard model with a number of individual attributes as covariates, with a Weibull baseline hazard function, and with two multiplicative lognormal error terms, as specified in the equation on page 117. The first error term represents individual-specific unmeasured heterogeneity and has the same value across all the spells experienced by the individual. The second error term represents spell-specific unmeasured heterogeneity and has a different value for each spell experienced by an individual. This error component specification is estimated using the same simulated maximum-likelihood techniques as described above (see discussion beginning on page 117).

Some of the estimated coefficients of this non-employment spell duration hazard function needed to be calibrated after disability events were added to PENSIM. Disability causes spells of non-employment, which were presumably included in the estimated coefficients. Adjustments to some of these coefficients are necessary in order to avoid double simulation of such non-employment spells.

The estimated and calibrated parameters of the non-employment spell duration model are presented in the Appendix C, Section C.9, beginning on page 180.

## B.19 Subsequent-Job-Start Event

Subsequent jobs are all those jobs held by an individual following the individual's first job. The subsequent-job-start event specifies several characteristics of that job and of the employer who offers the job, schedules the subsequent-job-finish event, the first earnings-adjustment event, and possibly one or more pension-eligibility events.

### B.19.1 Specify Job and Employer Characteristics

The subsequent job's characteristics are specified using the same kind of process used to specify first-job characteristics. This process involves sequential stochastic realization of six characteristic equations that together form a recursive model of job and employer characteristics. Each of the six characteristics is estimated using 1992 SIPP data — a merger of data on the 1991 panel (wave 7) and the 1992 panel (wave 4) — to be a function of individual attributes measured at the start of the subsequent job and other realized values of the that job's characteristics.

In order to induce the appropriate degree of correlation between the characteristics of the prior job and this subsequent job, we had planned to include prior job characteristics among the covariates of each of these six recursive equations. We could not do this because the available sample size was too small. We used an alternative estimation procedure involving two stages. First, we estimate the recursive equations without the covariates describing the characteristics of the prior job. And then in a second stage, the randomly generated variates used in PENSIM to realize each of the six job-characteristic equations have a correlation induced with the variate generated for the corresponding characteristic of the individual's prior job. The degree of induced correlation is calibrated so that PENSIM generates the desired degree of sample correlation between each prior-job and current-job characteristic. These target correlations are tabulated from longer-term longitudinal survey data, drawn from the Panel Study of Income Dynamics (PSID). This two-stage estimation strategy provides a way to combine information from SIPP and PSID in a manner that is usable in PENSIM simulations.

In addition, several constant terms are calibrated so that the prevalence rates for subsequent job characteristics are close to target rates derived from cross-sectional tabulations of SIPP data.

The calculation of the subsequent job's starting earnings is more involved than simply using the realized value of the earnings equation. The on-the-job earnings process represents annual earnings growth as the sum of two terms: a drift term (think of it as representing seniority pay increases) that all individuals experience, and a deviation term that varies across individuals (think of it as representing merit pay increases or bonuses). We will explain this on-the-job earnings adjustment process in more detail in Section B.21 later in this appendix.

The problem with using the realized value of the earnings equation to predict an individual's starting earnings on a new job is that its random element is in no way correlated with the individual's deviation term on the prior job. We need a way of calculating the individual's starting earnings that reflects earnings on the prior job as well as the expected effects of the individual's job change. An example of these expected job-change effects would be the drop in earnings associated with a move from a full-time job to a part-time job.

To accomplish this, we assume that when an individual changes jobs the starting value of log earnings, denoted by $LE$, is the sum of two terms: (1) the $LE$ value at the end of the prior job (which includes the cumulative effects of both the drift and deviation terms), and (2) a term that expresses the predicted change in $LE$ expected from the job change. This predicted change term is the drift-only value of $LE$ at the end of the prior job ($LE$ without any random shocks to the individual's earnings) minus the expected value of $LE$ at the start of the subsequent job (calculated from the earnings equation in the recursive job-characteristic equations). This process preserves much of the unexpectedly good or bad earnings experience the individual experienced on the prior job (represented by the cumulative random shocks in the deviation term), but adjusts that for systematic earnings effects expected from the change in jobs (as expressed in the estimated equation for starting earnings).

The structure of the six recursive equations, is therefore, identical to that described above on page 114 for first-job characteristics. The estimated and calibrated parameters of these six equations are presented in Appendix C, Section C.10, beginning on page 182.

### B.19.2 Possibly Specify Pension Plan Offering

If an employer is a pension sponsor, the employer can have one of three kinds of pension offerings: sponsoring only a defined-benefit (DB) pension plan, sponsoring only a defined-contribution (DC) pension plan, or sponsoring both a DB and DC pension plan.

The probabilities of a sponsoring employer having one of these three kinds of offerings, which are the same as those used for first jobs, are estimated as part of PENSIM's pension characteristics imputation model (Holmer and Janney 2003). See Appendix C, Section C.14, beginning on page 200, for more details.

### B.19.3 Possibly Specify Pension Plan Characteristics

The probabilities of a sponsored pension having a particular set of plan characteristics are estimated as part of PENSIM's pension characteristics imputation model (Holmer and Janney 2003). See Appendix C, Section C.14, beginning on page 200, for more details.

### B.19.4 Possibly Schedule Pension-Eligibility Event(s)

If the individual's employer is a pension sponsor and the individual will eventually be eligible for a pension, then a pension-eligibility event is scheduled. Separate events are scheduled for the DB plan and the DC plan, if both are offered.

The waiting time between the start of a subsequent job and a pension-eligibility event is calculated in the same way it is for first jobs (see Section B.17.4 in this appendix).

### B.19.5 Schedule Earnings-Adjustment Event

The subsequent-job-start event specifies the characteristics of the job, including its starting earnings level. This level is adjusted periodically by the activities of the earnings-adjustment event. In PENSIM, on-the-job earnings adjustments occur once a year on the anniversary of the individual's job start.

### B.19.6 Schedule Subsequent-Job-Finish Event

The subsequent-job-start event specifies the new job's and employer's characteristics, including the employer's pension offering. It also determines the duration of this subsequent job.

The waiting time between the subsequent-job-start and subsequent-job-finish events is estimated using survival analysis techniques on data from the 1992 SIPP data. The estimated model is a proportional hazard model with the same general structure as the equation on page 117. Because about seven percent of the subsequent-job spells are second spells for individuals, it is possible to estimate the error component version of the hazard function, in which there are separate error terms that represent individual-specific and spell-specific unmeasured heterogeneity in the sample. As described above beginning on page 117, simulated maximum-likelihood techniques are used to estimate the parameters of the hazard function.

It was not possible to include the characteristics of both the prior and subsequent job among the model's covariates because the number of spells without some missing data is extremely small. Even a less ambitious list of covariates, which includes the three individual attributes at the start of the subsequent job and the subsequent job's six characteristics, reduces the number of usable observations from 33,975 to 10,232 spells. The full sample of 33,975 spells consists of about 45 percent right-censored (or ongoing) spells. The mean incomplete duration of a subsequent-job spell in this full sample is 4.40 years. The mean incomplete duration of spells in the smaller subsample of spells with complete individual and job data is 9.96 years.

Clearly, the same estimation dilemma is present here as was faced in the estimation of the first-job duration model. The approach here is to use the same two-stage estimation approach described above on page 118. In the second calibration stage of estimating the duration of subsequent jobs, longer-term longitudinal PSID data are used to derive a target mean spell duration (that differs by age group) and a target correlation between prior-job duration and subsequent-job duration.

The estimated parameters of a proportional hazard model of subsequent-job duration, and their calibrated values, are presented in Appendix C, Section C.11, beginning on page 194.

## B.20 Subsequent-Job-Finish Event

The subsequent-job-finish event specifies the individual's age at which the subsequent job ends, handles appropriate pension calculations, schedules appropriate pension events, and schedules the event at which the individual will start the next subsequent job.

### B.20.1 Specify Job-Finish Age

At the subsequent-job-finish event, the age of the individual is recorded in a list of jobs held by the individual.

### B.20.2 Specify Vested Defined Pension Benefits

At the subsequent-job-finish event, any vested pension benefits in a defined-benefit plan are calculated and placed on a list of unaccepted pension claims.

### B.20.3 Possibly Schedule Pension-Contribution Event

If the individual is participating in a defined-contribution pension, schedule a pension-contribution event to occur immediately. This event, which normally occurs at the end of each calendar year, will handle partial year contributions because a job can end at any time during the year.

### B.20.4 Possibly Schedule Pension-Claim Event

If the individual has any unaccepted vested benefits from defined-benefit pension plans, a pension-claim event is scheduled to occur immediately.

### B.20.5 Schedule Subsequent-Job-Start Event

The waiting time between the end of a subsequent job and the start of the next job is generated using the same hazard function that is used to generate waiting times between the end of the first job and the start of the first subsequent job. The specification, estimation, and calibration of that hazard function is discussed in Appendix C, Section C.9, beginning on page 180.

## B.21 Earnings-Adjustment Event

The earnings-adjustment event simulates on-the-job pay increases and then schedules another earnings-adjustment event in another year. And every ten years after defined-contribution pension eligibility, a pension-participation event is scheduled if the individual is not participating in the plan.

### B.21.1 Specify Adjustment in Earnings

The size of an earnings adjustment is determined using a stochastic process that is often used in the earnings dynamics literature (Carroll 1992, for example) and is used by the Congressional Budget Office in its CBOLT model (Harris and Sabelhaus 2003). The basic approach employed in the PENSIM earnings-adjustment event is to use this standard method to simulate changes in no-growth earnings and then to scale up the adjusted level of no-growth earnings to nominal earnings using specified inflation and real wage growth assumptions. First, we explain in detail the concept of no-growth earnings and its relationship to nominal earnings, and then we describe in detail the stochastic process that generates changes in no-growth earnings.

*No-Growth Earnings and Nominal Earnings.* By no-growth earnings we mean earnings levels that are normalized for the effects of inflation and real wage growth. In an economy with zero rates of inflation and real wage growth, the only aggregate earnings adjustment from year to year would correspond to the age-related differences in a cross-sectional age-earnings profile. PENSIM works internally with no-growth earnings because the initial earnings level equations used in the first-job-start and subsequent-job-start events are estimated with cross-sectional data (as discussed above in Section B.17.1 and Section B.19.1, respectively). Our approach is to simulate changes in no-growth earnings using these earnings equations and the adjustment process described below, and then to convert no-growth earnings into nominal earnings. This conversion is accomplished by multiplying no-growth earnings by a factor that grows by the assumed aggregate inflation and real wage growth rates. Each individual's no-growth earnings are scaled in this manner because inflation and real wage growth are collective risks. The individual risks to earnings are discussed next.

*Adjustments in No-Growth Earnings.* Individuals experience widely varying no-growth earnings adjustments, with the variability being both across years for the same individual and across individuals at the same age. Following the earnings dynamics research literature, we represent annual on-the-job earnings adjustments using a stochastic process in the natural logarithm of no-growth earnings that has additive drift and deviation terms. All individuals experience the drift term at every earnings-adjustment event. The deviation term evolves stochastically, representing individual risks in earnings adjustments. The deviation term is the sum of two normally distributed variates: a permanent deviation, which is the same at each earnings-adjustment event throughout an individual's life, and a temporary deviation, which is different at each earnings-adjustment event. The resulting lognormal random variables, when they are used to adjust the logarithm of no-growth earnings, have one-half of their variance subtracted so that average earnings across individuals is not inflated by the undoing of the lognormal transformation.

The random earnings adjustment shocks — being a mixture of permanent and temporary shocks — cause individuals to experience widely differing earnings paths. In this way, PENSIM represents a key individual risk: that a person does not know at the start of life what kind of earnings path will play out in the future. Being able to represent this kind of risk is an essential feature of any simulation model that focuses on pension policy.

The value of the parameters used in the earnings adjustment process are presented in Appendix C, Section C.12, beginning on page 196.

### B.21.2 Schedule Another Earnings-Adjustment Event

In the current version of PENSIM, on-the-job earnings adjustments are assumed to occur annually on the anniversary of the job start. This means that each earnings-adjustment event concludes its activities by scheduling another earnings-adjustment event to occur a year later.

### B.21.3 Possibly Schedule Pension-Participation Event

At every tenth anniversary of becoming eligible for a defined-contribution plan, a non-participating individual is given another chance to participate in the plan. This is done by using the individual current earnings and age to generate a somewhat higher participation probability (because the assumed participation probability rises with age and relative earnings). If the individual is simulated to participate (using the same random number as was used in the initial participation decision), a waiting time until participation is generated and used to schedule the timing of a pension-participation event.

The nature of the defined-contribution participation probability function and the defined-contribution participation waiting-time function are presented in Appendix C, Section C.15, beginning on page 202.

## B.22 Pension-Eligibility Event

The pension-eligibility event — one each for defined-benefit pensions and for defined-contribution pensions — specifies the age at which the individual becomes eligible for a pension. It also calculates the waiting time until the individual will become vested in the pension and schedules a pension-vesting event at the current age plus that waiting time.

### B.22.1 Specify Pension Eligibility Age

If the individual is still on the same job for which the pension-eligibility event was scheduled, the individual's age is recorded as the eligibility age for this pension. This eligibility age is used in calculating the number of years for which the individual is eligible for this pension.

### B.22.2 Possibly Schedule Pension-Participation Event

If the pension-eligibility event is for a defined-benefit plan, the individual is assumed to participate immediately in the plan. (And, in the case of a defined-benefit cash-account plan, a pension-contribution event is scheduled at the end of the current calendar year.)

If the pension-eligibility event is for a defined-contribution plan that has a combination of no minimum employee contribution requirement and some nonmatching employer contribution (usually a profit-sharing or money-purchase plan), then full participation is assumed. For all other defined-contribution plans (mostly savings-thrift plans), an participation probability is used to determine participation in the plan. For an individual simulated to participate, a waiting time until participation is generated and used to schedule the timing of a pension-participation event.

The nature of the defined-contribution participation probability function and the defined-contribution participation waiting-time function are presented in Appendix C, Section C.15, beginning on page 202.

### B.22.3 Schedule Pension-Vesting Event

The waiting time until the next pension-vesting event is calculated given two parameters of the pension: the whole number of years to full vesting, and whether the plan has cliff or gradual vesting. In the case of zero vesting years, the pension-vesting event is scheduled to occur immediately. When the pension plan's number of vesting years is positive, a pension-vesting event is scheduled in the future. Under cliff vesting, the full-vesting event is scheduled to occur vesting years in the future, when the individual will be completely vested in the pension plan. If the pension plan calls for gradual vesting, the first pension-vesting event is scheduled to occur one year in the future, when the individual will become partially vested in the pension plan. The partial-vesting fraction is simply the inverse of the number of vesting years.

The waiting time to the next pension-vesting event contains, therefore, no stochastic elements.

## B.23 Pension-Participation Event

At the pension-participation event, a plan account is established for the individual and both a pension-contribution and an active-participation event are scheduled.

### B.23.1 Specify Plan Account

At the pension-participation event, an empty pension plan account is established for the individual.

### B.23.2 Schedule Pension-Contribution Event

At the pension-participation event, the first pension-contribution event is scheduled for the end of the current calendar year.

### B.23.3 Schedule Active-Pension-Participation Event

At the pension-participation event, the waiting time until the active-pension-participation event is generated and used to schedule an active-pension-participation event.

The nature of the active-pension-participation waiting-time function is presented in Appendix C, Section C.15, beginning on page 202.

## B.24 Active-Pension-Participation Event

At the active-pension-participation event, the individual is recorded as an active (that is, decision-making) participant in the defined-contribution plan. Defined-contribution plan participants are defined as being passive participants before the occurrence of this event. The contribution and asset-allocation behavior of passive participants are defined by the plan's default contribution rate and asset-allocation rules.

The active-pension-participation event schedules no other events.

## B.25 Pension-Contribution Event

The pension-contribution event specifies the realized returns on prior year investments, specifies the employee and employer contribution amounts to the plan account, and allocates plan assets between several investment options. And then the event schedules another pension-contribution event a year later (if the individual is not at the end of this job) or schedules a pension-rollover event to occur immediately (if the individual is at the end of this job).

### B.25.1 Specify Pension Contribution Amounts

Before simulating contributions to the plan account, the investment returns earned are calculated using the prior year's asset allocation and the simulated asset returns passed to PENSIM from SSASIM in the environment (env) file. Immediately after adjusting the account holdings to reflect realized returns, the contributions are simulated.

The employee contribution amount (as a percent of earnings) is specified using a pension contribution stochastic process. Given the employee contribution, the amount of the employer contribution (if any) is calculated using the plan characteristics. The contributions are limited by whatever federal regulations are effective for that year.

The nature of the employee pension-contribution stochastic process is presented in Appendix C, Section C.16, beginning on page 206.

And finally, immediately after the post-contribution account balance is determined, the total account balance is allocated among the investment options available in the plan account. This asset allocation process includes possible restrictions caused by employer contributions of company stock.

The manner in which the employee's pension-investment asset-allocation shares are calculated is discussed in Appendix C, Section C.17, beginning on page 209.

### B.25.2 Possibly Schedule Another Pension-Contribution Event

If the pension-contribution event does not occur at the end of the job, another pension-contribution event is scheduled exactly one year later.

### B.25.3 Possibly Schedule Pension-Rollover Event

If the pension-contribution event does occur at the end of the job, a pension-rollover event is scheduled to occur immediately.

## B.26 Cash-Account-Conversion Event

The cash-account-conversion event handles a plan conversion in which an employer terminates a terminal-earnings plan (often called a final-average-pay plan) and replaces it with a cash-account plan (often called a cash-balance plan). The event stores the old terminal-earnings plan characteristics (so that they are available for any grandfathering calculations at the end of the job) and specifies the ongoing characteristics of the new cash-account plan and well as the conversion provisions. The conversion provisions and the characteristics of the terminal-earnings plan are used to compute the opening balance. And finally, the cash-account-conversion event schedules the first pension-contribution event at the end of the current calendar year.

## B.27 Pension-Vesting Event

The pension-vesting event — one each for defined-benefit pensions and for defined-contribution pensions — specifies the individual's vested fraction for that pension, and schedules another pension-vesting event one year in the future whenever the individual's vested fraction is less than one.

### B.27.1 Specify Accrual of Pension Vesting Rights

If the individual is still on the same job for which the pension-vesting event was scheduled, the individual's vested fraction is increased by an amount that depends on the pension plan's vesting-years and vesting-scheme parameters. Under the cliff-vesting scheme or under the gradual-vesting scheme when vesting years is one or zero, the occurrence of the pension-vesting event results in complete vesting. In all other cases, the event results in a vested fraction of less than one. So, for example, when vesting years equals five under gradual vesting, the first pension-vesting event causes the individual's vested fraction to rise from zero to two-tenths.

### B.27.2 Possibly Schedule Another Pension-Vesting Event

Whenever the pension-vesting event leaves the individual's vested fraction below one, another pension-vesting event is scheduled a year in the future.

## B.28 Pension-Rollover Event

The pension-rollover event determines the vested account balance in the pension plan account, whether or not that balance is transfered to the individual's personal rollover account, and, if it is transferred, updates the rollover account and schedules a rollover-account-management event.

### B.28.1 Specify Vested Pension Account Balance

The fraction of the total account balance that is vested is calculated using the plan rules and the job tenure of the individual. If the vested balance is zero, nothing else is done in the pension-rollover event.

### B.28.2 Specify Rollover Decision

The probability of transferring the vested balance to the individual's person rollover account is calculated using a rollover probability function. If the individual is not simulated to rollover the vested balance, nothing else is done in the pension-rollover event (that is, the vested balance disappears from the PENSIM simulation because the individual consumes the vested balance rather than saving it for retirement in the rollover account).

The pension-rollover probability function is described in Appendix C, Section C.18, beginning on page 210.

### B.28.3 Possibly Schedule Rollover-Account-Management Event

If the individual decides to rollover the vested balance, the rollover account accrues partial-year investment returns, the vested balance is contributed, and the total rollover-account balance is allocated among the available investment options. The asset-allocation shares are determined are calculated in a manner discussed in Appendix C, Section C.17, beginning on page 209.

And finally, when the individual elects to rollover the vested balance, a rollover-account-management event is scheduled. That event is scheduled to occur either at the end of the current calendar year or immediately. The immediate rollover-account-management event occurs only if the individual has already started to make withdrawals from the rollover account. Such withdrawals could be made either to buy a retirement annuity or to support retirement consumption directly.

## B.29 Rollover-Account-Management Event

The rollover-account-management involves crediting the account with returns on assets held since the prior rollover-account-management event, allocating the total balance across the investment options available in the rollover account, scheduling an annuity-purchase event or a pension-payment event or both, if the individual is withdrawing from the account, and finally possibly schedule another rollover-account-management event.

### B.29.1 Specify Investment Returns and Asset Allocation

Investment returns earned since the last event are calculated using the prior event's asset allocation, the fraction of the year since the prior event, and the simulated asset returns passed to PENSIM from SSASIM in the environment (env) file.

Immediately after adjusting the account holdings to reflect realized returns, the account balance is allocated among a number of investment option using asset-allocation shares calculated for the individual. The manner in which the employee's pension-investment asset-allocation shares are calculated is discussed in Appendix C, Section C.17, beginning on page 209.

### B.29.2 Specify Account Withdrawal Timing

The age at which an individual is assumed to begin making rollover account withdrawals is simulated by following a series of rules. These rules are described in Appendix C, Section C.19, beginning on page 211.

### B.29.3 Possibly Schedule Annuity-Purchase Event

If the individual is assumed to have started making withdrawals from the rollover account and if the assumed fraction of withdrawals devoted to annuity purchases is greater than zero, an annuity-purchase event is scheduled to occur immediately.

### B.29.4 Possibly Schedule Pension-Payment Event

If the individual is assumed to have started making withdrawals from the rollover account and if the assumed fraction of withdrawals devoted to annuity purchases is less than one, a pension-payment event is scheduled to occur immediately (but a moment after the annuity-purchase event).

### B.29.5 Possibly Schedule Rollover-Account-Management Event

If the rollover account balance was positive at the start of this event, another rollover-account-management event is scheduled at the end of the current calendar year.

## B.30 Retirement Event

The retirement event involves withdrawing from employment by either finishing the current job or eliminating any pending job-related events that have been previously scheduled. The retirement event schedules a variety of pension-related events, which together move the individual from the pension accumulation phase of life into the pension withdrawal phase of life. In the current version of PENSIM, it is assumed that the individual holds no jobs after retiring.

The algorithm used to determine the timing of the retirement event is described in Appendix C, Section C.20, beginning on page 213.

## B.31 Annuity-Purchase Event

The annuity-purchase event involves determining the individual's dollar expenditure on an annuity, determining whether to buy a single-life or joint-and-survivor annuity, being quoted an annuity price (plus loading) by the annuity provider, and paying for the annuity by withdrawing the dollar expenditure from the rollover account, adding the purchased annuity to a list of annuity contracts owned by the individual, and possibly scheduling a pension-payment event.

### B.31.1 Specify Annuity Contract Characteristics

The fraction of the rollover-account withdrawal that is assumed to be devoted to purchasing an annuity and the total rollover account balance determine the dollar expenditure on the annuity contract. A married individual is assumed to purchase a joint-and-survivor annuity and a single individual is assumed to purchase a single-life annuity.

PENSIM has an annuity price function that uses standard actuarial methods to combine the ages and genders of the individual and spouse, as well as the individual's cohort birth year, with the age- and gender-specific mortality rates and mortality decline rates passed to PENSIM from SSASIM in the environment (env) file, to produce the unloaded price of a one-dollar-per-year annuity. An assumed gender-specific annuity loading factor is combined with this price to produce a total (that is, loaded) annuity price. The calculated annuity expenditures divided by this annuity price produces the quantity of one-dollar-per-year annuity purchased by the individual.

The purchased annuity contract is paid for out of the rollover account and added to a list of annuity contracts owned by the individual.

### B.31.2 Possibly Schedule Pension-Payment Event

If a pension-payment event has not yet been scheduled for this year, such an event is scheduled.

## B.32 Pension-Claim Event

The pension-claim event involves examining each of the unaccepted vested defined-benefit pension benefits earned on prior jobs to determine which benefits to accept (that is, claim).

### B.32.1 Specify Pension Benefit as Claimed

The individual claims vested benefits from plans that have a normal retirement age less than or equal to the individual's current age. Otherwise, the individual claims vested benefits from plans that have an early retirement age less than or equal to the individual's current age only if the plan does not impose benefit reductions for early retirement. If early retirement reductions are imposed, the individual is assumed to claim the vested benefit only if the individual is currently retired, disabled, or widowed. The individual is assumed not to claim vested benefits from plans that have an early retirement age greater than the individual's current age.

### B.32.2 Possibly Schedule Pension-Payment Event

If any vested defined benefits are claimed, the benefits are recorded as being claimed and a pension-payment event is scheduled to occur at the next earliest start-of-the-year pension-payment moment. If any vested defined benefits are not claimed, pension-claim events are scheduled in the future at the vested benefit's early or normal retirement age.

## B.33 Pension-Payment Event

The pension-payment event involves receiving claimed defined benefits, annuity contract payments, and annual withdrawals from the rollover account. These three types of payments are combined into the individual's employer-sponsored pension income, which is received at the start of each calendar year. And then another pension-payment event is scheduled for the start of the next year.

### B.33.1 Specify Annual Pension Income

An individual maintains a list of accepted vested defined benefits, a list of purchased annuity contracts, and a rollover account. The activity of the pension-payment event involves searching through the lists to determined defined-benefit pension income and annuity-contract pension income for the current year. In addition, if the rollover account balance is positive, is fraction of that balance is withdrawn to support retirement income needs. The withdrawal fraction is determined by a function that describes annual rollover-account withdrawal behavior. The details of that function are described in Appendix C, Section C.19, beginning on page 211.

### B.33.2 Schedule Another Pension-Payment Event

Another pension-payment event is scheduled for the start of the next year.

## B.34 Cross-Sectional-Survey Event

The cross-sectional-survey event involves recording personal attributes as well as current job and pension characteristics for the sampled individual. This information is similar to that obtained in the CPS or SIPP surveys when supplementary pension questionnaires are administered. The recorded information is written to an output file that can be used, after the PENSIM run has completed execution, to calculate any desired summary statistic or cross tabulation.

Because only one event of this kind is scheduled per individual, the cross-sectional sample contains at most the number of individuals specified to be born in the simulated birth cohort. The survey event is scheduled at birth to occur at a random age within a specified age range (for example, 14–64). This means that some members of the birth cohort will die before their cross-sectional-survey event is scheduled to occur, causing the survey sample to be somewhat smaller than the number of cohort births.

The sampling of individuals in the same birth cohort at different ages produces a sample that is similar in concept to a single-date, cross-sectional survey of a population that is composed of many birth cohorts. This statement is true only if we have assumed no inflation and no real wage growth over the course of individual's lives.

## B.35 Death Event

The death event specifies the individual's age at death, removes all other previously scheduled events from the pending event list, and triggers the collection of statistics that describe the individual's simulated lifetime experience.

A potential death event occurs at the end of each calendar year. The potential death event computes an annual mortality rate for the individual that depends on gender, age, birth cohort, education, and disability status. The general trend in gender- and age-specific mortality rates is passed to PENSIM from SSASIM via the environment (env) file. If a uniformly distributed random number is less than the scaled mortality rate for the individual, then the potential death event turns into an actual death event. Otherwise, another potential death event is scheduled for the end of the next year.

# C Appendix: PENSIM Input Parameters

This appendix contains a listing of the behavioral parameters that are being used as PENSIM input parameters. Most of the behavioral parameters have been estimated and a few have been calibrated so that results better fit statistics tabulated from data sets not used in their estimation. This appendix is best considered a technical supplement to Appendix B on behavioral events.

In this appendix, each section starts on a new page and sometimes tables containing estimation results will begin at the top of a page in order to avoid breaking the table across two pages.

## C.1 Mortality Experience

PENSIM uses gender-specific and age-specific period mortality rate tables for 1935, 1955, 1970, 1985, 1992, and 1997, drawn from Actuarial Study #107 by the SSA Office of the Chief Actuary. For any simulated cohort born before 1997, the starting mortality tables are interpolated from these historical tables. For lifetimes that extend beyond 1997, and for a cohort born after 1997, mortality tables are projected from the 1997 tables using mortality decline rates provided to PENSIM by SSASIM.

These all-population mortality rate are multiplicatively scaled for an individuals gender, education, and disability status.

The educational adjustment factors vary by birth cohort. For women (men) at all ages, high school dropouts have mortality rates that are 1.50 (1.30–1.40) times the overall rates for their gender, high school graduates have rates that are 0.90–1.00 (0.80–0.90) times their overall rates, college attendees have rates that are 0.90–0.96 (0.80–0.90) times their overall rates, and four-year college graduates (including those with graduate degrees) have mortality rates that are 0.70 (0.70) times the overall rates for their age and gender. These educational ratios have been specified to produce the variation in average life expectancy at age sixty reported in the MINT project (Panis and Lillard 1999, page 21).

When an individual is simulated to experience disability onset, we assume that individual's mortality rate is two times what it would be otherwise. The value of two is derived from an analysis of Disability Insurance (Holmer 2001) using overall mortality rates for the whole population and mortality rates among Disability Insurance beneficiaries (Zayatz 1999, Table 7). This factor of two compares with values of 2.94 for women and 2.45 for men reported in the MINT project (Panis and Lillard 1999, page 46). Note that this disability adjustment implies the need to scale down all mortality rates so that the average rate is close to the whole-population rates in the original tables. The all-gender adjustment factor used range from 0.95–1.00 (0.89–1.00) for women (men).

Note that PENSIM assumes deaths for a certain age occur at the end of that year of age, just before the individual has a birthday and becomes a year older. So, for example, those dying at age 72 all die moments before their 73rd birthday.

## C.2 Schooling Experience

The following four tables contain gender-specific non-parametric hazard functions for the timing of various schooling events, the decimal (not percentage) fraction of those finishing high school at each age who get a diploma, and the fraction of those finishing college each year after starting who get a four-year degree.

These four tables appear on the following four pages.

The first table contains the high-school-finish hazard functions and graduation probabilities. Age 41 hazard rates represent censoring.

***High School Finish Age and Outcome***

| Age | Female Hazard Rate | Female Graduation Rate | Male Hazard Rate | Male Graduation Rate |
|---|---|---|---|---|
| 14 | 0.0274 | 0.0586 | 0.0319 | 0.0625 |
| 15 | 0.0217 | 0.2639 | 0.0183 | 0.2430 |
| 16 | 0.0434 | 0.4320 | 0.0407 | 0.3963 |
| 17 | 0.1829 | 0.8558 | 0.1638 | 0.8256 |
| 18 | 0.7734 | 0.9765 | 0.7030 | 0.9705 |
| 19 | 0.6537 | 0.9567 | 0.6922 | 0.9585 |
| 20 | 0.3569 | 0.9381 | 0.4498 | 0.9233 |
| 21 | 0.2169 | 0.9015 | 0.3083 | 0.9065 |
| 22 | 0.1710 | 0.9389 | 0.1884 | 0.9046 |
| 23 | 0.1287 | 0.8980 | 0.2612 | 0.9222 |
| 24 | 0.1144 | 0.9024 | 0.1549 | 0.9192 |
| 25 | 0.1476 | 0.8955 | 0.1889 | 0.8269 |
| 26 | 0.1102 | 0.9001 | 0.1245 | 0.9481 |
| 27 | 0.1531 | 0.9374 | 0.1287 | 0.8549 |
| 28 | 0.1807 | 0.9096 | 0.1479 | 0.9175 |
| 29 | 0.1227 | 0.8361 | 0.1192 | 0.8623 |
| 30 | 0.1874 | 0.9444 | 0.2255 | 0.9258 |
| 31 | 0.1531 | 0.8961 | 0.1983 | 0.8321 |
| 32 | 0.1906 | 0.9437 | 0.2457 | 0.8841 |
| 33 | 0.2704 | 0.8964 | 0.2416 | 1.0000 |
| 34 | 0.1999 | 0.9165 | 0.2187 | 1.0000 |
| 35 | 0.1984 | 0.8638 | 0.2962 | 0.7978 |
| 36 | 0.2524 | 0.4989 | 0.3470 | 1.0000 |
| 37 | 0.3052 | 0.5248 | 0.2115 | 1.0000 |
| 38 | 0.2754 | 1.0000 | 0.1548 | 1.0000 |
| 39 | 0.3122 | 1.0000 | 0.5569 | 1.0000 |
| 40 | 0.5774 | 1.0000 | 0.0000 | 1.0000 |
| 41 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |

The second table contains the hazard functions for the waiting time (measured in years) between high school graduation and the start of college. Year 27 hazard rates represent censoring.

| ***College Start Waiting Time*** | | |
|---|---|---|
| | Female | Male |
| | Hazard | Hazard |
| Years | Rate | Rate |
| 0 | 0.3577 | 0.3427 |
| 1 | 0.2439 | 0.2445 |
| 2 | 0.0602 | 0.0622 |
| 3 | 0.0289 | 0.0333 |
| 4 | 0.0184 | 0.0297 |
| 5 | 0.0183 | 0.0232 |
| 6 | 0.0122 | 0.0151 |
| 7 | 0.0141 | 0.0171 |
| 8 | 0.0133 | 0.0121 |
| 9 | 0.0135 | 0.0069 |
| 10 | 0.0153 | 0.0079 |
| 11 | 0.0108 | 0.0060 |
| 12 | 0.0071 | 0.0077 |
| 13 | 0.0076 | 0.0077 |
| 14 | 0.0110 | 0.0039 |
| 15 | 0.0080 | 0.0062 |
| 16 | 0.0101 | 0.0005 |
| 17 | 0.0118 | 0.0034 |
| 18 | 0.0112 | 0.0025 |
| 19 | 0.0105 | 0.0040 |
| 20 | 0.0080 | 0.0000 |
| 21 | 0.0090 | 0.0030 |
| 22 | 0.0063 | 0.0094 |
| 23 | 0.0000 | 0.0000 |
| 24 | 0.0000 | 0.0000 |
| 25 | 0.0000 | 0.0000 |
| 26 | 0.0000 | 0.0000 |
| 27 | 1.0000 | 1.0000 |

The third table contains the hazard functions for the waiting time (measured in years) between college start and finish as well as the probability of finishing college with a four-year degree. Year 27 hazard rates represent censoring.

| ***College Finish Waiting Time and Outcome*** | | | | |
|---|---|---|---|---|
| Years | Female Hazard Rate | Female 4-yr Degree Rate | Male Hazard Rate | Male 4-yr Degree Rate |
| 0 | 0.0530 | 0.0273 | 0.0519 | 0.0342 |
| 1 | 0.1401 | 0.0086 | 0.1218 | 0.0129 |
| 2 | 0.1671 | 0.0580 | 0.1569 | 0.0657 |
| 3 | 0.2100 | 0.5335 | 0.1736 | 0.5272 |
| 4 | 0.4641 | 0.8389 | 0.4300 | 0.8342 |
| 5 | 0.3330 | 0.7907 | 0.3469 | 0.7962 |
| 6 | 0.2025 | 0.6746 | 0.2616 | 0.7387 |
| 7 | 0.1568 | 0.5815 | 0.2078 | 0.6265 |
| 8 | 0.1654 | 0.5451 | 0.1854 | 0.6461 |
| 9 | 0.1483 | 0.3897 | 0.1918 | 0.5196 |
| 10 | 0.1197 | 0.4224 | 0.1854 | 0.6834 |
| 11 | 0.1326 | 0.3297 | 0.1319 | 0.3157 |
| 12 | 0.1317 | 0.3480 | 0.1722 | 0.4408 |
| 13 | 0.1655 | 0.2020 | 0.1850 | 0.5304 |
| 14 | 0.1663 | 0.4540 | 0.1711 | 0.3266 |
| 15 | 0.1809 | 0.3701 | 0.2278 | 0.2906 |
| 16 | 0.2138 | 0.1982 | 0.2169 | 0.2984 |
| 17 | 0.1838 | 0.2521 | 0.2291 | 0.2252 |
| 18 | 0.1925 | 0.3547 | 0.1661 | 0.0000 |
| 19 | 0.1827 | 0.1746 | 0.2742 | 0.3106 |
| 20 | 0.2932 | 0.2315 | 0.2890 | 0.4782 |
| 21 | 0.4332 | 0.1094 | 0.2176 | 0.4439 |
| 22 | 0.3691 | 0.0000 | 0.5630 | 0.6098 |
| 23 | 1.0000 | 0.5544 | 0.0000 | 0.0000 |
| 24 | 1.0000 | 0.0000 | 0.0000 | 0.0000 |
| 25 | 1.0000 | 0.0000 | 0.0000 | 0.0000 |
| 26 | 1.0000 | 0.0000 | 0.0000 | 0.0000 |
| 27 | 1.0000 | 0.0000 | 1.0000 | 0.0000 |

The fourth table contains the hazard functions for the waiting time (measured in years) between college graduation (with a four-year degree) and the completion of a graduate degree. Year 27 hazard rates represent censoring.

| ***Graduate Degree Waiting Time*** | | |
|---|---|---|
| Years | Female Hazard Rate | Male Hazard Rate |
| 0 | 0.0030 | 0.0042 |
| 1 | 0.0163 | 0.0202 |
| 2 | 0.0507 | 0.0486 |
| 3 | 0.0421 | 0.0521 |
| 4 | 0.0478 | 0.0664 |
| 5 | 0.0403 | 0.0498 |
| 6 | 0.0334 | 0.0360 |
| 7 | 0.0258 | 0.0343 |
| 8 | 0.0246 | 0.0217 |
| 9 | 0.0175 | 0.0194 |
| 10 | 0.0158 | 0.0281 |
| 11 | 0.0108 | 0.0155 |
| 12 | 0.0135 | 0.0099 |
| 13 | 0.0185 | 0.0158 |
| 14 | 0.0157 | 0.0204 |
| 15 | 0.0071 | 0.0152 |
| 16 | 0.0121 | 0.0215 |
| 17 | 0.0097 | 0.0067 |
| 18 | 0.0000 | 0.0000 |
| 19 | 0.0744 | 0.0000 |
| 20 | 0.0000 | 0.0000 |
| 21 | 0.0000 | 0.0000 |
| 22 | 0.0000 | 0.0000 |
| 23 | 0.0000 | 0.0000 |
| 24 | 0.0000 | 0.0000 |
| 25 | 0.0000 | 0.0000 |
| 26 | 0.0000 | 0.0000 |
| 27 | 1.0000 | 1.0000 |

## C.3 Marital Experience

This section begins by citing the source of the wedding and divorce hazard functions that are used in PENSIM. The tables in this section show the PENSIM spouse-attributes probability distributions that are estimated with PSID data.

The waiting time between age 12 or the age of a divorce event until the age of the next wedding event is simulated in PENSIM using the gender-specific continuous-time log-linear hazard-function models estimated by RAND for the SSA MINT project (Panis and Lillard 1999, pages 22–27).

The waiting time between the wedding event and the divorce event is simulated using the gender-specific continuous-time log-linear hazard-function models estimated by RAND for the SSA MINT project (Panis and Lillard 1999, pages 28–30).

The attributes of the spouse — namely, the age difference and educational attainment of the spouse — are simulated in PENSIM using a set of probabilities that are estimated using thirty years of PSID data on the attributes of individuals getting married. There is a probability distribution for the age difference between the spouse and the cohort sample individual and their is a probability distribution for the educational level of the spouse.

There is a different probability distribution for each spouse attribute depending on the gender of the sample individual who is getting married, depending on the attributes of that sample individual (age of the individual for the age difference probabilities and education of the individual for the educational level probabilities), and depending on the birth year of the sample individual.

The spouse age-difference distributions are presented first followed by the spouse educational-level distributions.

***Spouse Age-Difference Distributions by Individual Birth Year, Gender, and Age Group***

| Birth Year | Gender | Age Group | Age Difference | Probability |
|---|---|---|---|---|
| <=1951 | Male | <=19 | -9 | 0 |
| <=1951 | Male | <=19 | -8 | 0 |
| <=1951 | Male | <=19 | -7 | 0 |
| <=1951 | Male | <=19 | -6 | 0 |
| <=1951 | Male | <=19 | -5 | 0.008 |
| <=1951 | Male | <=19 | -4 | 0.049 |

*table continued on next page*

***Spouse Age-Difference Distributions by Individual Birth Year, Gender, and Age Group***
*(continued)*

| Birth Year | Gender | Age Group | Age Difference | Probability |
|---|---|---|---|---|
| <=1951 | Male | <=19 | -3 | 0.086 |
| <=1951 | Male | <=19 | -2 | 0.204 |
| <=1951 | Male | <=19 | -1 | 0.228 |
| <=1951 | Male | <=19 | 0 | 0.202 |
| <=1951 | Male | <=19 | 1 | 0.120 |
| <=1951 | Male | <=19 | 2 | 0.017 |
| <=1951 | Male | <=19 | 3 | 0.035 |
| <=1951 | Male | <=19 | 4 | 0.017 |
| <=1951 | Male | <=19 | 5 | 0.010 |
| <=1951 | Male | <=19 | 6 | 0.010 |
| <=1951 | Male | <=19 | 7 | 0 |
| <=1951 | Male | <=19 | 8 | 0.001 |
| <=1951 | Male | <=19 | 9 | 0.013 |
| <=1951 | Male | 20–29 | -9 | 0.025 |
| <=1951 | Male | 20–29 | -8 | 0.021 |
| <=1951 | Male | 20–29 | -7 | 0.031 |
| <=1951 | Male | 20–29 | -6 | 0.066 |
| <=1951 | Male | 20–29 | -5 | 0.090 |
| <=1951 | Male | 20–29 | -4 | 0.119 |
| <=1951 | Male | 20–29 | -3 | 0.149 |
| <=1951 | Male | 20–29 | -2 | 0.160 |
| <=1951 | Male | 20–29 | -1 | 0.135 |
| <=1951 | Male | 20–29 | 0 | 0.101 |
| <=1951 | Male | 20–29 | 1 | 0.043 |
| <=1951 | Male | 20–29 | 2 | 0.022 |
| <=1951 | Male | 20–29 | 3 | 0.014 |
| <=1951 | Male | 20–29 | 4 | 0.008 |
| <=1951 | Male | 20–29 | 5 | 0.010 |
| <=1951 | Male | 20–29 | 6 | 0.004 |
| <=1951 | Male | 20–29 | 7 | 0.002 |
| <=1951 | Male | 20–29 | 8 | 0 |
| <=1951 | Male | 20–29 | 9 | 0 |
| <=1951 | Male | >=30 | -9 | 0.355 |

*table continued on next page*

***Spouse Age-Difference Distributions by Individual Birth Year, Gender, and Age Group***

*(continued)*

| Birth Year | Gender | Age Group | Age Difference | Probability |
|---|---|---|---|---|
| <=1951 | Male | >=30 | -8 | 0.060 |
| <=1951 | Male | >=30 | -7 | 0.057 |
| <=1951 | Male | >=30 | -6 | 0.083 |
| <=1951 | Male | >=30 | -5 | 0.045 |
| <=1951 | Male | >=30 | -4 | 0.048 |
| <=1951 | Male | >=30 | -3 | 0.057 |
| <=1951 | Male | >=30 | -2 | 0.051 |
| <=1951 | Male | >=30 | -1 | 0.053 |
| <=1951 | Male | >=30 | 0 | 0.033 |
| <=1951 | Male | >=30 | 1 | 0.042 |
| <=1951 | Male | >=30 | 2 | 0.028 |
| <=1951 | Male | >=30 | 3 | 0.027 |
| <=1951 | Male | >=30 | 4 | 0.015 |
| <=1951 | Male | >=30 | 5 | 0.017 |
| <=1951 | Male | >=30 | 6 | 0.008 |
| <=1951 | Male | >=30 | 7 | 0.005 |
| <=1951 | Male | >=30 | 8 | 0.004 |
| <=1951 | Male | >=30 | 9 | 0.012 |
| | | | | |
| <=1951 | Female | <=19 | -9 | 0 |
| <=1951 | Female | <=19 | -8 | 0 |
| <=1951 | Female | <=19 | -7 | 0 |
| <=1951 | Female | <=19 | -6 | 0 |
| <=1951 | Female | <=19 | -5 | 0 |
| <=1951 | Female | <=19 | -4 | 0 |
| <=1951 | Female | <=19 | -3 | 0.002 |
| <=1951 | Female | <=19 | -2 | 0.010 |
| <=1951 | Female | <=19 | -1 | 0.027 |
| <=1951 | Female | <=19 | 0 | 0.075 |
| <=1951 | Female | <=19 | 1 | 0.136 |
| <=1951 | Female | <=19 | 2 | 0.178 |
| <=1951 | Female | <=19 | 3 | 0.160 |
| <=1951 | Female | <=19 | 4 | 0.138 |

*table continued on next page*

**Spouse Age-Difference Distributions by Individual Birth Year, Gender, and Age Group**
*(continued)*

| Birth Year | Gender | Age Group | Age Difference | Probability |
|---|---|---|---|---|
| <=1951 | Female | <=19 | 5 | 0.089 |
| <=1951 | Female | <=19 | 6 | 0.061 |
| <=1951 | Female | <=19 | 7 | 0.043 |
| <=1951 | Female | <=19 | 8 | 0.033 |
| <=1951 | Female | <=19 | 9 | 0.048 |
| <=1951 | Female | 20–29 | -9 | 0 |
| <=1951 | Female | 20–29 | -8 | 0 |
| <=1951 | Female | 20–29 | -7 | 0 |
| <=1951 | Female | 20–29 | -6 | 0.006 |
| <=1951 | Female | 20–29 | -5 | 0.010 |
| <=1951 | Female | 20–29 | -4 | 0.017 |
| <=1951 | Female | 20–29 | -3 | 0.030 |
| <=1951 | Female | 20–29 | -2 | 0.039 |
| <=1951 | Female | 20–29 | -1 | 0.080 |
| <=1951 | Female | 20–29 | 0 | 0.132 |
| <=1951 | Female | 20–29 | 1 | 0.143 |
| <=1951 | Female | 20–29 | 2 | 0.136 |
| <=1951 | Female | 20–29 | 3 | 0.102 |
| <=1951 | Female | 20–29 | 4 | 0.081 |
| <=1951 | Female | 20–29 | 5 | 0.063 |
| <=1951 | Female | 20–29 | 6 | 0.041 |
| <=1951 | Female | 20–29 | 7 | 0.022 |
| <=1951 | Female | 20–29 | 8 | 0.014 |
| <=1951 | Female | 20–29 | 9 | 0.084 |
| <=1951 | Female | >=30 | -9 | 0.051 |
| <=1951 | Female | >=30 | -8 | 0.015 |
| <=1951 | Female | >=30 | -7 | 0.037 |
| <=1951 | Female | >=30 | -6 | 0.031 |
| <=1951 | Female | >=30 | -5 | 0.054 |
| <=1951 | Female | >=30 | -4 | 0.046 |
| <=1951 | Female | >=30 | -3 | 0.046 |
| <=1951 | Female | >=30 | -2 | 0.065 |
| <=1951 | Female | >=30 | -1 | 0.042 |

*table continued on next page*

***Spouse Age-Difference Distributions by Individual Birth Year, Gender, and Age Group***
*(continued)*

| Birth Year | Gender | Age Group | Age Difference | Probability |
|---|---|---|---|---|
| <=1951 | Female | >=30 | 0 | 0.078 |
| <=1951 | Female | >=30 | 1 | 0.053 |
| <=1951 | Female | >=30 | 2 | 0.058 |
| <=1951 | Female | >=30 | 3 | 0.053 |
| <=1951 | Female | >=30 | 4 | 0.054 |
| <=1951 | Female | >=30 | 5 | 0.050 |
| <=1951 | Female | >=30 | 6 | 0.049 |
| <=1951 | Female | >=30 | 7 | 0.040 |
| <=1951 | Female | >=30 | 8 | 0.032 |
| <=1951 | Female | >=30 | 9 | 0.146 |
| | | | | |
| >=1952 | Male | <=19 | -9 | 0 |
| >=1952 | Male | <=19 | -8 | 0 |
| >=1952 | Male | <=19 | -7 | 0 |
| >=1952 | Male | <=19 | -6 | 0.004 |
| >=1952 | Male | <=19 | -5 | 0.003 |
| >=1952 | Male | <=19 | -4 | 0.035 |
| >=1952 | Male | <=19 | -3 | 0.092 |
| >=1952 | Male | <=19 | -2 | 0.136 |
| >=1952 | Male | <=19 | -1 | 0.258 |
| >=1952 | Male | <=19 | 0 | 0.209 |
| >=1952 | Male | <=19 | 1 | 0.125 |
| >=1952 | Male | <=19 | 2 | 0.056 |
| >=1952 | Male | <=19 | 3 | 0.031 |
| >=1952 | Male | <=19 | 4 | 0.021 |
| >=1952 | Male | <=19 | 5 | 0.006 |
| >=1952 | Male | <=19 | 6 | 0.006 |
| >=1952 | Male | <=19 | 7 | 0.004 |
| >=1952 | Male | <=19 | 8 | 0.004 |
| >=1952 | Male | <=19 | 9 | 0.010 |
| >=1952 | Male | 20–29 | -9 | 0.007 |
| >=1952 | Male | 20–29 | -8 | 0.010 |
| >=1952 | Male | 20–29 | -7 | 0.019 |

*table continued on next page*

***Spouse Age-Difference Distributions by Individual Birth Year, Gender, and Age Group***
*(continued)*

| Birth Year | Gender | Age Group | Age Difference | Probability |
|---|---|---|---|---|
| >=1952 | Male | 20–29 | -6 | 0.033 |
| >=1952 | Male | 20–29 | -5 | 0.045 |
| >=1952 | Male | 20–29 | -4 | 0.089 |
| >=1952 | Male | 20–29 | -3 | 0.122 |
| >=1952 | Male | 20–29 | -2 | 0.166 |
| >=1952 | Male | 20–29 | -1 | 0.161 |
| >=1952 | Male | 20–29 | 0 | 0.131 |
| >=1952 | Male | 20–29 | 1 | 0.079 |
| >=1952 | Male | 20–29 | 2 | 0.039 |
| >=1952 | Male | 20–29 | 3 | 0.035 |
| >=1952 | Male | 20–29 | 4 | 0.017 |
| >=1952 | Male | 20–29 | 5 | 0.018 |
| >=1952 | Male | 20–29 | 6 | 0.007 |
| >=1952 | Male | 20–29 | 7 | 0.009 |
| >=1952 | Male | 20–29 | 8 | 0.004 |
| >=1952 | Male | 20–29 | 9 | 0.009 |
| >=1952 | Male | >=30 | -9 | 0.135 |
| >=1952 | Male | >=30 | -8 | 0.047 |
| >=1952 | Male | >=30 | -7 | 0.069 |
| >=1952 | Male | >=30 | -6 | 0.053 |
| >=1952 | Male | >=30 | -5 | 0.078 |
| >=1952 | Male | >=30 | -4 | 0.093 |
| >=1952 | Male | >=30 | -3 | 0.105 |
| >=1952 | Male | >=30 | -2 | 0.049 |
| >=1952 | Male | >=30 | -1 | 0.064 |
| >=1952 | Male | >=30 | 0 | 0.053 |
| >=1952 | Male | >=30 | 1 | 0.063 |
| >=1952 | Male | >=30 | 2 | 0.052 |
| >=1952 | Male | >=30 | 3 | 0.037 |
| >=1952 | Male | >=30 | 4 | 0.023 |
| >=1952 | Male | >=30 | 5 | 0.023 |
| >=1952 | Male | >=30 | 6 | 0.011 |
| >=1952 | Male | >=30 | 7 | 0.018 |

*table continued on next page*

***Spouse Age-Difference Distributions by Individual Birth Year, Gender, and Age Group***
*(continued)*

| Birth Year | Gender | Age Group | Age Difference | Probability |
|---|---|---|---|---|
| >=1952 | Male | >=30 | 8 | 0.003 |
| >=1952 | Male | >=30 | 9 | 0.024 |
| | | | | |
| >=1952 | Female | <=19 | -9 | 0 |
| >=1952 | Female | <=19 | -8 | 0 |
| >=1952 | Female | <=19 | -7 | 0 |
| >=1952 | Female | <=19 | -6 | 0 |
| >=1952 | Female | <=19 | -5 | 0 |
| >=1952 | Female | <=19 | -4 | 0 |
| >=1952 | Female | <=19 | -3 | 0 |
| >=1952 | Female | <=19 | -2 | 0.012 |
| >=1952 | Female | <=19 | -1 | 0.041 |
| >=1952 | Female | <=19 | 0 | 0.105 |
| >=1952 | Female | <=19 | 1 | 0.195 |
| >=1952 | Female | <=19 | 2 | 0.155 |
| >=1952 | Female | <=19 | 3 | 0.146 |
| >=1952 | Female | <=19 | 4 | 0.111 |
| >=1952 | Female | <=19 | 5 | 0.074 |
| >=1952 | Female | <=19 | 6 | 0.050 |
| >=1952 | Female | <=19 | 7 | 0.030 |
| >=1952 | Female | <=19 | 8 | 0.026 |
| >=1952 | Female | <=19 | 9 | 0.055 |
| >=1952 | Female | 20–29 | -9 | 0 |
| >=1952 | Female | 20–29 | -8 | 0 |
| >=1952 | Female | 20–29 | -7 | 0.002 |
| >=1952 | Female | 20–29 | -6 | 0.003 |
| >=1952 | Female | 20–29 | -5 | 0.010 |
| >=1952 | Female | 20–29 | -4 | 0.019 |
| >=1952 | Female | 20–29 | -3 | 0.024 |
| >=1952 | Female | 20–29 | -2 | 0.038 |
| >=1952 | Female | 20–29 | -1 | 0.090 |
| >=1952 | Female | 20–29 | 0 | 0.128 |
| >=1952 | Female | 20–29 | 1 | 0.149 |

*table continued on next page*

**Spouse Age-Difference Distributions by Individual Birth Year, Gender, and Age Group** *(continued)*

| Birth Year | Gender | Age Group | Age Difference | Probability |
|---|---|---|---|---|
| >=1952 | Female | 20–29 | 2 | 0.127 |
| >=1952 | Female | 20–29 | 3 | 0.096 |
| >=1952 | Female | 20–29 | 4 | 0.075 |
| >=1952 | Female | 20–29 | 5 | 0.049 |
| >=1952 | Female | 20–29 | 6 | 0.042 |
| >=1952 | Female | 20–29 | 7 | 0.032 |
| >=1952 | Female | 20–29 | 8 | 0.026 |
| >=1952 | Female | 20–29 | 9 | 0.090 |
| >=1952 | Female | >=30 | -9 | 0.029 |
| >=1952 | Female | >=30 | -8 | 0.019 |
| >=1952 | Female | >=30 | -7 | 0.026 |
| >=1952 | Female | >=30 | -6 | 0.015 |
| >=1952 | Female | >=30 | -5 | 0.025 |
| >=1952 | Female | >=30 | -4 | 0.058 |
| >=1952 | Female | >=30 | -3 | 0.067 |
| >=1952 | Female | >=30 | -2 | 0.044 |
| >=1952 | Female | >=30 | -1 | 0.070 |
| >=1952 | Female | >=30 | 0 | 0.074 |
| >=1952 | Female | >=30 | 1 | 0.059 |
| >=1952 | Female | >=30 | 2 | 0.090 |
| >=1952 | Female | >=30 | 3 | 0.071 |
| >=1952 | Female | >=30 | 4 | 0.083 |
| >=1952 | Female | >=30 | 5 | 0.044 |
| >=1952 | Female | >=30 | 6 | 0.048 |
| >=1952 | Female | >=30 | 7 | 0.054 |
| >=1952 | Female | >=30 | 8 | 0.016 |
| >=1952 | Female | >=30 | 9 | 0.108 |

The probability distributions for spouse educational level are presented in the following table.

| ***Spouse Education Distributions by Birth Year, Gender, and Education*** | | | | |
|---|---|---|---|---|
| Birth Year | Gender | Ind.Education | Sp.Education | Probability |
| <=1951 | Male | HSdropout | HSdropout | 0.590 |
| <=1951 | Male | HSdropout | HSgraduate | 0.344 |
| <=1951 | Male | HSdropout | someCollege | 0.051 |
| <=1951 | Male | HSdropout | 4+yrCollege | 0.015 |
| <=1951 | Male | HSgraduate | HSdropout | 0.206 |
| <=1951 | Male | HSgraduate | HSgraduate | 0.620 |
| <=1951 | Male | HSgraduate | someCollege | 0.130 |
| <=1951 | Male | HSgraduate | 4+yrCollege | 0.044 |
| <=1951 | Male | someCollege | HSdropout | 0.070 |
| <=1951 | Male | someCollege | HSgraduate | 0.476 |
| <=1951 | Male | someCollege | someCollege | 0.293 |
| <=1951 | Male | someCollege | 4+yrCollege | 0.161 |
| <=1951 | Male | 4+yrCollege | HSdropout | 0.024 |
| <=1951 | Male | 4+yrCollege | HSgraduate | 0.375 |
| <=1951 | Male | 4+yrCollege | someCollege | 0.271 |
| <=1951 | Male | 4+yrCollege | 4+yrCollege | 0.330 |
| | | | | |
| <=1951 | Female | HSdropout | HSdropout | 0.679 |
| <=1951 | Female | HSdropout | HSgraduate | 0.259 |
| <=1951 | Female | HSdropout | someCollege | 0.052 |
| <=1951 | Female | HSdropout | 4+yrCollege | 0.010 |
| <=1951 | Female | HSgraduate | HSdropout | 0.249 |
| <=1951 | Female | HSgraduate | HSgraduate | 0.450 |
| <=1951 | Female | HSgraduate | someCollege | 0.190 |
| <=1951 | Female | HSgraduate | 4+yrCollege | 0.111 |
| <=1951 | Female | someCollege | HSdropout | 0.090 |
| <=1951 | Female | someCollege | HSgraduate | 0.250 |
| <=1951 | Female | someCollege | someCollege | 0.367 |
| <=1951 | Female | someCollege | 4+yrCollege | 0.293 |
| <=1951 | Female | 4+yrCollege | HSdropout | 0.051 |
| <=1951 | Female | 4+yrCollege | HSgraduate | 0.119 |
| <=1951 | Female | 4+yrCollege | someCollege | 0.274 |
| <=1951 | Female | 4+yrCollege | 4+yrCollege | 0.556 |

*table continued on next page*

***Spouse Education Distributions by Birth Year, Gender, and Education***
*(continued)*

| Birth Year | Gender | Ind.Education | Sp.Education | Probability |
|---|---|---|---|---|
| >=1952 | Male | HSdropout | HSdropout | 0.412 |
| >=1952 | Male | HSdropout | HSgraduate | 0.470 |
| >=1952 | Male | HSdropout | someCollege | 0.100 |
| >=1952 | Male | HSdropout | 4+yrCollege | 0.018 |
| >=1952 | Male | HSgraduate | HSdropout | 0.120 |
| >=1952 | Male | HSgraduate | HSgraduate | 0.593 |
| >=1952 | Male | HSgraduate | someCollege | 0.230 |
| >=1952 | Male | HSgraduate | 4+yrCollege | 0.057 |
| >=1952 | Male | someCollege | HSdropout | 0.036 |
| >=1952 | Male | someCollege | HSgraduate | 0.331 |
| >=1952 | Male | someCollege | someCollege | 0.443 |
| >=1952 | Male | someCollege | 4+yrCollege | 0.190 |
| >=1952 | Male | 4+yrCollege | HSdropout | 0.012 |
| >=1952 | Male | 4+yrCollege | HSgraduate | 0.148 |
| >=1952 | Male | 4+yrCollege | someCollege | 0.345 |
| >=1952 | Male | 4+yrCollege | 4+yrCollege | 0.495 |
| >=1952 | Female | HSdropout | HSdropout | 0.455 |
| >=1952 | Female | HSdropout | HSgraduate | 0.442 |
| >=1952 | Female | HSdropout | someCollege | 0.089 |
| >=1952 | Female | HSdropout | 4+yrCollege | 0.014 |
| >=1952 | Female | HSgraduate | HSdropout | 0.160 |
| >=1952 | Female | HSgraduate | HSgraduate | 0.572 |
| >=1952 | Female | HSgraduate | someCollege | 0.212 |
| >=1952 | Female | HSgraduate | 4+yrCollege | 0.056 |
| >=1952 | Female | someCollege | HSdropout | 0.045 |
| >=1952 | Female | someCollege | HSgraduate | 0.318 |
| >=1952 | Female | someCollege | someCollege | 0.431 |
| >=1952 | Female | someCollege | 4+yrCollege | 0.206 |
| >=1952 | Female | 4+yrCollege | HSdropout | 0.035 |
| >=1952 | Female | 4+yrCollege | HSgraduate | 0.181 |
| >=1952 | Female | 4+yrCollege | someCollege | 0.398 |
| >=1952 | Female | 4+yrCollege | 4+yrCollege | 0.386 |

## C.4 Childbirth Experience

The tables in this section show the results of fitting Cox regression models to the waiting time until childbirth of women in four groups according to their education level at age thirty. The four education levels are high-school dropout, graduated high school, some college, and graduated college. These models are estimated with data merged together from several PSID files.

Cox regression is a parametric method of fitting proportional hazard functions. We apply the results of fitting a Cox model by having SPSS generate the baseline hazard functions from each regression. Each baseline hazard function is a schedule that shows how the estimated hazard rate changes over duration or waiting time for the default case where the values of all the dummy variables are equal to zero. To generate the hazard function for a simulated individual we multiply the appropriate baseline hazard function by the exponential of $Xb$ where $X$ is the set of values for the independent variables and $b$ is the vector of coefficients from the regression.

The explanatory variables and factors included in the regression specifications for the three lower education levels are age, age squared, husband's educational level, woman's age cohort, and two factors relating to birth order. If the woman was unmarried at the time of birth, all four of the factors for the husband's educational level were set equal to zero. The summary regression results are displayed in the next three tables. The two digits at the end of the name for the birth cohorts indicate the first year of the woman's six-year birth cohort. For example, Cohort26 stands for the 1926–1931 cohort. Birthorder1 indicates that the birth is the woman's first, and Birthorder2 indicates the second. If the birth order is greater than two, both of these variables are set to zero. There is no constant term because one of the birth cohort factors is always equal to one.

***Time to Next Birth for Women Who Dropped Out of High School***

*Observations = 6716; Log-L = -31,638,205*

| Parameter | Estimate | Std. Error | p-Value |
|---|---|---|---|
| Age | 0.322 | .001 | .000 |
| AgeSquared | -0.008 | .000 | .000 |
| HusbandEduc1 | 1.141 | .003 | .000 |
| HusbandEduc2 | 0.985 | .003 | .000 |
| HusbandEduc3 | 0.924 | .004 | .000 |

*table continued on next page*

| ***Time to Next Birth for Women Who Dropped Out of High School*** *(continued)* | | | |
|---|---|---|---|
| Parameter | Estimate | Std. Error | p-Value |
| HusbandEduc4 | 1.237 | .007 | .000 |
| Cohort26 | 0.249 | .003 | .000 |
| Cohort32 | 0.170 | .003 | .000 |
| Cohort38 | -0.042 | .003 | .000 |
| Cohort44 | -0.079 | .003 | .000 |
| Cohort50 | -0.055 | .003 | .000 |
| Cohort56 | -0.170 | .003 | .000 |
| Cohort62 | 0.002 | .003 | .657 |
| BirthOrder1 | 0.471 | .002 | .000 |
| BirthOrder2 | 0.393 | .002 | .000 |

| ***Time to Next Birth for Women Who Graduated High School*** *Observations = 8921; Log-L = -52,109,667* | | | |
|---|---|---|---|
| Parameter | Estimate | Std. Error | p-Value |
| Age | 0.422 | .001 | .000 |
| AgeSquared | -0.010 | .000 | .000 |
| HusbandEduc1 | 1.312 | .003 | .000 |
| HusbandEduc2 | 1.201 | .003 | .000 |
| HusbandEduc3 | 1.309 | .003 | .000 |
| HusbandEduc4 | 1.159 | .003 | .000 |
| Cohort26 | 0.316 | .002 | .000 |
| Cohort32 | 0.116 | .002 | .000 |
| Cohort38 | 0.062 | .002 | .000 |
| Cohort44 | -0.069 | .002 | .000 |
| Cohort50 | -0.137 | .002 | .000 |
| Cohort56 | -0.020 | .002 | .000 |
| Cohort62 | -0.034 | .003 | .000 |
| BirthOrder1 | 0.714 | .002 | .000 |
| BirthOrder2 | 0.593 | .001 | .000 |

| ***Time to Next Birth for Women With Some College*** | | | |
|---|---|---|---|
| *Observations = 3807; Log-L = -19,241,416* | | | |
| Parameter | Estimate | Std. Error | p-Value |
| Age | 0.423 | .001 | .000 |
| AgeSquared | -0.009 | .000 | .000 |
| HusbandEduc1 | 1.357 | .005 | .000 |
| HusbandEduc2 | 1.085 | .004 | .000 |
| HusbandEduc3 | 0.943 | .004 | .000 |
| HusbandEduc4 | 1.176 | .004 | .000 |
| Cohort26 | 0.016 | .005 | .003 |
| Cohort32 | 0.559 | .005 | .000 |
| Cohort38 | -0.021 | .005 | .000 |
| Cohort44 | -0.343 | .005 | .000 |
| Cohort50 | -0.131 | .005 | .000 |
| Cohort56 | -0.019 | .005 | .000 |
| Cohort62 | -0.181 | .005 | .000 |
| BirthOrder1 | 0.647 | .003 | .000 |
| BirthOrder2 | 0.838 | .002 | .000 |

The fourth table in this set shows the results for the equation for women who graduated college at age thirty. The set of explanatory factors in this model is different from those in the first three models. For these women the factor MaritalStatus is set to one for a married woman; zero, otherwise. For this group of women only, the father's educational level was not significant in explaining variations in waiting time or overall fertility, but such women do have significantly more children when married than they do when they are single. These women tend to marry later and then have their children spaced closer together than the women in the other education categories. The age, age squared, marital status, and birth order factors in combination account for these differences.

| ***Time to Next Birth for Women with College Degree*** | | | |
|---|---|---|---|
| *Observations = 1730; Log-L = -10,282,489* | | | |
| Parameter | Estimate | Std. Error | p-Value |
| Age | 0.415 | .002 | .000 |

*table continued on next page*

| ***Time to Next Birth for Women with College Degree*** *(continued)* | | | |
|---|---|---|---|
| Parameter | Estimate | Std. Error | p-Value |
| AgeSquared | -0.009 | .000 | .000 |
| MaritalStatus | 2.310 | .009 | .000 |
| Cohort26 | -0.154 | .006 | .000 |
| Cohort32 | -0.789 | .007 | .000 |
| Cohort38 | -0.756 | .006 | .000 |
| Cohort44 | -1.017 | .005 | .000 |
| Cohort50 | -0.646 | .005 | .000 |
| Cohort56 | -0.728 | .006 | .000 |
| Cohort62 | -1.096 | .007 | .000 |
| BirthOrder1 | 0.579 | .004 | .000 |
| BirthOrder2 | 1.035 | .003 | .000 |

To work with the regression results we have SPSS generate the cumulative baseline hazard schedule. These four vectors, multiplied by 100 for easier reading, are displayed in the next table.

| ***Cumulative Baseline Hazard Schedules (times 100) For Childbirth Event In Four Educational Categories*** | | | | |
|---|---|---|---|---|
| Duration in Years | High-School Dropouts | Graduated High School | Some College | Graduated College |
| 1 | 0.19123 | 0.02883 | 0.01514 | 0.00854 |
| 2 | 0.62678 | 0.08903 | 0.05002 | 0.02707 |
| 3 | 0.96132 | 0.14430 | 0.08392 | 0.05448 |
| 4 | 1.21796 | 0.18896 | 0.12147 | 0.07567 |
| 5 | 1.42377 | 0.22283 | 0.14461 | 0.08896 |
| 6 | 1.61112 | 0.24807 | 0.16377 | 0.10559 |
| 7 | 1.76323 | 0.27513 | 0.17842 | 0.11706 |
| 8 | 1.84253 | 0.29581 | 0.19034 | 0.12344 |
| 9 | 1.99749 | 0.31060 | 0.20312 | 0.13030 |
| 10 | 2.07122 | 0.32349 | 0.21322 | 0.13370 |
| 11 | 2.12642 | 0.33756 | 0.21932 | 0.13510 |
| 12 | 2.19417 | 0.34496 | 0.22913 | 0.13699 |

*table continued on next page*

***Cumulative Baseline Hazard Schedules (times 100)***
***For Childbirth Event***
***In Four Educational Categories***
*(continued)*

| Duration in Years | High-School Dropouts | Graduated High School | Some College | Graduated College |
|---|---|---|---|---|
| 13 | 2.20654 | 0.34973 | 0.23218 | 0.13906 |
| 14 | 2.22196 | 0.35574 | 0.23502 | 0.13906 |
| 15 | 2.23239 | 0.35961 | 0.24290 | 0.14166 |
| 16 | 2.23718 | 0.36064 | 0.24519 | 0.14999 |
| 17 | 2.26097 | 0.36306 | 0.24519 | 0.15017 |
| 18 | 2.26187 | 0.36314 | 0.24519 | 0.15017 |
| 19 | 2.26187 | 0.36403 | 0.24519 | 0.15017 |
| 20 | 2.26285 | 0.36415 | 0.24519 | 0.15225 |
| 21 | 2.29443 | 0.36495 | 0.24519 | 0.15462 |
| 22 | 2.29443 | 0.36495 | 0.24519 | 0.15462 |
| 23 | 2.29443 | 0.36495 | 0.24634 | 0.15462 |
| 24 | 2.29443 | 0.36495 | 0.24634 | 0.15462 |
| 25 | 2.29520 | 0.36504 | 0.24634 | 0.15462 |
| 26 | 2.29520 | 0.37063 | 0.24634 | 0.15462 |
| 27 | 2.29520 | 0.37063 | 0.24634 | 0.15462 |

While the model actually works with the baseline hazard functions, the differences among the educational levels is see more easily in the survival functions displayed in the next table. Because the baseline functions are computed with all the explanatory variables and factors set to zero, they do not correspond to any of the women used in the analysis. The age variable, for example, is set to zero. Therefore, one cannot draw conclusions about the differences simply by examining the schedules shown in the table above. The survival curves, on the other hand, are computed by SPSS at the means of the factors and variables. The mean ages are all in the early twenties, for example. Comparing the four schedules shows immediately the general tendency for women with more education to have fewer children. For example, the survival rate at ten years for women who have graduated college is 70.0 percent; for women who dropped out of high school it is only 40.9 percent. In other words, it is about twice as likely that a high-school dropout in her early twenties will have a child in the next ten years as for a college-educated woman of the same age.

***Estimated Survival Schedules***
***For Childbirth Event***
***In Four Educational Categories***

| Duration in Years | High-School Dropouts | Graduated High School | Some College | Graduated College |
|---|---|---|---|---|
| 1 | 92.1 | 94.7 | 96.1 | 97.8 |
| 2 | 76.3 | 84.6 | 87.8 | 93.0 |
| 3 | 66.0 | 76.3 | 80.4 | 86.5 |
| 4 | 59.1 | 70.2 | 73.0 | 81.7 |
| 5 | 54.1 | 65.9 | 68.7 | 78.9 |
| 6 | 49.9 | 62.8 | 65.4 | 75.5 |
| 7 | 46.7 | 59.7 | 62.9 | 73.2 |
| 8 | 45.1 | 57.5 | 61.0 | 72.0 |
| 9 | 42.2 | 55.9 | 59.0 | 70.7 |
| 10 | 40.9 | 54.6 | 57.5 | 70.0 |
| 11 | 39.9 | 53.1 | 56.6 | 69.8 |
| 12 | 38.8 | 52.4 | 55.2 | 69.4 |
| 13 | 38.6 | 51.9 | 54.7 | 69.1 |
| 14 | 38.3 | 51.4 | 54.3 | 69.1 |
| 15 | 38.1 | 51.0 | 53.2 | 68.6 |
| 16 | 38.1 | 50.9 | 52.9 | 67.1 |
| 17 | 37.7 | 50.7 | 52.9 | 67.0 |
| 18 | 37.7 | 50.7 | 52.9 | 67.0 |
| 19 | 37.7 | 50.6 | 52.9 | 67.0 |
| 20 | 37.6 | 50.6 | 52.9 | 66.7 |
| 21 | 37.1 | 50.5 | 52.9 | 66.2 |
| 22 | 37.1 | 50.5 | 52.9 | 66.2 |
| 23 | 37.1 | 50.5 | 52.8 | 66.2 |
| 24 | 37.1 | 50.5 | 52.8 | 66.2 |
| 25 | 37.1 | 50.5 | 52.8 | 66.2 |
| 26 | 37.1 | 50.0 | 52.8 | 66.2 |
| 27 | 37.1 | 50.0 | 52.8 | 66.2 |

## C.5 Disability Experience

The functional disability hazard function parameters are drawn from the MINT project (Panis and Lillard 1999, pages 31–32). The estimated coefficients of the two-piece Gompertz hazard function are shown in the following table, along with the coefficients used in PENSIM. The calibrated constant coefficient is formed from the estimated constant and ethnicity coefficients by assuming blacks and Hispanics each constitute ten percent of the population while the other two groups constitute zero percent.

***Functional Disability Hazard Function Coefficients***

| Covariate | Panis-Lillard Estimates | Assumed PENSIM Values |
|---|---:|---:|
| Constant | -7.3766 | -7.3656 |
| Age slope, 30–45 | 0.0526 | 0.0526 |
| Age slope, 45+ | 0.1746 | 0.1746 |
| Male | 0.0062 | 0.0062 |
| High school dropout | 0.7312 | -0.7312 |
| College graduate | -0.6668 | -0.6668 |
| Black | 0.2779 | — |
| American Indian, Eskimo or Aleut | -0.5446 | — |
| Asian or Pacific Islander | -0.5249 | — |
| Hispanic | -0.1674 | — |

The Disability Insurance (DI) onset annual hazard rates are calibrated so that PENSIM produces DI prevalence rates that are similar to those projected in the OASDI Trustees Report. Using the social security Trustees' Report (2005) intermediate-cost projection as the calibration target, the annual probability of DI onset among those who are functionally disabled are shown in the table below by calendar year. Note that the DI program was established in 1958, and therefore, PENSIM assumes there were no DI beneficiaries until after 1960. The annual probability of DI onset for years not listed in the following table are calculated using linear interpolation and the closest surrounding calendar years; probabilities beyond the last calendar year are all assumed to be the same as for the last calendar year listed.

***Annual Probability of DI Onset***

*Note: values in unlisted years are calculated using linear interpolation*

| Calendar Year | Probability |
|---|---|
| 1940 | 0.000 |
| 1960 | 0.000 |
| 1980 | 0.120 |
| 2020 | 0.120 |
| 2030 | 0.165 |

## C.6 First-Job Timing

The following two tables contain gender-specific non-parametric hazard functions for starting the first job in a work career. The two sets of hazard functions pertain to two schooling phases: the time while attending high school and the time while attending college.

These two tables appear on the following two pages.

The initial first-job-timing table presents the gender-specific hazard rates for starting a first job before finishing high school. Age 41 hazard rates represent censoring.

| ***First-Job-Start Age During High School*** | | |
|---|---|---|
| Age | Female Hazard Rate | Male Hazard Rate |
| 14 | 0.0229 | 0.0339 |
| 15 | 0.0584 | 0.0683 |
| 16 | 0.1215 | 0.1293 |
| 17 | 0.1272 | 0.1267 |
| 18 | 0.0557 | 0.0783 |
| 19 | 0.0516 | 0.0777 |
| 20 | 0.0743 | 0.0929 |
| 21 | 0.1332 | 0.1021 |
| 22 | 0.0529 | 0.0672 |
| 23 | 0.0266 | 0.0994 |
| 24 | 0.0564 | 0.0287 |
| 25 | 0.0000 | 0.1017 |
| 26 | 0.0949 | 0.0698 |
| 27 | 0.0294 | 0.1364 |
| 28 | 0.0506 | 0.0000 |
| 29 | 0.0660 | 0.1457 |
| 30 | 0.0775 | 0.3540 |
| 31 | 0.0566 | 0.0000 |
| 32 | 0.0000 | 0.0000 |
| 33 | 0.0876 | 0.0000 |
| 34 | 0.0000 | 1.0000 |
| 35 | 0.1963 | 1.0000 |
| 36 | 0.0000 | 1.0000 |
| 37 | 0.0000 | 1.0000 |
| 38 | 0.0000 | 1.0000 |
| 39 | 0.0000 | 1.0000 |
| 40 | 0.0000 | 1.0000 |
| 41 | 1.0000 | 1.0000 |

The second first-job-timing table presents the gender-specific hazard rates for starting a first job a certain number of years after starting college and before finishing college. Year 27 hazard rates represent censoring.

| ***First-Job-Start Waiting Time During College*** | | |
|---|---|---|
| Years | Female Hazard Rate | Male Hazard Rate |
| 0 | 0.1862 | 0.1806 |
| 1 | 0.1690 | 0.1596 |
| 2 | 0.1495 | 0.1399 |
| 3 | 0.1675 | 0.1559 |
| 4 | 0.3858 | 0.3594 |
| 5 | 0.3247 | 0.3162 |
| 6 | 0.2170 | 0.3081 |
| 7 | 0.2250 | 0.2156 |
| 8 | 0.1225 | 0.3078 |
| 9 | 0.1769 | 0.3717 |
| 10 | 0.2238 | 0.1639 |
| 11 | 0.3054 | 0.1773 |
| 12 | 0.0812 | 0.1941 |
| 13 | 0.1907 | 0.6657 |
| 14 | 0.1025 | 0.0000 |
| 15 | 0.1540 | 0.0000 |
| 16 | 0.0000 | 0.0000 |
| 17 | 0.0000 | 0.0000 |
| 18 | 0.0000 | 0.0000 |
| 19 | 0.0000 | 0.0000 |
| 20 | 0.0000 | 0.0000 |
| 21 | 0.0000 | 0.0000 |
| 22 | 0.0000 | 0.0000 |
| 23 | 0.0000 | 0.0000 |
| 24 | 0.0000 | 0.0000 |
| 25 | 0.0000 | 0.0000 |
| 26 | 0.0000 | 0.0000 |
| 27 | 1.0000 | 1.0000 |

## C.7 First-Job Characteristics

The following six tables contain the estimated parameters of the equations that specify the characteristics of first jobs and the employers offering those jobs. The tables contain maximum-likelihood estimates of the parameters of models in which the dependent variable is employer industry (10 categories), job unionization (2 categories), job hours (2 categories), job earnings (continuous), employer firm size (3 categories), and employer pension sponsorship (2 categories).

In all six of the following tables, C denotes the constant term, AGE the age of the individual at the start of the first job, FEMALE denotes a dummy variable for gender (male is the omitted category), HSA is a dummy variable indicating highest level of schooling at the start of the first job is attended or attending high school, COA is a dummy variable indicating attended or attending college is the highest level of schooling attained at the start of the first job, and COG is a dummy variable indicating the highest level of schooling attained at the start of the first job is either a four-year college degree or a graduate degree. The category of high school graduate is the omitted schooling dummy variable. This means that the constant term represents men who have high school diplomas as their highest level of schooling when they start their first jobs.

The first of the six first-job-characteristics tables presents the multinomial logit parameters of employer industry, where the one-digit SIPP industry codes are as follows: (0) agriculture, mining, and construction, (1) product manufacturing, (2 & 3) equipment manufacturing, (4) transportation and communication, (5) wholesale trade, (6) retail trade, (7) finance, insurance, real estate, and personal services, (8) professional services, and (9) government. The last category (government) is the category whose parameters are normalized to zero. In the following table the number appended to each variable indicates the industry category to which it corresponds.

***First-Job Employer Industry***

*Observations = 3689; log-L = -7274.78*

| Parameter | Estimate | Std. Error | t-Statistic |
|---|---|---|---|
| C0 | -0.424319 | 0.321711 | -1.31895 |
| AGE0 | 0.042394 | 0.011564 | 3.66616 |
| FEMALE0 | -2.21739 | 0.187737 | -11.8112 |
| HSA0 | 0.522009 | 0.200735 | 2.60048 |

*table continued on next page*

***First-Job Employer Industry***
*(continued)*

| Parameter | Estimate | Std. Error | t-Statistic |
|---|---|---|---|
| COA0 | -1.04908 | 0.218038 | -4.81146 |
| COG0 | -1.52598 | 0.310264 | -4.91834 |
| | | | |
| C1 | -0.073509 | 0.336969 | -0.218147 |
| AGE1 | -0.007042 | 0.012335 | -0.570885 |
| FEMALE1 | -0.314271 | 0.171991 | -1.82725 |
| HSA1 | 0.229770 | 0.211192 | 1.08797 |
| COA1 | -0.956886 | 0.222909 | -4.29272 |
| COG1 | -0.985762 | 0.284041 | -3.47049 |
| | | | |
| C2 | -0.894028 | 0.434190 | -2.05907 |
| AGE2 | 0.014655 | 0.015932 | 0.919845 |
| FEMALE2 | -1.44058 | 0.232482 | -6.19653 |
| HSA2 | 0.512706 | 0.264076 | 1.94151 |
| COA2 | -0.769152 | 0.291376 | -2.63973 |
| COG2 | -1.45728 | 0.463144 | -3.14649 |
| | | | |
| C3 | -0.224525 | 0.376518 | -0.596319 |
| AGE3 | 0.009158 | 0.014033 | 0.652589 |
| FEMALE3 | -1.13880 | 0.196741 | -5.78830 |
| HSA3 | -0.301282 | 0.247293 | -1.21832 |
| COA3 | -0.937467 | 0.235718 | -3.97708 |
| COG3 | -1.79577 | 0.399463 | -4.49545 |
| | | | |
| C4 | 0.201746 | 0.439691 | 0.458836 |
| AGE4 | -0.027102 | 0.017253 | -1.57088 |
| FEMALE4 | -1.02746 | 0.217984 | -4.71346 |
| HSA4 | -0.519744 | 0.295504 | -1.75884 |
| COA4 | -0.569841 | 0.252938 | -2.25289 |
| COG4 | -0.946713 | 0.365688 | -2.58886 |
| | | | |
| C5 | 0.893727 | 0.333283 | 2.68159 |
| AGE5 | -0.040640 | 0.012741 | -3.18979 |
| FEMALE5 | -0.294544 | 0.164665 | -1.78875 |

*table continued on next page*

| ***First-Job Employer Industry*** *(continued)* | | | |
|---|---|---|---|
| Parameter | Estimate | Std. Error | t-Statistic |
| HSA5 | 0.061650 | 0.204626 | 0.301283 |
| COA5 | -0.748287 | 0.201133 | -3.72037 |
| COG5 | -1.83481 | 0.375168 | -4.89062 |
| | | | |
| C6 | 2.54599 | 0.253615 | 10.0388 |
| AGE6 | -0.074077 | 0.009547 | -7.75961 |
| FEMALE6 | -0.135304 | 0.123088 | -1.09925 |
| HSA6 | 0.725314 | 0.158714 | 4.56996 |
| COA6 | -0.575083 | 0.150935 | -3.81013 |
| COG6 | -1.14881 | 0.217291 | -5.28696 |
| | | | |
| C7 | 0.174053 | 0.268885 | 0.647314 |
| AGE7 | 0.004702 | 0.009528 | 0.493487 |
| FEMALE7 | -0.057771 | 0.137894 | -0.418954 |
| HSA7 | 0.312965 | 0.176556 | 1.77261 |
| COA7 | -0.516491 | 0.163108 | -3.16657 |
| COG7 | -0.792776 | 0.209815 | -3.77845 |
| | | | |
| C8 | 0.067294 | 0.266652 | 0.252365 |
| AGE8 | -0.005939 | 0.009338 | -0.636028 |
| FEMALE8 | 0.311667 | 0.136502 | 2.28323 |
| HSA8 | 0.252616 | 0.181671 | 1.39051 |
| COA8 | 0.078899 | 0.155543 | 0.507248 |
| COG8 | -0.017469 | 0.188660 | -0.092595 |

The second table begins on the next page.

The second of the six first-job-characteristics tables contains the binary logit parameters of job unionization, where non-unionization is the category whose parameters are normalized to zero. The individual attribute variables are defined in the same way they are above. The industry codes described above are translated into a set of nine dummy variables with industry code 9 (government) as the omitted dummy variable.

***First-Job Unionized***

*Observations = 3659; log-L = -928.597*

| Parameter | Estimate | Std. Error | t-Statistic |
|---|---|---|---|
| C | -1.57632 | 0.275501 | -5.72166 |
| AGE | 0.028560 | 0.009282 | 3.07678 |
| FEMALE | -0.081057 | 0.135443 | -0.598459 |
| HSA | -0.605035 | 0.172296 | -3.51160 |
| COA | -0.202939 | 0.155779 | -1.30274 |
| COG | -0.238779 | 0.208670 | -1.14429 |
| IND0 | -1.27560 | 0.259778 | -4.91035 |
| IND1 | -0.884686 | 0.237605 | -3.72334 |
| IND2 | -0.850377 | 0.321086 | -2.64844 |
| IND3 | -0.639611 | 0.252027 | -2.53787 |
| IND4 | -0.096964 | 0.248809 | -0.389713 |
| IND5 | -2.55588 | 0.429022 | -5.95746 |
| IND6 | -1.95248 | 0.206428 | -9.45838 |
| IND7 | -2.92679 | 0.353918 | -8.26968 |
| IND8 | -1.81324 | 0.216054 | -8.39254 |

The third table begins on the next page.

The third of the six first-job-characteristics tables contains the binary logit parameters of part-time hours, where full-time hours is the category whose parameters are normalized to zero. The individual attribute and job characteristic variables are defined in the same way they are above. The UNION dummy variable indicates a unionized job.

***First-Job Part-time Hours***

*Observations = 3651; log-L = -2092.08*

| Parameter | Estimate | Std. Error | t-Statistic |
|---|---|---|---|
| C | 0.002255 | 0.190598 | 0.011833 |
| AGE | -0.024964 | 0.005917 | -4.21925 |
| FEMALE | 0.678232 | 0.082143 | 8.25677 |
| HSA | 0.758363 | 0.095745 | 7.92069 |
| COA | -0.183360 | 0.098633 | -1.85902 |
| COG | -0.562173 | 0.143578 | -3.91544 |
| IND0 | -0.540235 | 0.178100 | -3.03333 |
| IND1 | -1.38773 | 0.203540 | -6.81795 |
| IND2 | -1.92757 | 0.320823 | -6.00820 |
| IND3 | -1.62300 | 0.261324 | -6.21070 |
| IND4 | -0.386000 | 0.232593 | -1.65955 |
| IND5 | 0.439038 | 0.169865 | 2.58463 |
| IND6 | 1.03374 | 0.129868 | 7.95990 |
| IND7 | -0.285710 | 0.139907 | -2.04215 |
| IND8 | 0.113690 | 0.133718 | 0.850221 |
| UNION | -0.995348 | 0.157936 | -6.30221 |

The fourth table begins on the next page.

The fourth of the six first-job-characteristics tables contains the linear regression parameters of log-earnings. The individual attribute and job characteristic variables are defined in the same way they are above. The PTIME dummy variable indicates a part-time job with the omitted category being a full-time job.

***First-Job Log-Earnings***
*Observations = 3589; R-squared = 0.3819*
*Mean of dependent variable = 6.11152*
*Standard error of regression = 0.794303*
*Calibrated standard error of regression = 0.0*

| Parameter | Estimate | Std. Error | t-Statistic |
|---|---|---|---|
| C | 6.30903 | 0.070912 | 88.9693 |
| AGE | 0.011170 | 0.002121 | 5.26723 |
| FEMALE | -0.103378 | 0.028900 | -3.57707 |
| HSA | -0.166271 | 0.034025 | -4.88677 |
| COA | 0.202226 | 0.035741 | 5.65804 |
| COG | 0.443941 | 0.050137 | 8.85463 |
| IND0 | 0.086454 | 0.064474 | 1.34091 |
| IND1 | 0.025928 | 0.066140 | 0.392022 |
| IND2 | 0.068924 | 0.087462 | 0.788041 |
| IND3 | 0.110850 | 0.074976 | 1.47847 |
| IND4 | 0.188966 | 0.083481 | 2.26359 |
| IND5 | 0.094613 | 0.064033 | 1.47757 |
| IND6 | -0.056375 | 0.048001 | -1.17447 |
| IND7 | -0.057092 | 0.052666 | -1.08405 |
| IND8 | 0.046261 | 0.050531 | 0.915494 |
| UNION | 0.168294 | 0.049881 | 3.37391 |
| PTIME | -0.986125 | 0.030049 | -32.8170 |

Note that for about 27 percent of all the first-job observations the beginning of the first-job spell began before the first SIPP interview. In these cases, the only earnings information available was for the time just before the first SIPP interview. These mid-spell data were adjusted to represent start-of-spell earnings levels using the following procedure. First, the number of years between the start of the first job and the first SIPP interview is determined. For about 73 percent of the first-job observations, the number of adjustment years is zero or less, and therefore, no earnings adjustment is done. And second, the positive number of adjustment years and the gender of the individual are used in a look-up table to obtain an earnings adjustment factor (less than one in value), which is multiplied by the reported mid-spell earnings to produce an estimate of start-of-spell earnings. The entries in the look-up table are calculated using aggregate historical data on annual real (not nominal) wage growth and on gender-specific career-wage growth derived from an analysis of March 1995 CPS age-earnings profiles. These profiles imply annual age-related growth rates in earnings of 3.77 percent for women and 5.40 percent for men over the 21–46 age range.

The fifth table begins on the next page.

The fifth of the six first-job-characteristics tables presents the multinomial logit parameters of employer firm size. Employer size is a categorical variable defined as follows: (1) 1–24 employees, (2) 25–99 employees, and (3) 100 or more employees. The parameters of the third size category are normalized to zero. The individual attribute and job characteristic variables are defined in the same way they are above. The LEARN variable denotes log-earnings at the start of the first job. In the following table the number appended to each variable indicates the size category to which it corresponds.

***First-Job Employer Firm Size***

*Observations = 1131; log-L = -903.582*

| Parameter | Estimate | Std. Error | t-Statistic |
|---|---|---|---|
| C1 | -2.93757 | 0.762756 | -3.85126 |
| AGE1 | 0.036093 | 0.012218 | 2.95399 |
| FEMALE1 | -0.059782 | 0.167211 | -0.357522 |
| HSA1 | -0.014327 | 0.205760 | -0.069631 |
| COA1 | -0.270116 | 0.186420 | -1.44896 |
| COG1 | -0.524042 | 0.304581 | -1.72053 |
| IND01 | 2.47022 | 0.413429 | 5.97495 |
| IND11 | 1.23229 | 0.408864 | 3.01392 |
| IND21 | 2.10665 | 0.446400 | 4.71920 |
| IND31 | 1.04108 | 0.428463 | 2.42979 |
| IND41 | 1.00991 | 0.492230 | 2.05171 |
| IND51 | 1.28953 | 0.446079 | 2.89080 |
| IND61 | 2.03891 | 0.359489 | 5.67170 |
| IND71 | 1.90561 | 0.360238 | 5.28987 |
| IND81 | 2.14436 | 0.351032 | 6.10873 |
| UNION1 | -0.482958 | 0.246447 | -1.95968 |
| PTIME1 | 0.068956 | 0.206429 | 0.334044 |
| LEARN1 | -0.050357 | 0.096306 | -0.522878 |
| | | | |
| C2 | -2.58685 | 0.994564 | -2.60099 |
| AGE2 | 0.001201 | 0.017201 | 0.069821 |
| FEMALE2 | -0.095135 | 0.215985 | -0.440470 |
| HSA2 | 0.241291 | 0.294989 | 0.817965 |
| COA2 | -0.068803 | 0.248524 | -0.276846 |
| COG2 | 0.231358 | 0.330710 | 0.699579 |

*table continued on next page*

| ***First-Job Employer Firm Size*** *(continued)* | | | |
|---|---|---|---|
| Parameter | Estimate | Std. Error | t-Statistic |
| IND02 | -0.391546 | 0.665977 | -0.587927 |
| IND12 | 0.329358 | 0.393241 | 0.837548 |
| IND22 | 0.320921 | 0.551796 | 0.581594 |
| IND32 | -0.774302 | 0.572311 | -1.35294 |
| IND42 | -0.315869 | 0.572730 | -0.551515 |
| IND52 | -0.138867 | 0.539004 | -0.257636 |
| IND62 | 0.215001 | 0.384293 | 0.559471 |
| IND72 | 0.531347 | 0.359174 | 1.47936 |
| IND82 | 1.03075 | 0.318373 | 3.23755 |
| UNION2 | 0.063697 | 0.270223 | 0.235720 |
| PTIME2 | 0.030119 | 0.279272 | 0.107847 |
| LEARN2 | 0.074036 | 0.132529 | 0.558637 |

The sixth and final first-job characteristics table begins on the next page.

The last of the six first-job-characteristics tables presents the binary logit parameters of employer pension sponsorship. Pension non-sponsorship is the category for which parameters are normalized to zero. The individual attribute and job characteristic variables are defined in the same way they are above. The FSIZE1 and FSIZE2 dummy variables indicate the 1–24 and the 25–99 employer size categories, with the largest of the three size categories being the omitted category.

***First-Job Employer Pension Sponsorship***

*Observations = 777; log-L = -344.332*

| Parameter | Estimate | Std. Error | t-Statistic |
|---|---|---|---|
| C | 1.43778 | 0.878627 | 1.63640 |
| AGE | -0.029120 | 0.015669 | -1.85848 |
| FEMALE | 0.145960 | 0.214617 | 0.680095 |
| HSA | -0.906740 | 0.274306 | -3.30558 |
| COA | 0.056613 | 0.230014 | 0.246130 |
| COG | 0.662001 | 0.391692 | 1.69010 |
| IND0 | -1.57632 | 0.476171 | -3.31041 |
| IND1 | -0.495171 | 0.426383 | -1.16133 |
| IND2 | -0.645245 | 0.501457 | -1.28674 |
| IND3 | -0.569705 | 0.448800 | -1.26940 |
| IND4 | -0.169042 | 0.588493 | -0.287246 |
| IND5 | -0.055521 | 0.554706 | -0.100091 |
| IND6 | -1.65038 | 0.405287 | -4.07214 |
| IND7 | -1.17153 | 0.385754 | -3.03700 |
| IND8 | -0.745021 | 0.372908 | -1.99787 |
| UNION | 1.51103 | 0.351164 | 4.30292 |
| PTIME | -0.298596 | 0.279708 | -1.06753 |
| LEARN | 0.213043 | 0.113321 | 1.88000 |
| FSIZE1 | -1.53685 | 0.218841 | -7.02266 |
| FSIZE2 | -0.777624 | 0.294316 | -2.64214 |

## C.8 First-Job Duration

The following four tables contain estimated parameters for different versions of the proportional hazard model of first-job duration. The tables contain simulated maximum-likelihood estimates of the parameters of a no-covariates version of the model, an all-covariates version, and a constrained all-covariates version of the model. The original and calibrated value of the constant term's parameter are both presented for the constrained all-covariates version of the model. It is the calibrated, constrained all-covariates version that is used in PENSIM.

The first of the three tables presents estimated parameters for the no-covariates version of the model. The S_VAR parameter is the variance of the normal variate (with mean equal to minus one-half the variance) that is used to generate the lognormal error term that has mean equal to one and represents spell-specific unmeasured heterogeneity. There is no individual-specific error term in this model because there is only one first-job spell for each individual. The B_CONST parameter is the coefficient of the proportional hazard model's constant term. And ALPHAM1 (i.e., $\alpha - 1$) is the Weibull shape parameter minus one. This variable is estimated rather than $\alpha$ to facilitate hypothesis testing on whether the shape parameter is significantly different from one, in which case the Weibull baseline hazard function simplifies to an exponential function.

***First-Job Duration: No Covariates***
*Observations = 2909; log-L = -3533.54*
*Mean duration in estimation sample = 2.28 years*

| Parameter | Estimate | Std. Error | t-Statistic |
|---|---|---|---|
| S_VAR | 2.111100 | 0.097583 | 21.63400 |
| B_CONST | 1.337370 | 0.074629 | 17.92020 |
| ALPHAM1 | 0.699066 | 0.127954 | 5.46341 |

The second table begins on the next page.

The second of the three tables contains estimated parameters for the all-covariates version of the model. The covariates used in the model are the same as those defined above in the first-job characteristics equations. The PENSION dummy variable indicates whether the first-job employer sponsors a pension plan; non-sponsorship is the omitted category.

***First-Job Duration: All Covariates***
*Observations = 214; log-L = -624.990*
*Mean duration in estimation sample = 9.67 years*

| Parameter | Estimate | Std. Error | t-Statistic |
|---|---|---|---|
| S_VAR | 0.000633 | 0.039901 | 0.015863 |
| B_CONST | -4.98141 | 0.823507 | -6.04902 |
| AGE | 0.023187 | 0.012252 | 1.89254 |
| FEMALE | -0.121171 | 0.175953 | -0.688658 |
| HSA | 0.865231 | 0.226306 | 3.82328 |
| COA | 0.658331 | 0.224080 | 2.93793 |
| COG | 0.231639 | 0.241742 | 0.958208 |
| IND0 | -0.014555 | 0.373555 | -0.038962 |
| IND1 | -0.523439 | 0.385016 | -1.35953 |
| IND2 | -0.816004 | 0.519594 | -1.57046 |
| IND3 | -0.603283 | 0.403901 | -1.49364 |
| IND4 | -0.410571 | 0.451255 | -0.909841 |
| IND5 | 0.256409 | 0.420288 | 0.610081 |
| IND6 | 0.450078 | 0.353318 | 1.27386 |
| IND7 | 0.207888 | 0.333174 | 0.623961 |
| IND8 | 0.106107 | 0.313230 | 0.338751 |
| UNION | -0.650290 | 0.294718 | -2.20648 |
| PTIME | 0.842008 | 0.202655 | 4.15488 |
| LEARN | 0.359284 | 0.110824 | 3.24194 |
| FSIZE1 | -0.251794 | 0.196141 | -1.28374 |
| FSIZE2 | -0.155060 | 0.305674 | -0.507274 |
| PENSION | -0.431943 | 0.177579 | -2.43240 |
| ALPHAM1 | 0.094469 | 0.070216 | 1.34540 |

Notice two things about these results: the variance of the unmeasured heterogeneity error is close to zero and the Weibull shape parameter is close to one. A joint significance test on these two parameters is performed after the next table of results is presented.

The last of the three tables contains estimated parameters for the constrained all-covariates version of the model. This version of the model constrains the lognormal variance to zero and the Weibull shape parameter to one (assuming an exponential baseline hazard function).

***First-Job Duration: All Covariates***
*Observations = 214; log-L = -626.298*
*Mean duration in estimation sample = 9.67 years*

| Parameter | Estimate | Std. Error | t-Statistic |
|---|---|---|---|
| B_CONST | -4.84971 | 0.938908 | -5.16527 |
| B_CONST | -3.68 | *calibrated value* | |
| AGE | 0.024338 | 0.014117 | 1.72407 |
| FEMALE | -0.117439 | 0.202408 | -0.580210 |
| HSA | 0.859306 | 0.259259 | 3.31447 |
| COA | 0.659166 | 0.260235 | 2.53297 |
| COG | 0.241205 | 0.279633 | 0.862574 |
| IND0 | 0.004777 | 0.433644 | 0.011015 |
| IND1 | -0.524801 | 0.446252 | -1.17602 |
| IND2 | -0.811086 | 0.594681 | -1.36390 |
| IND3 | -0.604807 | 0.470803 | -1.28463 |
| IND4 | -0.404806 | 0.511437 | -0.791508 |
| IND5 | 0.272163 | 0.476633 | 0.571011 |
| IND6 | 0.455786 | 0.411441 | 1.10778 |
| IND7 | 0.229142 | 0.381881 | 0.600035 |
| IND8 | 0.126562 | 0.360502 | 0.351071 |
| UNION | -0.664565 | 0.337970 | -1.96634 |
| PTIME | 0.828312 | 0.235272 | 3.52065 |
| LEARN | 0.339948 | 0.128066 | 2.65447 |
| FSIZE1 | -0.270245 | 0.226175 | -1.19485 |
| FSIZE2 | -0.159222 | 0.346162 | -0.459964 |
| PENSION | -0.437177 | 0.195223 | -2.23937 |

The likelihood ratio statistic for this constrained all-covariates model (versus the unconstrained all-covariates model) is $\chi^2(2) = 2.616$, while 5.99 is the 95% level for the $\chi^2(2)$ statistic. This indicates that there is no significant deterioration in the data-fitting ability of an all-covariate version of the model that assumes an exponential (rather than Weibull) baseline hazard function and assumes that there is no unmeasured heterogeneity in the sample.

## C.9 Non-Employment Duration

The following table contains estimated and calibrated parameters for the proportional hazard model of between-job duration. The parameters are estimated using a large sample of between-job spells, about one-third of which are right censored, drawn from the 1992 SIPP data, a merger of data on the 1991 panel (wave 7) and the 1992 panel (wave 4).

The S_VAR parameter is the variance of the normal variate (with mean equal to minus one-half its variance) that is used to generate the lognormal spell-specific error term that has mean equal to one. The I_VAR parameter is the variance of the normal variate (with mean equal to minus one-half its variance) that is used to generate the lognormal individual-specific error term that has mean equal to one. The B_CONST parameter is the coefficient of the proportional hazard model's constant term. And ALPHAM1 (i.e., $\alpha - 1$) is the Weibull shape parameter minus one. This variable is estimated rather than $\alpha$ to facilitate hypothesis testing on whether the shape parameter is significantly different from one.

The covariates of the proportional hazard model include a set of dummy variables that represent the individual's gender, age, and schooling level at the start of the between-job spell. They also include dummy variables that interact the age dummy variables with the FEMALE dummy variable and the schooling dummy variables with the FEMALE dummy variable. The individual's age at the start of the spell is categorized into six age ranges: teens (AGE1), twenties (AGE2), thirties (AGE3), forties (AGE4), fifties (AGE5), and sixties and above (AGE6); the omitted category is 30–39 years old (AGE3). The schooling categories are defined as above, with high school graduate again being the omitted category. The interaction terms are denoted by appending an F onto the name of the age or schooling dummy variable.

Some of the estimated coefficient have been calibrated to account for the fact that disability-related non-employment spells are being explicitly simulated in PENSIM. Also, for the same reason, the variance of the predicted waiting-time distribution has been reduced (by larger amounts at older ages) in a way that preserves the mean of the distribution.

The table begins on the next page.

| ***Between-Job Duration*** | | | |
|---|---|---|---|
| *Observations = 9379; log-L = -4571.21* | | | |
| *Mean duration in estimation sample = 0.62 years* | | | |
| Parameter | Estimate | Std. Error | t-Statistic |
| S_VAR | 0.064077 | 0.019179 | 3.34104 |
| S_VAR | 0.00 | *calibrated value* | |
| I_VAR | 1.18510 | 0.039177 | 30.2498 |
| I_VAR | 0.00 | *calibrated value* | |
| B_CONST | 1.42385 | 0.053466 | 26.6308 |
| B_CONST | -0.40 | *calibrated value* | |
| FEMALE | -0.539523 | 0.063528 | -8.49272 |
| FEMALE | -1.10 | *calibrated value* | |
| AGE1 | -0.201664 | 0.061450 | -3.28174 |
| AGE1 | 0.75 | *calibrated value* | |
| AGE2 | 0.036911 | 0.052954 | 0.697040 |
| AGE2 | 0.40 | *calibrated value* | |
| AGE4 | -0.022828 | 0.065201 | -0.350126 |
| AGE4 | -0.30 | *calibrated value* | |
| AGE5 | -0.544049 | 0.069406 | -7.83868 |
| AGE5 | -1.40 | *calibrated value* | |
| AGE6 | -1.46654 | 0.076008 | -19.2946 |
| AGE1F | 0.528750 | 0.085310 | 6.19799 |
| AGE2F | 0.122193 | 0.070928 | 1.72278 |
| AGE4F | 0.121362 | 0.088874 | 1.36556 |
| AGE5F | 0.410076 | 0.096482 | 4.25027 |
| AGE6F | 0.513555 | 0.105191 | 4.88214 |
| HSA | -0.170805 | 0.044776 | -3.81466 |
| COA | -0.018870 | 0.049546 | -0.380861 |
| COG | 0.067478 | 0.062264 | 1.08373 |
| HSAF | 0.010161 | 0.064549 | 0.157416 |
| COAF | 0.153069 | 0.065451 | 2.33869 |
| COGF | 0.250949 | 0.083741 | 2.99672 |
| ALPHAM1 | 1.49751 | 0.146846 | 10.1978 |
| ALPHAM1 | 0.00 | *calibrated value* | |

## C.10 Subsequent-Job Characteristics

The following six tables contain the estimated and calibrated parameters of the equations that specify the characteristics of subsequent jobs and the employers offering those jobs. As described above, the term subsequent job refers to any job held by an individual following that individual's first job. The tables contain maximum-likelihood estimates of the parameters of models in which the dependent variable is employer industry (10 categories), job unionization (2 categories), job hours (2 categories), job earnings (continuous), employer firm size (3 categories), and employer pension sponsorship (2 categories).

In all six of the following tables, C denotes the constant term and FEMALE denotes a dummy variable for gender (male is omitted category). Age at the start of the subsequent job is categorized into several age-range dummy variables: AGE1 denotes a dummy variable indicating age in the teens, AGE2 twenties, AGE2B denotes a dummy variable indicating age in the teens or twenties, AGE3 thirties, AGE4 forties, AGE5 fifties, AGE6 sixties, and AGE5A denotes the individual is fifty or more at the start of the subsequent job. The 30–39 age range (AGE3) is the omitted category. HSA is a dummy variable indicating highest level of schooling at the start of the subsequent job is attended or attending high school, COA is a dummy variable indicating attended or attending college is the highest level of schooling attained at the start of the subsequent job, and COG is a dummy variable indicating the highest level of schooling attained at the start of the subsequent job is either a four-year college degree or a graduate degree. The category of high school graduate is the omitted schooling dummy variable. This means that the constant term represents men in their thirties who have high school diplomas as their highest level of schooling when they start their subsequent jobs.

As described in Appendix B, Section B.19.1, some constant terms have been calibrated. In addition, correlation has been induced in some random variates used to realize the subsequent job characteristics so that the simulated correlations between the characteristics of pairs of subsequent jobs is similar to the correlations calculated in long-term job data drawn from PSID data. The degree of calibrated correlation often differs across the categories of a job characteristic variable.

The first of the six subsequent-job-characteristics tables presents the multinomial logit parameters of employer industry, where the one-digit SIPP industry codes are as follows: (0) agriculture, mining, and construction, (1) product manufacturing, (2 & 3) equipment manufacturing, (4) transportation and communication, (5) wholesale trade, (6) retail trade, (7) finance, insurance, real estate, and personal services, (8) professional services, and (9) government. The last category (government) is the category whose parameters are normalized to zero. In the following table the number appended to each variable indicates the industry category to which it corresponds.

***Subsequent-Job Employer Industry***

*Observations = 31034; log-L = -63986.2*

| Parameter | Estimate | Std. Error | t-Statistic |
|---|---|---|---|
| C0 | 0.993715 | 0.061259 | 16.2215 |
| C0 | 1.20 | *calibrated value* | |
| FEMALE0 | -2.18641 | 0.064759 | -33.7622 |
| AGE2B0 | -0.341343 | 0.061531 | -5.54749 |
| AGE40 | -0.011093 | 0.079602 | -0.139362 |
| AGE5A0 | -0.111698 | 0.090957 | -1.22803 |
| HSA0 | 0.603774 | 0.080368 | 7.51257 |
| COA0 | -1.26899 | 0.064684 | -19.6183 |
| COG0 | -2.09478 | 0.091886 | -22.7976 |
| $\epsilon$ correl0 | 0.45 | *calibrated value* | |
| | | | |
| C1 | -0.337500 | 0.072851 | -4.63273 |
| C1 | -0.90 | *calibrated value* | |
| FEMALE1 | -0.265479 | 0.056170 | -4.72633 |
| AGE2B1 | 0.028903 | 0.067749 | 0.426624 |
| AGE41 | 0.085403 | 0.087349 | 0.977719 |
| AGE5A1 | -0.049849 | 0.102912 | -0.484384 |
| HSA1 | 0.607992 | 0.089245 | 6.81264 |
| COA1 | -0.826850 | 0.069215 | -11.9461 |
| COG1 | -1.13607 | 0.085632 | -13.2668 |
| $\epsilon$ correl1 | 0.99 | *calibrated value* | |
| | | | |
| C2 | -0.181322 | 0.080820 | -2.24352 |
| C2 | -0.80 | *calibrated value* | |
| FEMALE2 | -1.34768 | 0.075023 | -17.9636 |
| AGE2B2 | -0.090436 | 0.081073 | -1.11548 |

*table continued on next page*

**_Subsequent-Job Employer Industry_**
*(continued)*

| Parameter | Estimate | Std. Error | t-Statistic |
|---|---|---|---|
| AGE42 | -0.083637 | 0.108574 | -0.770324 |
| AGE5A2 | -0.247375 | 0.125806 | -1.96632 |
| HSA2 | 0.687450 | 0.097361 | 7.06084 |
| COA2 | -1.24031 | 0.088486 | -14.0171 |
| COG2 | -1.69495 | 0.118048 | -14.3581 |
| $\epsilon$ correl2 | 0.99 | *calibrated value* | |
| | | | |
| C3 | 0.166838 | 0.068201 | 2.44629 |
| C3 | -0.60 | *calibrated value* | |
| FEMALE3 | -1.06006 | 0.056967 | -18.6082 |
| AGE2B3 | -0.013438 | 0.065077 | -0.206489 |
| AGE43 | -0.039766 | 0.086588 | -0.459257 |
| AGE5A3 | -0.106228 | 0.101134 | -1.05037 |
| HSA3 | 0.128711 | 0.094388 | 1.36364 |
| COA3 | -0.768664 | 0.065867 | -11.6699 |
| COG3 | -0.986703 | 0.078154 | -12.6251 |
| $\epsilon$ correl3 | 0.99 | *calibrated value* | |
| | | | |
| C4 | -0.120477 | 0.076057 | -1.58403 |
| C4 | -0.80 | *calibrated value* | |
| FEMALE4 | -1.34667 | 0.067338 | -19.9987 |
| AGE2B4 | 0.003440 | 0.073598 | 0.046734 |
| AGE44 | -0.101629 | 0.100497 | -1.01127 |
| AGE5A4 | 0.062495 | 0.110278 | 0.566705 |
| HSA4 | 0.150492 | 0.103970 | 1.44747 |
| COA4 | -0.623831 | 0.072673 | -8.58408 |
| COG4 | -1.13566 | 0.093000 | -12.2114 |
| $\epsilon$ correl4 | 0.99 | *calibrated value* | |
| | | | |
| C5 | -0.005251 | 0.068275 | -0.076907 |
| C5 | -0.80 | *calibrated value* | |
| FEMALE5 | -0.577637 | 0.053553 | -10.7862 |
| AGE2B5 | 0.127860 | 0.063685 | 2.00769 |
| AGE45 | -0.018995 | 0.085499 | -0.222162 |

*table continued on next page*

| ***Subsequent-Job Employer Industry*** *(continued)* | | | |
|---|---|---|---|
| Parameter | Estimate | Std. Error | t-Statistic |
| AGE5A5 | -0.240107 | 0.104228 | -2.30367 |
| HSA5 | 0.170531 | 0.091965 | 1.85430 |
| COA5 | -0.700902 | 0.063089 | -11.1097 |
| COG5 | -1.22428 | 0.081529 | -15.0165 |
| $\epsilon$ correl5 | 0.99 | *calibrated value* | |
| | | | |
| C6 | 0.240466 | 0.058420 | 4.11613 |
| C6 | -0.20 | *calibrated value* | |
| FEMALE6 | -0.055346 | 0.042513 | -1.30185 |
| AGE2B6 | 0.632009 | 0.051937 | 12.1689 |
| AGE46 | -0.153556 | 0.073388 | -2.09240 |
| AGE5A6 | -0.188171 | 0.085424 | -2.20279 |
| HSA6 | 0.689609 | 0.073709 | 9.35585 |
| COA6 | -0.696742 | 0.050772 | -13.7230 |
| COG6 | -1.41357 | 0.066965 | -21.1091 |
| $\epsilon$ correl6 | 0.97 | *calibrated value* | |
| | | | |
| C7 | 0.254304 | 0.057459 | 4.42582 |
| C7 | -0.50 | *calibrated value* | |
| FEMALE7 | 0.055883 | 0.042455 | 1.31630 |
| AGE2B7 | 0.087138 | 0.050676 | 1.71951 |
| AGE47 | 0.087366 | 0.065868 | 1.32638 |
| AGE5A7 | 0.106340 | 0.077200 | 1.37745 |
| HSA7 | 0.243043 | 0.079071 | 3.07374 |
| COA7 | -0.438147 | 0.051023 | -8.58721 |
| COG7 | -0.603848 | 0.058564 | -10.3109 |
| $\epsilon$ correl7 | 0.98 | *calibrated value* | |
| | | | |
| C8 | -0.451438 | 0.060232 | -7.49503 |
| C8 | -1.20 | *calibrated value* | |
| FEMALE8 | 0.688823 | 0.043198 | 15.9458 |
| AGE2B8 | 0.024374 | 0.049414 | 0.493270 |
| AGE48 | 0.058427 | 0.064040 | 0.912358 |
| AGE5A8 | 0.190880 | 0.075204 | 2.53816 |

*table continued on next page*

| ***Subsequent-Job Employer Industry*** *(continued)* | | | |
|---|---|---|---|
| Parameter | Estimate | Std. Error | t-Statistic |
| HSA8 | 0.209599 | 0.085058 | 2.46419 |
| COA8 | 0.173653 | 0.051541 | 3.36921 |
| COG8 | 0.233762 | 0.056585 | 4.13119 |
| $\epsilon$ correl8 | 0.97 | *calibrated value* | |
| | | | |
| $\epsilon$ correl9 | 0.57 | *calibrated value* | |

The second table begins on the next page.

The second of the six subsequent-job-characteristics tables contains the binary logit parameters of job unionization, where non-unionization is the category whose parameters are normalized to zero. The individual attribute variables are defined in the same way they are above. The industry code described above is translated into a set of nine dummy variables with industry code 9 (government) as the omitted dummy variable.

| ***Subsequent-Job Unionized*** | | | |
|---|---|---|---|
| *Observations = 30935; log-L = -10547.3* | | | |
| Parameter | Estimate | Std. Error | t-Statistic |
| C | -0.301607 | 0.055718 | -5.41309 |
| C | -0.80 | *calibrated value* | |
| FEMALE | -0.388040 | 0.039030 | -9.94220 |
| AGE2B | -0.223048 | 0.043280 | -5.15364 |
| AGE4 | 0.027180 | 0.055750 | 0.487539 |
| AGE5A | 0.245580 | 0.061606 | 3.98631 |
| HSA | -0.247817 | 0.056706 | -4.37020 |
| COA | -0.109704 | 0.043373 | -2.52930 |
| COG | -0.341599 | 0.054963 | -6.21507 |
| IND0 | -0.928611 | 0.063174 | -14.6993 |
| IND1 | -1.05380 | 0.073409 | -14.3552 |
| IND2 | -0.773057 | 0.082689 | -9.34897 |
| IND3 | -0.935027 | 0.067891 | -13.7724 |
| IND4 | -0.588711 | 0.071202 | -8.26818 |
| IND5 | -2.23383 | 0.101569 | -21.9932 |
| IND6 | -2.29259 | 0.075081 | -30.5351 |
| IND7 | -2.66996 | 0.086982 | -30.6955 |
| IND8 | -1.85272 | 0.064074 | -28.9154 |
| $\epsilon$ correl | 0.80 | *calibrated value* | |

The third table begins on the next page.

The third of the six subsequent-job-characteristics tables contains the binary logit parameters of part-time hours, where full-time hours is the category whose parameters are normalized to zero. The individual attribute and job characteristic variables are defined in the same way they are above. The UNION dummy variable indicates a unionized job.

| ***Subsequent-Job Part-time Hours*** | | | |
|---|---|---|---|
| *Observations = 30769; log-L = -16259.9* | | | |
| Parameter | Estimate | Std. Error | t-Statistic |
| C | -1.51068 | 0.053589 | -28.1899 |
| C | -0.30 | *calibrated value* | |
| FEMALE | 0.826681 | 0.029289 | 28.2250 |
| AGE2B | 0.388463 | 0.034067 | 11.4028 |
| AGE4 | 0.042133 | 0.046325 | 0.909517 |
| AGE5A | 0.519917 | 0.050724 | 10.2499 |
| HSA | 0.486468 | 0.041544 | 11.7097 |
| COA | -0.006393 | 0.033251 | -0.192260 |
| COG | -0.267474 | 0.042515 | -6.29125 |
| IND0 | -0.519369 | 0.066703 | -7.78630 |
| IND1 | -1.05410 | 0.078217 | -13.4766 |
| IND2 | -1.49913 | 0.118746 | -12.6247 |
| IND3 | -1.48117 | 0.091435 | -16.1992 |
| IND4 | -0.572486 | 0.082120 | -6.97134 |
| IND5 | -0.067816 | 0.061820 | -1.09700 |
| IND6 | 0.754597 | 0.047677 | 15.8274 |
| IND7 | -0.242857 | 0.050004 | -4.85676 |
| IND8 | 0.206470 | 0.047069 | 4.38658 |
| UNION | -0.769295 | 0.051613 | -14.9051 |
| $\epsilon$ correl | 0.96 | *calibrated value* | |

The fourth table begins on the next page.

The fourth of the six subsequent-job-characteristics tables contains the linear regression parameters of log-earnings. The individual attribute and job characteristic variables are defined in the same way they are above. The PTIME dummy variable indicates a part-time job with the omitted category being a full-time job.

***Subsequent-Job Log-Earnings***
*Observations = 30118; R-squared = 0.398503*
*Mean of dependent variable = 6.71327*

| Parameter | Estimate | Std. Error | t-Statistic |
|---|---|---|---|
| C | 7.03221 | 0.017276 | 407.042 |
| C | 7.10 | *calibrated value* | |
| FEMALE | -0.175914 | 0.009542 | -18.4363 |
| AGE2B | -0.215187 | 0.010765 | -19.9892 |
| AGE1 | -1.30 | *calibrated value* | |
| AGE2 | -0.10 | *calibrated value* | |
| AGE4 | 0.054753 | 0.014340 | 3.81824 |
| AGE4 | 0.20 | *calibrated value* | |
| AGE5A | 0.032890 | 0.016700 | 1.96942 |
| AGE5 | 0.10 | *calibrated value* | |
| AGE6 | -2.20 | *calibrated value* | |
| HSA | -0.245781 | 0.013802 | -17.8079 |
| COA | 0.211904 | 0.010810 | 19.6021 |
| COG | 0.473375 | 0.013392 | 35.3470 |
| IND0 | 0.102652 | 0.019880 | 5.16364 |
| IND1 | 0.051974 | 0.021676 | 2.39772 |
| IND2 | 0.047078 | 0.026230 | 1.79480 |
| IND3 | 0.168565 | 0.021056 | 8.00564 |
| IND4 | 0.164877 | 0.023614 | 6.98207 |
| IND5 | 0.019190 | 0.020750 | 0.924809 |
| IND6 | -0.138776 | 0.016772 | -8.27401 |
| IND7 | 0.008106 | 0.016658 | 0.486650 |
| IND8 | 0.062944 | 0.016233 | 3.87750 |
| UNION | 0.207882 | 0.013744 | 15.1254 |
| PTIME | -1.06144 | 0.010463 | -101.451 |

Note that for about forty percent of all the subsequent-job observations the beginning of the subsequent-job spell began before the first SIPP interview. In these cases, the only earnings information available was for the time just before the first SIPP interview. These mid-spell data were adjusted to represent start-of-spell earnings levels using the same procedure described above on page 172.

The fifth of the six subsequent-job-characteristics tables presents the multinomial logit parameters of employer firm size. Employer size is a categorical variable defined as follows: (1) 1–24 employees, (2) 25–99 employees, and (3) 100 or more employees. The parameters of the third size category are normalized to zero. The individual attribute and job characteristic variables are defined in the same way they are above. The LEARN variable denotes log-earnings at the start of the subsequent job. In the following table the number appended to each variable indicates the size category to which it corresponds.

The random variate correlation values are arbitrarily assigned (because the PSID data have no job information about firm size) to produce job-to-job correlations of about 0.5 for the 1–24 size category, 0.2 for the 24–99 category, and 0.7 for the 100+ category.

***Subsequent-Job Employer Firm Size***

*Observations = 13995; log-L = -11500.0*

| Parameter | Estimate | Std. Error | t-Statistic |
|---|---|---|---|
| C1 | -0.052154 | 0.230663 | -0.226106 |
| FEMALE1 | -0.222091 | 0.046502 | -4.77593 |
| AGE2B1 | -0.324071 | 0.049642 | -6.52814 |
| AGE41 | 0.149741 | 0.065124 | 2.29930 |
| AGE5A1 | 0.273105 | 0.083875 | 3.25610 |
| HSA1 | 0.059529 | 0.072778 | 0.817962 |
| COA1 | -0.274325 | 0.050713 | -5.40932 |
| COG1 | -0.301503 | 0.066213 | -4.55351 |
| IND01 | 2.41279 | 0.105798 | 22.8057 |
| IND11 | 0.705153 | 0.110290 | 6.39362 |
| IND21 | 0.893486 | 0.126465 | 7.06507 |
| IND31 | 0.511323 | 0.109241 | 4.68068 |
| IND41 | 0.865398 | 0.115133 | 7.51648 |
| IND51 | 1.17119 | 0.103081 | 11.3619 |
| IND61 | 1.30963 | 0.090918 | 14.4045 |
| IND71 | 1.28488 | 0.088494 | 14.5194 |
| IND81 | 1.70398 | 0.085018 | 20.0425 |
| UNION1 | -1.09770 | 0.075252 | -14.5870 |
| PTIME1 | 0.208314 | 0.062415 | 3.33758 |
| LEARN1 | -0.211804 | 0.030203 | -7.01278 |

*table continued on next page*

| ***Subsequent-Job Employer Firm Size*** *(continued)* | | | |
|---|---|---|---|
| Parameter | Estimate | Std. Error | t-Statistic |
| $\epsilon$ correl1 | 0.4 | *arbitrary* | |
| | | | |
| C2 | -1.56743 | 0.314435 | -4.98489 |
| FEMALE2 | -0.018210 | 0.061061 | -0.298219 |
| AGE2B2 | -0.212786 | 0.065687 | -3.23937 |
| AGE42 | 0.156458 | 0.085021 | 1.84021 |
| AGE5A2 | 0.126195 | 0.114919 | 1.09813 |
| HSA2 | 0.314756 | 0.094323 | 3.33700 |
| COA2 | -0.163500 | 0.068465 | -2.38808 |
| COG2 | -0.137916 | 0.086632 | -1.59197 |
| IND02 | 1.23627 | 0.136559 | 9.05303 |
| IND12 | 0.561440 | 0.122741 | 4.57419 |
| IND22 | 0.550762 | 0.148794 | 3.70151 |
| IND32 | 0.318616 | 0.122964 | 2.59113 |
| IND42 | -0.157905 | 0.166110 | -0.950604 |
| IND52 | 0.507334 | 0.132126 | 3.83977 |
| IND62 | 0.286616 | 0.121576 | 2.35751 |
| IND72 | 0.661763 | 0.106157 | 6.23382 |
| IND82 | 1.16540 | 0.096701 | 12.0516 |
| UNION2 | -0.300892 | 0.079592 | -3.78041 |
| PTIME2 | -0.028530 | 0.088006 | -0.324185 |
| LEARN2 | -0.068705 | 0.041751 | -1.64560 |
| $\epsilon$ correl2 | 0.8 | *arbitrary* | |
| | | | |
| $\epsilon$ correl3 | 0.6 | *arbitrary* | |

The sixth and final subsequent-job characteristics table begins on the next page.

The last of the six characteristics tables for subsequent jobs presents the binary logit parameters of employer pension sponsorship. Pension non-sponsorship is the category for which parameters are normalized to zero. The individual attribute and job characteristic variables are defined in the same way they are above. The FSIZE1 and FSIZE2 dummy variables indicate the 1–24 and the 25–99 employer size categories, with the largest of the three size categories being the omitted category.

***Subsequent-Job Employer Pension Sponsorship***

*Observations = 10232; log-L = -4079.02*

| Parameter | Estimate | Std. Error | t-Statistic |
|---|---|---|---|
| $\epsilon$ correl(non-sponsor) | 0.60 | *calibrated value* | |
| $\epsilon$ correl(sponsor) | 0.90 | *calibrated value* | |
| | | | |
| C | -0.120760 | 0.330437 | -0.365456 |
| C | -0.25 | *calibrated value* | |
| FEMALE | 0.069040 | 0.062377 | 1.10683 |
| AGE2B | 0.290390 | 0.065867 | 4.40874 |
| AGE4 | -0.225618 | 0.084375 | -2.67401 |
| AGE5A | -0.463128 | 0.111060 | -4.17008 |
| HSA | -0.275555 | 0.095391 | -2.88870 |
| COA | 0.281677 | 0.067466 | 4.17509 |
| COG | 0.355287 | 0.087751 | 4.04880 |
| IND0 | -2.27088 | 0.151321 | -15.0070 |
| IND1 | -1.17732 | 0.148546 | -7.92564 |
| IND2 | -0.997243 | 0.170707 | -5.84185 |
| IND3 | -0.950064 | 0.148910 | -6.38014 |
| IND4 | -1.46376 | 0.156594 | -9.34745 |
| IND5 | -1.31418 | 0.145273 | -9.04626 |
| IND6 | -2.23088 | 0.130270 | -17.1250 |
| IND7 | -1.79079 | 0.126732 | -14.1306 |
| IND8 | -1.20897 | 0.123921 | -9.75600 |
| UNION | 0.918042 | 0.098644 | 9.30659 |
| PTIME | -0.147942 | 0.086895 | -1.70254 |
| LEARN | 0.443643 | 0.043385 | 10.2257 |
| FSIZE1 | -2.05834 | 0.062234 | -33.0744 |
| FSIZE2 | -1.12426 | 0.082147 | -13.6860 |

## C.11 Subsequent-Job Duration

The following table contains estimated parameters for the proportional hazard model of subsequent-job duration. The covariates are expressed using the same names as described above in the presentation of the subsequent-job characteristics equations (see Section C.10 in this appendix), while the hazard function parameters and error terms are denoted in the same way as described above in the presentation on the non-employment duration model (see Section C.9 in this appendix).

As described in Appendix B, Section B.19.6, the variances of both the individual-specific and spell-specific error terms and the age-specific constant terms of the subsequent-job duration hazard function are calibrated to produce mean duration by age group, standard deviation of duration, and correlation of duration statistics that closely match those tabulated from over 20 years of longitudinal PSID data. The calibration target statistics from the PSID are: 8.9, 11.6, 12.7, and 6.3 years mean completed duration for those aged 14–29, 30–39, 40–49, and 50–64 when their job began, respectively; 9.6 years for the standard deviation of duration (all-age mean is about 9.9); and 0.1 for the correlation between the duration of the current and prior jobs. The following table includes calibrated values of the hazard function parameters that come close to matching those target statistics.

The table begins at the top of the next page.

| ***Subsequent-Job Duration*** | | | |
|---|---|---|---|
| *Observations = 10232; log-L = -11737.8* | | | |
| *Mean duration in estimation sample = 9.96 years* | | | |
| Parameter | Estimate | Std. Error | t-Statistic |
| S_VAR | 0.815702 | 0.176528 | 4.62081 |
| S_VAR | 0.00 | *calibrated value* | |
| I_VAR | 0.219685 | 0.132458 | 1.65853 |
| I_VAR | 0.32 | *calibrated value* | |
| B_CONST | -4.02206 | 0.253493 | -15.8665 |
| B_CONST | -3.68 | *calibrated value* | |
| FEMALE | 0.062291 | 0.045598 | 1.36610 |
| AGE2B | -0.035570 | 0.049542 | -0.717987 |
| AGE2B | 0.46 | *calibrated value* | |
| AGE4 | 0.164870 | 0.069750 | 2.36373 |
| AGE4 | -0.66 | *calibrated value* | |
| AGE5A | 0.715115 | 0.092784 | 7.70729 |
| AGE5A | -0.79 | *calibrated value* | |
| HSA | 0.193861 | 0.071491 | 2.71168 |
| COA | 0.188881 | 0.051687 | 3.65431 |
| COG | 0.414842 | 0.064326 | 6.44908 |
| IND0 | 0.941534 | 0.099257 | 9.48583 |
| IND1 | 0.285675 | 0.098743 | 2.89311 |
| IND2 | 0.372189 | 0.113917 | 3.26720 |
| IND3 | 0.180939 | 0.095725 | 1.89019 |
| IND4 | 0.236679 | 0.107332 | 2.20512 |
| IND5 | 0.459658 | 0.100314 | 4.58220 |
| IND6 | 0.879064 | 0.084392 | 10.4164 |
| IND7 | 0.679962 | 0.080624 | 8.43377 |
| IND8 | 0.335913 | 0.079337 | 4.23402 |
| UNION | -0.531991 | 0.065143 | -8.16648 |
| PTIME | 0.929116 | 0.063251 | 14.6894 |
| LEARN | 0.079637 | 0.031436 | 2.53336 |
| FSIZE1 | -0.048653 | 0.052295 | -0.930338 |
| FSIZE2 | 0.151846 | 0.067100 | 2.26299 |
| PENSION | -0.586494 | 0.052046 | -11.2689 |
| ALPHAM1 | 0.110754 | 0.042196 | 2.62476 |

## C.12 On-the-Job Earnings Adjustments

The parameters assumed in the PENSIM validation run for the earnings adjustment process are shown in the following table.

***Parameters of On-the-Job Earnings Adjustment Process***

| Parameter | PENSIM value |
|---|---|
| *Conversion of No-Growth to Nominal Earnings:* | |
| Inflation rate | 0.0 |
| Real wage growth rate | 0.0 |
| *Adjustments to No-Growth Earnings:* | |
| Annual non-stochastic drift term | 0.005 |
| Std. deviation of annual permanent-deviation shock | 0.015 |
| Std. deviation of annual temporary-deviation shock | 0.060 |

The standard deviations of the permanent deviation and temporary deviation shocks are constrained to be in a four-to-one ratio. This assumption is approximately consistent with the findings for men of Harris and Sabelhaus (2003). The level of the standard deviations is much lower than those estimated by Harris and Sabelhaus (2003) for CBO's CBOLT model because the PENSIM first-job and subsequent-job earnings equations explain more cross-sectional variation in earnings than does the mean earnings equation used in CBOLT.

The level of the standard deviations has been set so that simulated earnings exhibit the correct level of earnings inequality (as measured by the Gini coefficient) at each age of the simulated cohort. The calibrated levels of the earnings shock variances produce the rising degree of earnings inequality shown in longitudinal data and in repeated-cross-sectional data.

## C.13 Job Industry Translation

This section describes the translation from one-digit SIPP industry to one-digit SIC industry. As part of simulating the first-job-start and subsequent-job-start events, PENSIM assigns an individual six job and employer characteristics (see Appendix B, Section B.17.1 and Section B.19.1, respectively). The first characteristic to be assigned is employer industry, which is one of ten one-digit SIPP industry classes. After the SIPP industry class is assigned, it is used as an explanatory variable in the other five equations, the final one of which assigns the employer's pension sponsorship status.

If the employer sponsors a pension, PENSIM uses the pension characteristics imputation model (Holmer and Janney 2003) to assign the pension offering and detailed characteristics of each offered pension plan (see Section B.17.2 and Section B.17.3 for first jobs, and Section B.19.2 and Section B.19.3 for subsequent jobs). Many of the equations in the pension characteristics imputation model use the one-digit SIC industry as an explanatory factor. As these two industry codes are not identical, the SIPP industry code has to be translated into an SIC industry code. This translation involves the use of both deductive logic and two estimated probability models.

The table shows that eight of the SIPP industry codes translate directly into SIC industry codes. Four SIPP industry codes (1, 4, 8, and 9) have a one-to-one match with SIC industry codes. Two pairs of SIPP industry codes (2-3 and 5-6) translate into a single SIC industry code.

***One-Digit SIPP and SIC Industry Codes***

| SIPP Code | SIC Code | Industry Description |
|---|---|---|
| 0 | 0 | Agriculture |
| | 1 | Mining & construction |
| 1 | 2 | Product manufacturing |
| 2–3 | 3 | Equipment manufacturing |
| 4 | 4 | Transportation & communications |
| 5–6 | 5 | Wholesale & retail trade |
| 7 | 6 | Finance, insurance, & real estate |
| | 7 | Personal services |
| 8 | 8 | Professional services |
| 9 | 9 | Government |

The remaining two SIPP industry codes, 0 and 7, are more of a problem;

both of these must be split between two SIC industry codes. To develop a rational rule for making this split, we estimate a binomial logit probability model for each. The data for the models come from SIPP household interview data. The SIC industry code has been added to each record having SIPP industry code 0 or 7. According to the documentation available at the SIPP web site, the two industry codes are related as follows:

SIPP codes 000–039 correspond to SIC 0,
SIPP codes 040–099 correspond to SIC 1,
SIPP codes 700–720 correspond to SIC 6, and
SIPP codes 721–799 correspond to SIC 7.

For each probability model the independent effects included the other five job and employer characteristics: union status, work schedule (PTIME), log of job earnings, firm size, and whether the employer sponsors a pension plan (SPONSORS). Three individual attributes — gender, age, and education level — were also included as explanatory variables in each probability model and tested for significance.

The first probability model predicts the SIC code using all records with SIPP industry code 0. SIC industry code 0 is the omitted level of the dependent variable, so the model predicts the probability that the SIC industry code is 1 when the SIPP industry code is 0. The first table on the next page shows the estimation results. All of the effects that are actually levels of a single factor, such as FSIZE1 (small-sized firm) and FSIZE2 (medium-sized firm), were tested as a group and were found to be significant. All the rest of the effects are individually significant. All factors, such as age, that do not appear are not significant.

The second table on the next page shows results for the probability model that model that is estimated using all records with SIPP industry code 7. SIC industry code 6 is the omitted level of the dependent variable, so the model predicts the probability that the SIC industry code is 7 when the SIPP industry code is 7. Again, all of the effects are either individually significant or significant as part of a group. Age is significant in this model, but firm size and work schedule (PTIME) are not.

We have no a priori expectations regarding the signs or magnitudes of the estimated coefficients of these two probability models. It was reasonable, however, to expect that some of the various job and individual characteristics would be coincidentally associated with the alternative SIC industries and could provide a method for accurately assigning employees to SIC industries.

***Probability that a Job is in SIC Industry 1 when its SIPP Industry Code is 0***

*Observations = 2311; Log-L = -879*

| Parameter | Estimate | Std. Error | p-Value |
|---|---|---|---|
| C | -3.585 | 0.631 | 0.000 |
| FEMALE | -1.112 | 0.154 | 0.000 |
| LEARN | 0.767 | 0.091 | 0.000 |
| SPONSORS | 0.444 | 0.158 | 0.005 |
| UNION | 0.410 | 0.215 | 0.057 |
| PTIME | -0.540 | 0.195 | 0.006 |
| HSA | -0.588 | 0.153 | 0.000 |
| COA | -0.126 | 0.163 | 0.441 |
| COG | -0.841 | 0.205 | 0.000 |
| FSIZE1 | 0.171 | 0.157 | 0.278 |
| FSIZE2 | 0.444 | 0.209 | 0.034 |

***Probability that a Job is in SIC Industry 7 when its SIPP Industry Code is 7***

*Observations = 4840; Log-L = -2692*

| Parameter | Estimate | Std. Error | p-Value |
|---|---|---|---|
| C | 6.122 | 0.377 | 0.000 |
| FEMALE | -0.990 | 0.072 | 0.000 |
| LEARN | -0.640 | 0.049 | 0.000 |
| SPONSORS | -0.816 | 0.071 | 0.000 |
| UNION | 0.631 | 0.164 | 0.000 |
| AGE | -0.006 | 0.003 | 0.036 |
| HSA | 0.924 | 0.134 | 0.000 |
| COA | -0.084 | 0.083 | 0.308 |
| COG | -0.214 | 0.091 | 0.019 |

## C.14 Pension Plan Offerings and Characteristics

PENSIM includes a pension characteristics imputation model that imputes the number and kind of employer pension plan offerings as well as the detailed characteristics of each offered plan. The imputation model has been estimated with 1996–98 job-level data from the National Compensation Survey (formerly the Employee Benefits Survey) conducted by the Bureau of Labor Statistics. There is a separate document (Holmer and Janney 2003) that describes the PENSIM pension characteristics imputation model in detail. A short description of the imputation model is provided in the next paragraph.

National Compensation Survey (formerly Employee Benefits Survey) data for 1996–98 (BLS 1999a, BLS 1999b, BLS 2000b) are used to estimate a recursive system of statistical equations that allow the stochastic imputation of the characteristics of pensions sponsored by each individual's employer. The covariates in these equations include job characteristics (industry, firm size, unionization, part/full-time), but not the individual's personal attributes (gender, age, and education). The simulated correlation between individual attributes and sponsored pension characteristics are induced indirectly through the interaction of two sets of correlations: (1) the correlations between attributes and job characteristics, which are present in job characteristics equations, and (2) the correlations between job characteristics and pension characteristics that are embedded in the recursive system of pension-characteristic equations.

There is one known problem with the imputation model. Because providing a defined-contribution thrift plan with no matching employer contribution generates no employee compensation, BLS does not consider such a plan an employee benefit. Given that view, such plans were not included in the raw BLS pension data used to estimate the imputation model. In the pension community, however, such plans are considered pensions and are included in statistical tabulations based on Form 5500 data and government-sponsored household surveys.

Published tabulations of 1996-98 data from the Employee Benefit Survey (BLS 1999a, BLS 1999b, BLS 2000b), which has since been integrated into the National Compensation Survey, show that the percent of full-time private-sector employees who participated in a no-employer-matching thrift plan was 6.5 percent, while the percent of full-time state-and-local government employees who participated in a no-employer-matching thrift plan was

21.9 percent. Given the number of full-time employees in each sector, these percents translate into 5.11 million in the private sector and 3.14 million in the state-and-local government sector. We have no idea from these published tabulations how many full-time employees were offered no-employer-matching thrift plans by their employer (but declined to participate), and we know nothing about the offering and participation experience of part-time employees. Moreover, we have no concrete information about how often such plans are paired with other kinds of pension plans offered by the employer.

## C.15 Pension Participation Behavior

This section describes (a) the defined-contribution participation probability function, (b) the hazard function that determines the waiting time between the eligibility event and the participation event when the plan has standard-enrollment procedures, and (c) the hazard function that determines the waiting time between the participation event and the active-participation event when the plan has automatic-enrollment procedures.

**Participation Probability.** The participation probability is generated by a binary logit equation that is estimated using tabular data on participation in 2001 reported by Hewitt and used by the Government Accountability Office in a recent study of pensions (GAO 2003, page 45). The covariates in the equation are age, relative earnings (annualized current earnings divided by the SSA average wage index in that year), and several interactions of age and relative earnings. The omitted age category is 60+ and the omitted relative earnings category is the top category.

In addition to these estimated coefficients, three additional coefficients have been added based on a recent review of participation rate studies (Munnell et al. 2002). Each of these coefficients have been set to a value that produces a three percentage point decline in the average predicted participation rate. The constant coefficient has been increased from 3.22051 to maintain the same average participation rate as observed when these three coefficients are zero. The three coefficients represent the effect on participation of the defined-contribution plan being offered along with a defined-benefits plan (BOTH), the participation effect of the defined-contribution plan having no employer matching of employee contributions (NOMATCH), and the effect of the interaction of these two variables.

The estimated coefficients of the participation logit equation are in the following table, where the relative earnings categories are defined using percentage ratios.

***Probability of Defined-Contribution Pension Participation***

| Parameter | Estimate |
|---|---|
| C | 3.48000 |
| C_DEL_AE | see text |
| AGE<20 | -2.12326 |
| AGE20-29 | -0.20591 |

*table continued on next page*

***Probability of Defined-Contribution Pension Participation***
*(continued)*

| Parameter | Estimate |
|---|---|
| AGE30-39 | 0.36015 |
| AGE40-49 | 0.39182 |
| AGE50-59 | 0.50893 |
| REARN<59 | -2.71795 |
| REARN59-118 | -1.27301 |
| REARN118-177 | -1.70877 |
| REARN177-236 | -0.83155 |
| REARN236-295 | -0.36750 |
| AGE20-29*REARN<118 | -1.06428 |
| AGE30-39*REARN<118 | -1.00381 |
| AGE40-49*REARN<118 | -0.93602 |
| AGE50-59*REARN<118 | -0.57443 |
| BOTH | -0.13300 |
| NOMATCH | -0.13300 |
| BOTH*NOMATCH | -0.13300 |

The C_DEL_AE parameter in the participation probability equation specifies the impact of automatic-enrollment procedures on the participation rate. The C_DEL_AE parameter is coefficient of a dummy variable whose value is one when the defined-contribution plan has automatic-enrollment procedures and zero when the plan has standard-enrollment procedures. Its value is irrelevant to most PENSIM runs because there are no defined-contribution plans with automatic-enrollment procedures. The baseline pension characteristics imputation model is estimated using data from 1996–1998, before BLS started gathering information about that plan characteristic. The imputation model has been enhanced so that the probability of a plan having automatic-enrollment procedures can be greater than zero.

**Eligibility-to-Participation Waiting Time.** It is assumed that, under standard-enrollment procedures, even if an individual is simulated to participate in the defined-contribution plan, there is a waiting time between the eligibility event and the participation event. (Of course, under automatic-enrollment procedures, there is no waiting time between the eligibility and participation events.) The standard-enrollment waiting time is generated by a piecewise log-linear hazard function with no covariates. The hazard function parameters are specified so that the simulated waiting time distribution is similar to those described by Choi, Laibson, and Madrian (2004). The calibrated coefficients of the two-piece log-linear hazard function produce a high hazard rate during the first 0.3 year after the eligibility event and then a lower hazard rate after that initial period. The calibrated coefficients are shown in the following table.

***Piecewise Log-Linear Hazard Function for Eligibility-to-Participation Waiting Time***

| Parameter | Estimate |
|---|---|
| A1 | 0.693 |
| S1 | 0.300 |
| A2 | -0.693 |

**Participation-to-Active-Participation Waiting Time.** It is assumed that, under automatic-enrollment procedures, there is a waiting time between the participation event and the active-participation event. During this waiting time the participant's contribution and investment decisions are determined by the automatic-enrollment default options. (Of course, under standard-enrollment procedures, there is no waiting time between the participation and active-participation events.) The automatic-enrollment waiting time is generated by a piecewise log-linear hazard function. The hazard function parameters are specified so that the simulated waiting time distribution is similar to those described by Choi, Laibson, and Madrian (2004). The calibrated coefficients of the two-piece log-linear hazard function produce a relatively high hazard rate during the first 0.3 year after the participation event and then a relatively lower hazard rate after that initial period. In addition, the hazard function has a term that is the product of a coefficient and the individual's simulated participation probability. This term is added to both the piecewise linear coefficients. This means that individuals with high participation probabilities become active participants faster than individuals with low participation probabilities. The calibrated coefficients are shown in the following table.

| ***Piecewise Log-Linear Hazard Function for Participation-to-Active-Participation Waiting Time*** | |
|---|---|
| Parameter | Estimate |
| A1 | -2.25 |
| S1 | 0.30 |
| A2 | -4.10 |
| PARTPROB | 3.00 |

## C.16 Pension Contribution Behavior

This section describes how before-tax employee contributions to defined-contribution plans are simulated. PENSIM does not simulate contributions of after-tax employee contributions to defined-contribution plans.

The method currently used by PENSIM to simulate employee contributions is rudimentary, largely because of the lack of quantitative research on the dynamics of employee contribution behavior. The fraction of earnings contributed, the contribution rate, is assigned each year using a stochastic process that produces rate variation across age-earnings groups, variation across individuals within age-earnings groups, and variation across years for the same individual.

An active participant in a defined-contribution plan always contributes a positive amount to the plan in the first year of active participation. In subsequent years of active participation, the individual moves between a zero-contribution state and a positive-contribution state according to Markov transition probabilities. In the current version of PENSIM, the ergodic probability of being in the zero-contribution state is assumed to be eight percent (Holden and VanDerhei 2005, footnote 21) and the annual transition probability between the zero-contribution state and the positive-contribution state is assumed to be fifty percent.

For an active participant who is in the positive-contribution state, the desired employee contribution rate is drawn from a lognormal distribution. The mean of this lognormal distribution is the sum of two terms: the logarithm of the decimal average rate for the age-earnings group of the individual that year (denoted by $avg$), and a lifetime deviation term for the individual that is drawn from a normal distribution (denoted by $dev$) that does not vary from year to year. The positive contribution rate assigned to the individual is $exp(avg + dev)$, where the positive rate is rounded to the nearest tenth of a percent.

In the current version of PENSIM, the mean and standard deviation of $dev$ are assumed to be 0.03 and 0.90, respectively. The average rate for each of thirty age-earnings groups is shown in the following table, which is based largely on data from Holden and VanDerhei (2001) as described by GAO (2003, page 45). Note that the table transforms the original dollar earnings categories into relative earnings categories (defined relative to the SSA average wage index), that relative earnings values are expressed in percentage terms, and that the contribution rate are expressed in percentage terms before applying the logarithmic transformation.

***Mean Before-Tax Employee Contribution Rates to Defined-Contribution Plans***

| Age | Relative Earnings (see text) | Rate |
|---|---|---:|
| <=29 | <66 | 5.1 |
| <=29 | 66–131 | 5.3 |
| <=29 | 131–197 | 6.8 |
| <=29 | 197–263 | 7.4 |
| <=29 | 263–328 | 6.8 |
| <=29 | 328+ | 4.8 |
| 30–39 | <66 | 6.4 |
| 30–39 | 66–131 | 6.2 |
| 30–39 | 131–197 | 6.8 |
| 30–39 | 197–263 | 7.2 |
| 30–39 | 263–328 | 6.9 |
| 30–39 | 328+ | 5.1 |
| 40–49 | <66 | 6.9 |
| 40–49 | 66–131 | 6.7 |
| 40–49 | 131–197 | 7.1 |
| 40–49 | 197–263 | 7.3 |
| 40–49 | 263–328 | 6.8 |
| 40–49 | 328+ | 5.0 |
| 50–59 | <66 | 7.8 |
| 50–59 | 66–131 | 7.6 |
| 50–59 | 131–197 | 8.3 |
| 50–59 | 197–263 | 8.2 |
| 50–59 | 263–328 | 7.3 |
| 50–59 | 328+ | 5.1 |
| 60+ | <66 | 9.0 |
| 60+ | 66–131 | 8.5 |
| 60+ | 131–197 | 9.3 |
| 60+ | 197–263 | 9.0 |

*table continued on next page*

***Mean Before-Tax Employee Contribution Rates to Defined-Contribution Plans***
*(continued)*

| Age | Relative Earnings (see text) | Rate |
|---|---|---|
| 60+ | 263–328 | 7.9 |
| 60+ | 328+ | 5.1 |

These assumptions about the distribution of positive contribution rates for active participants produces a distribution of contribution rates that is similar to the distribution of observed rates (Holden and VanDerhei 2001, page 4).

PENSIM uses this process to assign a contribution rate to each participant who is not passively accepting the default contribution rate associated with any automatic-enrollment procedures in the plan. The assigned contribution rate does not depend on the plan match rate or match up-to level, which reflects the ambiguous findings of past research (Holden and VanDerhei 2001, page 15). Also, the assigned contribution rate is not adjusted for individuals whose plans have no employer matching because there is evidence that the contribution rates are roughly the same for those who make contributions to no-employer-matching plans (Holden and VanDerhei 2001, page 12).

## C.17 Pension Investment Behavior

This section describes how the allocation of assets in defined-contribution plan accounts is simulated. Individuals are assumed to rebalance their account holdings each year and invest the total account balance in either Treasury bonds or corporate equities (represented as the S&P 500 index of U.S. stocks). The fraction of assets invested in bonds and equities can vary by age and gender. The user of PENSIM is responsible for setting the values of these age- and gender-specific asset allocation fractions when specifying a PENSIM run.

**Company Stock Returns.** Company stock, which is the employer contribution in some defined-contribution plans, is assumed to have a rate of return equal to that of this broad corporate equity index plus an annual random element that is drawn from a normal distribution with a mean of zero and a standard deviation of thirty-one percent, an assumption based on 1962–1995 results on the returns of individual stocks included in the S&P 500 index (Ikenberry et al. 1998, Table 1).

## C.18 Pension Rollover Behavior

This section describes how PENSIM simulates the rollover of pension account balances at the end of a job. If the job ends with the disability, retirement, or death event, all pension account balances are rolled over into a personal rollover account and are withdrawn, sooner or later, to generate retirement income in some manner. In all other cases, there is some probability that the end-of-job pension balance will not be rolled over into the rollover account, but will be used for some purpose other than saving for retirement. If the balance is not rolled over (that is, if it is cashed out), it disappears completely from the simulation.

The probability of rolling over the pension balance is assumed to rise with size of the pension balance as shown in Fidelity administrative data used by the Government Accountability Office in a prior pension study (GAO 2003, page 46). In order to permit the rollover probability equation to be used in all simulation years, the pension account balance is transformed from nominal dollars to a relative balance, defined as the nominal balance divided by the SSA average wage index (AWI) in that year.

In PENSIM, very large balances (that is, more than ten times AWI) are assumed to be always rolled over. The probability for other balances is calculated using a logit equation that has a constant and the natural logarithm of the relative balance as the only covariate.

The constant and log-relative-balance coefficients are estimated by calibrating those coefficient so that PENSIM replicates Health and Retirement Survey (HRS) findings on pension rollovers: a person cash-out rate of 20 percent and a dollar cash-out rate of 9.3 percent among those more the fifty years old at job end (Hurd and Panis 2006?). The calibration, which is conducted using a two percent sample of the 1970 birth cohort, produces a constant coefficient of 1.2800 and a log-relative-balance coefficient of 0.5625.

Using these calibrated coefficients and the two percent sample of those born in 1970, the simulated incidence of not rolling over (that is, cashing out) pension account balances at job endings that occurred at age fifty or more is 19.9 percent person-weighted and 9.1 percent dollar-weighted. In this same simulation, the incidence of cashing out pension account balances is higher among those ending a job at any age because at ages before fifty there is a higher likelihood of a small relative balance. The simulated incidence of cashing out at all ages is 34.4 percent person-weighted and 12.9 percent dollar-weighted.

## C.19 Pension Withdrawal Behavior

This section describes (a) the timing of claims of defined pension benefits, (b) the timing of first withdrawal of funds in the rollover account, and (c) the use made of rollover account withdrawals.

**Claiming of Traditional Defined Pension Benefits.** Traditional pensions (that is, defined-benefit plans that are not of the cash-account type) earned on prior jobs are claimed by disabled, widowed, and retired individuals at their earliest availability, which is determined by the early retirement age of each plan. Employed individuals are assumed to wait to claim defined-benefit pensions earned on prior jobs until there is no early retirement reduction in those pensions.

**Timing of First Account Withdrawal.** All pension account balances that are preserved for retirement income (rather than being cashed-out) at the end of jobs are transferred to a personal rollover account. Hence, withdrawals are made only from that rollover account. The first withdrawal from this rollover account is not made until the individual is older than the first non-penalty withdrawal age, which is established by government policy, and the individual is either disabled, widowed, or retired.

**Use of Account Withdrawals.** An individual making a rollover account withdrawal can use the money to buy an annuity or to directly support consumption needs in the year of the withdrawal. In PENSIM, the model user specifies what percent of the account balance at the first account withdrawal is used to purchase an annuity. The remaining balance, if any, continues to be invested in the rollover account and provides a balance against which to make annual withdrawals.

If the whole rollover account balance is not used to purchase an annuity, the fraction of the rollover balance withdrawn each year is a function of the minimum withdrawal fraction specified for that age by federal government regulations (Holmer 2003, page 10).

If part or all of the rollover account balance is used to purchase an annuity, then price of that annuity is calculated using assumed annuity loading factors for men and women and assumed annuity features, including whether the annuity payments are indexed for inflation and the size of the survivor payment in a joint-and-survivor annuity. The unloaded annuity price is calculated using mortality information used in PENSIM (which is imported from SSASIM via the environment file) and standard actuarial methods. The model user sets gender-specific loading factors that ensure the solvency of the annuity provider, as measured by the annuity provider revenue and expenditure statistics contained in the PENSIM .arc output results file.

## C.20 Retirement Timing

Beginning at age 62 each individual considers retiring, by which is meant withdrawing from employment (that is, quitting a job or quitting looking for a job) and claiming a social security retirement benefit. At each age between 62 and 68, the individual decides whether or not to defer retirement another year. An individual always defers retirement at an age that is less than the social security early retirement age. For an individual whose age is greater than or equal to the social security early retirement age, there is a probability of deferring retirement. In the current version of PENSIM, the generosity of pensions on the current job or past jobs does not affect the timing of retirement.

The retirement deferral probabilities used in the PENSIM retirement-timing algorithm are flexible data inputs that can be changed when simulating substantially different social security or pension environments. Using these reduced-form probabilities eliminates the need to include a complex retirement-timing model in PENSIM, while maintaining the flexibility of changing the age pattern of retirement in runs that assume significant shifts in social security policy or employer-sponsored pension plan designs.

The following table presents the retirement deferral probabilities, which can vary by birth cohort, age, and gender, used in PENSIM to simulate current social security policy and current pension plan design. These probabilities produce the age pattern of social security retirement benefit claiming observed in recent data published in the *Annual Statistical Supplement to the Social Security Bulletin*.

***Retirement Deferral Probabilities***

| Year Born | Age | Female | Male |
|---|---|---|---|
| <=1938 | 62 | 0.40 | 0.43 |
| <=1938 | 63 | 0.80 | 0.85 |
| <=1938 | 64 | 0.70 | 0.70 |
| <=1938 | 65 | 0.70 | 0.70 |
| <=1938 | 66 | 0.70 | 0.70 |
| <=1938 | 67 | 0.70 | 0.70 |
| <=1938 | 68 | 0.70 | 0.70 |
| <=1938 | 69 | 0.00 | 0.00 |
| | | | |
| >=1939 | 62 | 0.50 | 0.53 |

*table continued on next page*

***Retirement Deferral Probabilities***

*(continued)*

| Year Born | Age | Female | Male |
|---|---|---|---|
| >=1939 | 63 | 0.80 | 0.85 |
| >=1939 | 64 | 0.70 | 0.70 |
| >=1939 | 65 | 0.70 | 0.70 |
| >=1939 | 66 | 0.70 | 0.70 |
| >=1939 | 67 | 0.70 | 0.70 |
| >=1939 | 68 | 0.70 | 0.70 |
| >=1939 | 69 | 0.00 | 0.00 |

# References


Abowd, John M. and Card, David. "On the Covariance Structure of Earnings and Hours Changes." *Econometrica*, 1989, *57*(March), pp. 411–445.

Aho, Alfred V.; Kernighan, Brian W.; and Weinberger, Peter J. *The AWK Programming Language*. Reading, MA: Addison-Wesley, 1988.

Bell, Felicitie C.; Wade, Alice H.; and Goss, Stephen C. *Life Tables for the United States Social Security Area 1900–2080*. Actuarial Study No. 107. SSA Pub. No. 11-11536. Baltimore, MD: Social Security Administration, August 1992.

Benítez-Silva, Hugo. "A Dynamic Model of Labor Supply, Consumption/Saving, and Annuity Decisions Under Uncertainty." SUNY-Stony Brook working paper, September 2000.
⟨http://www.econ.yale.edu/ hugo/dynpn.pdf⟩

Brooks, Frederick P. *The Mythical Man-Month: Essays on Software Engineering, 20th Anniversary Edition*. Reading, MA: Addison-Wesley, 1995.

Bureau of Labor Statistics. *Employee Benefits in Small Private Establishments, 1996*. Bulletin 2507. Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics, April 1999.

Bureau of Labor Statistics. *Employee Benefits in Medium and Large Private Establishments, 1997*. Bulletin 2517. Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics, September 1999.

Bureau of Labor Statistics. "Employee Tenure in 2000" Press Release 00-245, August 29, 2000a.

Bureau of Labor Statistics. *Employee Benefits in State and Local Governments, 1998*. Bulletin 2531. Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics, December 2000b.

Carroll, Christopher D. "The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence." *Brookings Papers on Economic Activity*, 1992(2):61–135.

CBO. *The Outlook for Social Security*. Washington, DC: Congressional Budget Office, June 2004.

Centers for Disease Control and Prevention, *Morbidity and Mortality Weekly Report*, "State-Specific Variation in Rates of Twin Births – United States, 1992-1994," February 14, 1997, Vol. 46, No. 6.

Choi, James J.; Laibson, David; and Madrian, Bridgitte C. "Plan Design and 401(k) Savings Outcomes" written for the *National Tax Journal* Forum on Pensions, June 2004.
⟨http://www.polsim.com/PENSIM/plandesign.pdf⟩

Citro, Constance F. and Hanushek, Eric A., eds. *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling*, *Volume I: Review and Recommendations*. Washington, DC: National Academy Press, 1991.

Citro, Constance F. and Hanushek, Eric A., eds. *Assessing Policies for Retirement Income: Needs for Data, Research, and Models*. Washington, DC: National Academy Press, 1997.

Cohen, Michael L. "Statistical Matching and Microsimulation Models," in Constance F. Citro and Eric A. Hanushek, eds., *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling*, *Volume II: Technical Papers*. Washington, DC: National Academy Press, 1991.

Copeland, Craig. "Retirement Plan Participation and Retirees' Perception of Their Standard of Living." *EBRI Issue Brief No. 289*. Washington, DC: Employee Benefit Research Institute, January 2006.

Coplien, James O. *Advanced C++ Programming Styles and Idioms*. Reading, MA: Addison-Wesley, 1992.

Cotton, Paul and Sadowsky, George. "Future Computing Environments for Microsimulation Modeling," in Constance F. Citro and Eric A. Hanushek, eds., *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling*, *Volume II: Technical Papers*. Washington, DC: National Academy Press, 1991, pp. 141–234.

Date, C. J. *An Introduction to Database Systems, 7th edition*. Reading, MA: Addison-Wesley, 1999.

Deaton, Angus and Paxson, Christina. "Intertemporal Choice and Inequality." *Journal of Political Economy*, 1994, *102*(3), pp. 437–467.

Dickens, Richard. "The Evolution of Individual Male Earnings in Great Britain: 1975–95." *Economic Journal*, 2000, *110*(January), pp. 27–49.

Fidelity Institutional Retirement Services Company. *Building Futures, Volume VI.* Available at ⟨http://buildingfutures.fidelity.com⟩.

Fishman, George S. *Principles of Discrete Event Simulation.* New York: John Wiley & Sons, 1978.

GAO. *Retirement Income: Intergenerational Comparisons of Wealth and Future Income.* GAO-03-429. Washington, DC: Government Accountability Office, April 2003.

GAO. *Private Pensions: Information on Cash Balance Pension Plans.* GAO-06-42. Washington, DC: Government Accountability Office, October 2005.

Gourinchas, Pierre-Olivier and Parker, Jonathan A. "Consumption Over the Life Cycle." *NBER Working Paper 7271.* Cambridge, MA: National Bureau of Economic Research, July 1999.

Hammersley, J.M. and Handscomb, D.C. *Monte Carlo Methods.* London: Chapman and Hall, 1964.

Hanushek, Eric A. and Maritato, Nancy L., eds. *Assessing Knowledge of Retirement Behavior.* Washington, DC: National Academy Press, 1996.

Harris, Amy Rehder and Sabelhaus, John. "Projecting Longitudinal Earnings Patterns for Long-Run Policy Analysis." Technical Paper 2003-2. Washington, DC: Congressional Budget Office, April 2003.

Holden, Sarah and VanDerhei, Jack. "Contribution Behavior of 401(k) Plan Participants." *ICI Perspective Vol. 7, No. 4.* Washington, DC: Investment Company Institute, October 2001.

Holden, Sarah and VanDerhei, Jack. “The Influence of Automatic Enrollment, Catch-Up, and IRA Contributions on 401(k) Accumulations at Retirement.” *EBRI Issue Brief No. 283*. Washington, DC: Employee Benefit Research Institute, July 2005.

Holmer, Martin R. “The Asset/Liability Management Strategy System at Fannie Mae.” *Interfaces*, May-June 1994.

Holmer, Martin R. *Validating Simulated Earnings Histories*. Washington, DC: Policy Simulation Group, September 2000. ⟨http://www.polsim.com/ehvalid.pdf⟩

Holmer, Martin R. *The Value of Social Security Disability Insurance*. Issue Paper #2001–09. Washington, DC: AARP Public Policy Institute, June 2001.

Holmer, Martin R. “Simulation Analysis of the Decision to Annuitize Pension Balances.” Washington, DC: Policy Simulation Group, September 2003.

Holmer, Martin R. “Decomposing Differences Between SSASIM OLC Analysis Run 154 and 2005TR Intermediate-Cost Projection.” Washington, DC: Policy Simulation Group, January 23, 2006. ⟨http://www.polsim.com/PENSIM/tr05v154.pdf⟩

Holmer, Martin R. *Introductory Guide to GEMINI*. Washington, DC: Policy Simulation Group, current version available at ⟨http://www.polsim.com/guide2.pdf⟩

Holmer, Martin R. *Introductory Guide to SSASIM*. Washington, DC: Policy Simulation Group, current version available at ⟨http://www.polsim.com/guide.pdf⟩

Holmer, Martin R. and Janney, Asa M. III. *Characteristics of Pension Plans in the United States, 1996–98*. Washington, DC: Policy Simulation Group, December 2003.

Hurd, Michael and Panis, Constantijn. “The Choice to Cash Out Pension Rights at Job Change or Retirement.” *Journal of Public Economics* vol(issue), pp-pp, forthcoming 2006?.

Ikenberry, David L.; Shockley, Richard L.; and Womack, Kent L. "Why Active Fund Managers Underperform the S&P 500: The Impact of Size and Skewness." *Journal of Private Portfolio Management*, Spring 1998, *1*(1), pp. 13–26.

Lancaster, Tony. *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press, 1990.

LifePaths homepage at Statistics Canada website.
⟨http://www.statcan.ca/english/spsd/LifePaths.htm⟩

Maddala, G.S. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press, 1983.

Marquis, M. Susan and Holmer, Martin R. "Alternative Models of Choice Under Uncertainty and Demand for Health Insurance." *Review of Economics and Statistics*, August 1996, *78*(3), pp. 421–427.

Munnell, Alicia H.; Sundén, Annika; and Taylor, Catherine. "What Determines 401(k) Participation and Contributions?" *Social Security Bulletin*, 2002, *64*(3), pp. 64–75.

Orcutt, Guy; Caldwell, Steven; Wertheimer, Richard; Franklin, S.; Hendericks, Gary; Peabody, G.; Smith, James; and Zedlewski, Sheila. *Policy Exploration Through Microanalytic Simulation*. Washington, DC: Urban Institute Press, 1976.

Panis, Constantijn and Lillard, Lee. *Final Report: Near Term Model Development: Part II*. Contract report to the Social Security Administration on MINT development. Santa Monica, CA: RAND, August 15, 1999.

Polisim homepage at Urban Institute website.
⟨http://polisim.urban.org⟩

Ross, Christine M. "DYNASIM2 and PRISM: Examples of Dynamic Modeling," in Constance F. Citro and Eric A. Hanushek, eds., *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling, Volume II: Technical Papers*. Washington, DC: National Academy Press, 1991, pp. 121–137.

Rust, John. "A Dynamic Programming Model of Retirement Behavior," in David Wise, ed., *The Economics of Aging.* Chicago: University of Chicago Press, 1989, pp. 359–398.

Rust, John. "Structural Estimation of Markov Decision Processes," in R. Engle and D. McFadden, eds., *Handbook of Econometrics, Volume IV.* Amsterdam: North-Holland, 1994, pp. 3082–3139.

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

Stern, Steven. "Simulation-Based Estimation." *Journal of Economic Literature*, December 1997, *35*(4), pp. 2006–2039.

Toder, Eric; Uccello, Cori; O'Hare, John; Favreault, Melissa; Ratcliffe, Caroline; Smith, Karen; Burtless, Gary; and Bosworth, Barry. *Final Report: Modeling Income in the Near Term — Projections of Retirement Income Through 2020 for the 1931–60 Birth Cohorts.* Contract report to the Social Security Administration on MINT development. Washington, DC: The Urban Institute, September, 1999.

Trustees of the Federal OASDI Trust Funds. *Trustees' Report on the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds.* Washington, DC: U.S. Government Printing Office, 2001.

Trustees of the Federal OASDI Trust Funds. *Trustees' Report on the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds.* Washington, DC: U.S. Government Printing Office, 2004.

Trustees of the Federal OASDI Trust Funds. *Trustees' Report on the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds.* Washington, DC: U.S. Government Printing Office, 2005.

Vanguard Center for Retirement Research. *How America Saves 2005: A Report on Vanguard 2004 Defined Contribution Plan Data.* Available at ⟨http://www.vanguardretirementresearch.com⟩.

Wolfson, Michael C. "Socio-Economic Statistics and Public Policy: A New Role For Microsimulation Modeling." Statistics Canada, Analytical Studies Branch (Socio-Economic Modeling Group), Research Paper No. 81,

July 1995.
⟨http://www.statcan.ca/english/research/
11F0019MIE/11F0019MIE1995081.pdf⟩.

Wolfson, Michael C.; Rowe, Geoff; Lin, X.; and Gribble, Steven F. "Historical Generational Accounting with Heterogeneous Populations." Statistics Canada, Analytical Studies Branch (Socio-Economic Modeling Group), working draft, February 11, 1997.

Wolfson, Michael C.; Rowe, Geoff; Lin, X.; and Gribble, Steven F. "Historical Generational Accounting with Heterogeneous Populations" in Miles Corak, editor, *Government Finances and Generational Equity*, Ottawa: Statistics Canada, 1998.

Zayatz, Tim. "Social Security Disability Insurance Program Worker Experience." *Actuarial Study No. 114*. Baltimore, MD: Social Security Administration, Office of the Chief Actuary, June 1999.