

ORES Working Paper Series

Number 91

Methods in Modeling Income
in the Near Term (MINT I)

Barbara A. Butrica, Howard M. Iams, James H. Moore, and Mikki D. Waid*

Division of Policy Evaluation

June 2001

Social Security Administration
Office of Policy
Office of Research, Evaluation, and Statistics

* Social Security Administration, Office of Policy
9th Floor, ITC Building, 500 E Street, SW, Washington, DC 20254-0001

Working Papers in this series are preliminary materials circulated for review and comment. The views expressed are the authors' and do not necessarily represent the position of the Social Security Administration. The papers have not been cleared for publication and should not be quoted without permission.

Introduction

The Social Security Administration (SSA), with substantial assistance from the Brookings Institution, RAND, and the Urban Institute, has developed a model that projects retirement income for current and future beneficiaries. The model is the result of SSA's project on Modeling Income in the Near Term (MINT I). In order to examine the Social Security benefits of individuals and couples in the baby boom retiring over the next 22 years, the model projects their retirement income as new beneficiaries at the age when they first receive benefits. It also projects the income of the beneficiary population in 2020. The MINT I model makes independent projections for each retiree's income from Social Security benefits, pensions, assets, and earnings (for working beneficiaries) using the Census Bureau's Survey of Income and Program Participation (SIPP) data for 1990 through 1993 matched with SSA administrative records for earnings in 1951 through 1996, benefits, and death.

The policy universe in the MINT I model is the baby-boom generation expected to receive Social Security retirement and survivor benefits.¹ The MINT I model therefore focuses on persons born between 1931 and 1960 and projects their expected date of death, marital history, and year of Social Security benefit take-up.² The model also statistically creates expected former and future spouses not directly observable from the SIPP panels. Thus, the policy universe for retirement income estimates is the surviving population born

¹ Although the MINT I model projects the expected year that individuals will start Social Security disability benefits, DI beneficiaries are currently not included in its universe for policy analysis.

² The marital, mortality, and earnings projections in MINT I are made on a larger sample of respondents born between 1926 and 1965. Pension and asset projections in MINT I are made on the sample of individuals born between 1931 and 1960. MINT I omits baby boomers born between 1961 and 1965 from its analysis of retirement income policy because reliable income projections cannot be made with data from the SIPP panels in the early 1990s for persons that young.

between 1931 and 1960 that is expected to reach age 62 and to receive Social Security retirement and survivor benefits.

This paper summarizes the work completed by the Brookings Institution, RAND, and the Urban Institute for the MINT I model. In most cases, several methods of estimating and projecting demographic characteristics and income were researched and tested; however, this appendix describes only those methods eventually used in the MINT I model. More detailed information can be found in RAND's final report by Panis and Lillard (1999) and the Urban Institute's final report by Toder and others (1999).

Demographic Projections

Because an individual's Social Security benefit depends not only on his or her earnings history but also, to a large extent, on his or her marital history and spouse's earnings history, the MINT I model projects mortality, marital status, and disability status for all individuals in the MINT I data system.

Mortality and Marital Status

RAND researchers projected marital status and mortality simultaneously for respondents in the 1990-1993 SIPP who participated in all eight waves of the surveys. The hazards of own and spousal mortality were based on model parameters estimated from the Panel Study of Income Dynamics (PSID) and corrected for differences between the PSID and the U.S. population using United States Vital Statistics data. The hazards of (re)marriage and divorce were based on model parameters estimated from the SIPP. Parameters from the hazards of marital change and mortality were then used to project

dates of marital change and death for respondents and their spouses. The processes that determined projected dates of (re)marriage and death included the hazard of:

- Own mortality,
- Spousal mortality (leading to widowhood),
- Getting married or remarried, and
- Getting divorced.

Methodology: Mortality. Mortality was estimated using data from the 1968-94 PSID for individuals aged 30 and older in a continuous time hazard model (failure-time model) that had a piecewise-linear Gompertz form. The model specification was:

$$\ln h_i^m(t) = \Gamma_m(t) + \Theta'_m X_i,$$

where $\ln h_i^m(t)$ denoted the log-hazard of mortality at time t , $\Gamma(t)$ captured the piecewise-linear age duration dependency and a linear calendar time, and X_i represented the regressors (race, educational attainment, marital status, and permanent income).³ The model was estimated separately for males and females. Table 1 describes the results.

Using United States Vital Statistics data, these estimates were corrected to be representative of the U.S. population. Vital Statistics data were collected at 10-year intervals between 1901 and 1994 and converted into mortality hazard spells. Mortality hazard models were estimated for individuals aged 30 and older using sex, age, calendar

³ Permanent income in the mortality estimation was based on an individual's long-run position in the distribution of household log-income. Using the PSID, permanent income was computed by using the first three years of income after a respondent reached age 30. Annual log-income in each year was then regressed on age (piecewise linear with different slopes before and after age 65), marital status interacted with sex, and the number of adult-equivalents in the household. The residuals were calculated for each respondent and each of his or her three annual incomes. An individual's measure of permanent income was defined as the average of the three residuals.

Table 1. Estimates of PSID Mortality Hazard		
	Males	Females
Constant	-9.6619*** (0.2603)	-10.0891*** (.3314)
Age Slope 30-65	0.0879*** (0.0044)	0.0869*** (0.0057)
Age Slope 65+	0.0793*** (0.0042)	0.0867*** (0.0048)
Calendar Time	-0.0119*** (0.0038)	-0.0152*** (0.0047)
Black	0.1768** (0.0804)	0.3219*** (0.0953)
High School Dropout	0.3778*** (0.0704)	0.0934 (0.0778)
College Graduate	-0.0513 (0.1040)	-0.2514* (0.1427)
Never Married	0.2183* (0.1132)	0.0184 (0.1421)
Divorced	0.4343*** (0.1146)	-0.1185 (0.1527)
Widowed	0.1080 (0.0905)	-0.0041 (0.0805)
Permanent Income	-0.1591*** (0.0435)	-0.2675*** (0.0477)
Income Missing	-0.4083 (1.1828)	-2.1304 (3.9271)
Log-Likelihood	-14424.95	
* indicates p < .10; ** indicates p < .05; *** indicates p < .01		

time, and race as determinants. The same model was estimated using the PSID. Table 2 describes the results of those estimations and the differences between PSID and Vital Statistics coefficients. Respondent and spousal dates of death were projected for the SIPP sample using the coefficients in Table 1 and subtracting the PSID-VS coefficients in Table 2. Social Security administrative data from the Numident file were used to obtain information on actual deaths that occurred between the last survey date and March 1998.

Methodology: Marriage and Divorce. The Marital History module in Wave 2 of the SIPP contains the marital history of respondents through the Wave 2 interview date. The SIPP collects marital history only for the first two and the most recent marriages. Information on the number of marriages and on marital changes occurring between the third and most recent marriage were not available in the SIPP and were imputed using the PSID. To impute the number of marriages, RAND researchers estimated an ordered probit model of the number of marriages, using the period between the dissolution of the second marriage and the most recent marriage date as the sole explanatory variable. The number of marriages was then imputed in the SIPP using the coefficients from that equation.

To impute marital changes occurring between the third marriage and the most recent marriage, RAND researchers randomly assigned marital changes to reflect the percentage of marriages ending in divorce (85.1 percent) or widowhood (14.9 percent) in the PSID. Transition dates were selected such that marriages were spread evenly between the dissolution of the second marriage and the most recent marriage date.

Once an individual's marital history, up to and including the most recent marriage, was completed, core files in Wave 3 through Wave 8 of the SIPP were used to update the marriage history through the last interview date. After updating the marriage history,

Table 2. Differences Between Vital Statistics and PSID			
	VS	PSID	PSID-VS
Males			
Constant	-8.3597*** (0.0013)	-9.6791*** (0.2480)	-1.3195*** (0.2480)
Age Slope 30-65	0.0721*** (.0000)	0.0909*** (.0042)	0.0187*** (.0042)
Age Slope 65+	0.0821*** (0.0000)	0.0838*** (0.0039)	0.0017 (0.0039)
Time 1901-1994	-0.0081*** (0.0000)	-0.0179*** (0.0037)	-0.0099*** (0.0037)
Black	0.2815*** (0.0004)	0.3913*** (0.0778)	0.1097 (0.0778)
Females			
Constant	-8.7528*** (0.0016)	-10.2761*** (0.3260)	-1.5233*** (0.3260)
Age Slope 30-65	0.0685*** (0.0000)	0.0902*** (0.0055)	0.0217*** (0.0055)
Age Slope 65+	0.0954*** (0.0000)	0.0862*** (0.0043)	-0.0093** (0.0043)
Time 1901-1994	-0.0141*** (0.0000)	-0.0181*** (0.0044)	-0.0040 (0.0044)
Black	0.3325*** (0.0005)	0.5323*** (0.0912)	0.1998** (0.0912)
Log-Likelihood	-222314824.9	-14498.74	-14498.74
* indicates p < .10; ** indicates p < .05; *** indicates p < .01			

RAND researchers estimated marital changes using the 1990-1991 SIPP panels.

Transitions into marriage and divorce were modeled using a continuous time hazard model. Like the mortality hazard model, those hazard models had a piecewise-linear Gompertz form. The specifications were:

$$\ln h_{ij}^w(t) = \Gamma_w(t) + \Theta'_w X_{ij} \text{ and}$$

$$\ln h_{ij}^d(t) = \Gamma_d(t) + \Theta'_d X_{ij},$$

where $\ln h_{ij}^w(t)$ was the log-hazard that individual i married at time t for the j -th time (w is for wedding) and $\ln h_{ij}^d(t)$ was the log-hazard that individual i divorced at time t for the j th time (d is for divorce); $\Gamma_w(t)$ captured duration dependencies on respondent age, calendar time, and duration since the previous marriage dissolved; $\Gamma_d(t)$ captured duration dependencies on respondent age, calendar time pre- and post-1980, and duration since the current marriage began; and X_{ij} represented the regressors.⁴

The regressors in the marriage hazard were number of marriages, race, education, whether the individual was widowed, and permanent income. The model was estimated separately for males and females. The results are described in Table 3.

The regressors in the divorce hazard were number of marriages, education, and race. The model was estimated separately for males and females, and the results are described in Table 4.

Projections of Mortality, Marriage, and Divorce. Using the coefficients from the mortality, marriage, and divorce hazard models (described above), RAND researchers

⁴ Calendar time pre- and post-1980 was included because RAND researchers found that divorce rates sharply decreased after 1980.

Table 3. Estimates of Getting Married		
	Males	Females
Constant	-23.7332*** (1.2834)	-21.9557*** (0.5813)
Age Slope 0-16	1.1847*** (0.0813)	1.1783*** (0.0370)
Age Slope 16-20	0.6211*** (0.0121)	0.3855*** (0.0072)
Age Slope 20-25	0.0840*** (0.0041)	-0.0545*** (0.0038)
Age Slope 25+	-0.0496*** (0.0010)	-0.0751*** (0.0012)
Slope on Duration Unmarried, 0-3 years	0.1208*** (0.0153)	0.0789*** (0.0146)
Slope on Duration Unmarried, 3-8 years	-0.1086*** (0.0101)	-0.0726*** (0.0094)
Slope on Duration Unmarried, 8+ years	-0.0382*** (0.0074)	-0.0223*** (0.0061)
Calendar Time	-0.0079*** (0.0004)	-0.0036*** (0.0003)
Married Once Before	0.4325*** (0.0327)	0.3590*** (0.0304)
Married Twice Before	0.6669*** (0.0425)	0.6248*** (0.0395)
Married Three + Times Before	1.2981*** (0.0576)	1.2017*** (0.0506)
Black	-0.3587*** (0.0208)	-0.5179*** (0.0183)
American Indian, Eskimo, or Aleut	-0.1756** (0.0750)	-0.0543 (0.0647)
Asian or Pacific Islander	-0.2368*** (0.0491)	-0.2276*** (0.0425)
Hispanic	-0.0592*** (0.0241)	-0.3009*** (0.0232)
High School Dropout	-0.0744*** (0.0153)	0.1284*** (0.0134)
College Graduate	-0.1733*** (0.0153)	-0.4313*** (0.0173)
Widowed	0.2856*** (0.0399)	-0.3813*** (0.0356)
Permanent Income	0.0164*** (0.0059)	-0.0279*** (0.0049)
Log-Likelihood	-328842.25	
* indicates $p < .10$; ** indicates $p < .05$; *** indicates $p < .01$		

Table 4. Estimates of Getting Divorced		
	Male	Female
Constant	-1.0198*** (0.1100)	-1.7268*** (0.0946)
Age Slope 0-30	-0.1193*** (0.0038)	-.1021*** (0.0032)
Age Slope 30+	-0.0400*** (0.0015)	-0.0523*** (0.0015)
Marriage Duration, 0-1 years	0.4439*** (0.0724)	0.7350*** (0.0694)
Marriage Duration, 1-4 years	0.2395*** (0.0117)	0.1526*** (0.0107)
Marriage Duration, 4-15 years	-0.0228*** (0.0032)	-0.0156*** (0.0030)
Marriage Duration, 15-25 years	-0.0386*** (0.0048)	-0.0275*** (0.0044)
Marriage Duration, 25+ years	-0.0875*** (0.0060)	-0.0832*** (0.0052)
Calendar Time, pre-1980	0.0401*** (0.0010)	0.0429*** (0.0008)
Calendar Time, post-1980	-0.0025 (0.0020)	0.0058*** (0.0019)
Second Marriage	0.5737*** (0.0248)	0.6368*** (0.0232)
Third or Higher Marriage	1.2503*** (0.0396)	1.3584*** (0.0338)
High School Dropout	-0.0274 (0.0208)	-0.0085 (0.0186)
College Graduate	-0.2117*** (0.0204)	-0.1068*** (0.0215)
Black	0.1198** (0.0276)	0.1786** (0.0240)
American Indian, Eskimo, or Aleut	0.3339*** (0.0766)	0.3237*** (0.0611)
Asian or Pacific Islander	-0.6198*** (0.0692)	-0.6378*** (0.0610)
Hispanic	-0.3015*** (0.0343)	-0.2076*** (0.0314)
Log-Likelihood	-687975.70	
* indicates p < .10; ** indicates p < .05; *** indicates p < .01		

projected mortality and marital status dates.⁵ For example, the probability of living through time t was equal to the value of the survivor function at time t , $S^m_i(t)$. For each respondent, a random number was generated from a uniform (0,1) distribution. Respondent i was projected to die at time t^d_i such that $S^m_i(t^d_i)$ was equal to the random draw. At t_0 , the first projection year, the respondent was still alive and thus $S^m_i(t_0)=1$. The survivor function decreased monotonically after t_0 and was zero at the oldest age to which a given individual would live. Using this method, marriage and divorce dates were also determined.

After the dates of each of these events were determined, the dates were compared. The earliest event determined the respondent's first transition. If the respondent did not die, a new draw was taken from each of the possible events described above, and the dates were then compared a second time to determine the respondent's second transition. The process of redrawing and comparing dates continued until the respondent died.

Onset of Disability

Using retrospective disability information from the 1990 and 1991 SIPP panels, RAND researchers estimated the hazard of individual i becoming disabled at time t .⁶ The hazard model had a piecewise-linear Gompertz form that captured duration dependencies on respondent age. The regressors in the disability hazard were sex, education, and race. The dependent variable had a value of 1 if the individual had a health problem that limited

⁵ Permanent income in the SIPP panels was computed using the same methodology as previously described, except that annual income was constructed from monthly income information. Annual household income in the first year was computed by summing the first 10 months of reported household income. Annual household income was computed in the second year by summing the next 12 months of reported household income and in the third year by summing the last 10 months of reported household income.

⁶ Given that the respondent was not disabled before or at age 30.

the kind or amount of work that he or she could perform (and 0 otherwise). The results of the estimation are described in Table 5. RAND researchers used the coefficients from the disability hazard (reported in Table 5) to impute missing disability status and date of disability onset for respondents. Missing values for disability status were imputed by comparing the projected date of disability onset with the respondent's age at the time of the last interview.

Simulated Marriage Partners

After RAND researchers imputed the number of marriages an individual had in his or her lifetime, they imputed the spousal characteristics of those marriages. Thus, the MINT I data system includes characteristics of spouses who were observed at the time of the SIPP interview, as well as those of spouses who were never part of the SIPP panels. For former and future spouses, RAND researchers imputed birth date, race, Hispanic ethnicity, education, disability status, and date of disability onset. RAND researchers imputed spousal age (birth date) using the empirical distribution of age differences between spouses. They imputed spousal race and Hispanic ethnicity based on the bivariate empirical distribution of husbands' and wives' race combinations. Educational attainment was imputed based on spousal education combinations, and spousal disability was imputed based on spousal characteristics that predicted disability status. The coefficients from the disability hazard were used to impute disability status and date of disability onset for former, current, and future spouses (based on spousal age, education, and race). Missing values for disability status were imputed by projecting a date of disability onset and comparing it with the spouse's age in each year.

Constant	-7.3766*** (0.1786)
Age Slope 30-65	0.0526*** (0.0045)
Age Slope 45+	0.1746*** (0.0047)
Male	0.0062 (0.0348)
High School Dropout	0.7312*** (0.0389)
College Graduate	-0.6668** (0.0577)
Black	0.2779*** (0.0487)
American Indian, Eskimo, or Aleut	0.5446*** (0.1465)
Asian or Pacific Islander	-0.5249*** (0.1378)
Hispanic	-0.1674** (0.0681)
Log-Likelihood	-25736.61
* indicates $p < .10$; ** indicates $p < .05$; *** indicates $p < .01$	

Although RAND researchers imputed demographic characteristics of former and future spouses, they did not actually impute spouses for each unobserved marriage. Researchers from the Urban Institute used a statistical matching algorithm to identify a spouse with the characteristics specified by researchers from RAND. Urban Institute researchers imputed proxy spouses based on statistically imputed characteristics and the closeness (minimizing a distance function) of other respondents to the imputed spouse characteristics. The pool of potential spouses was limited to individuals of the proper sex born within two years of the desired birth year and who survived until at least age 70 or the marriage termination date.⁷

Within the pool of potential spouses, the “best” individual was selected to be the spouse, where “best” was defined to be the individual with the smallest distance measured by a distance function. The distance function was:

$$D_d = \sum_{j=1}^n w_j * [(X_{dj} - X_{rj}) / \sigma_j]^2,$$

where j was the number of measured attributes in the distance function, w was the weight factor, X was a characteristic measure, σ was the standard deviation of the j th X variable in the data set, d denoted the characteristic of the donor, and r denoted the characteristic of the recipient. The characteristics measured in the distance function included spouse’s birth date, Hispanic ethnicity, education, race, death date, disability date, disability status, permanent income, marriage start date, marriage end date, and marriage termination status (divorce, widow, death). Characteristic measures that were deemed more

⁷ One of the reasons for imputing spouses using donors was to “borrow” the earnings histories of donors. Age 70 was chosen as a cutoff since earnings histories are completed for most individuals by age 70.

important to match with were given more weight in the distance function through the weight factor w_j .⁸ Table 6 lists the values of the weights and standard deviations used in the distance function, and Table 7 describes the quality of the matches for imputed spouses.

Receipt of Social Security Retirement Benefits

For each year from 1997 through 2031, the MINT I model projects the age at which individuals will first collect Social Security retirement benefits. Predictions were derived using the 1990-1993 SIPP panels matched to Social Security data from the Summary Earnings Records (SER) and Master Beneficiary Records (MBR).

Methodology. The probability of receiving benefits was estimated for individuals aged 62 and over. Individuals who received disabled-worker benefits before age 62 were deleted from the sample. The probability of receiving Social Security at age t given that individual i did not receive benefits at age $t-1$ was estimated using a logit specification:

$$d_{it}^* = \alpha + \beta X_{it} + \varepsilon_{it}$$

$$d_{it} = 1 \quad \text{if } d_{it}^* > 0, \text{ and } 0 \text{ otherwise.}$$

Where d_{it} indicated whether the individual i received Social Security benefits at time t , X_{it} represented the characteristics of individual i in year t , and ε_{it} was a random error term.

The regressors in the equation included the respondent's age, education, race, pension coverage, earnings, sex, marital status, nonhousing wealth, an indicator of home

⁸ Because the primary reason for imputing spouses was to ascertain the earnings records of spouses who were never part of the SIPP panels, larger weights were assigned to characteristics most likely to affect earnings. Those characteristics included date of death, date of disability, and permanent income. The standard deviation in the distance function scaled the differences to some common unit and reduced the impact of highly variable characteristics.

Table 6. Weights and Standard Deviations Used in the Distance Function		
Characteristic	Weight	Std. Dev.
Birth Date	3	4320.99
Hispanic	1	0.2761
Education	1	0.6358
Race	1	0.6011
Death Date	2	6580.35
Disability Date	2	6258.08
Disability Status	1	0.2978
Permanent Income	5	0.7632
Marriage Start Date	1	5170.19
Marriage End Date	5	8431.54
Marriage Termination Status	1	0.8090

Table 7. Percentage of Spouses Who Match the Characteristics for Imputed Spouses	
Characteristic	Imputed Spouse
Birth Date (within 2 years)	90.5
Hispanic	99.9
Education	93.0
Race	96.4
Death Date (within 3 years)	85.0
Disability Date (within 3 years)	67.3
Disability Status (within 3 years)	99.5
Permanent Income (within .03)	87.9
Marriage Start Date (within 3 years)	37.8
Marriage End Date (within 3 years)	46.3
Marriage Termination Status	92.4

ownership, and the home value, as well as the spouse's age, education, earnings, and pension coverage. Researchers from the Urban Institute estimated the above equation for three groups—married males, married females, and unmarried males and females pooled together. The results of the estimation are described in Table 8. Coefficients from the equations can be interpreted as the effects of each variable on the log-odds of taking up Social Security benefits at age t given nonreceipt at age $t-1$.

Projections. The estimated coefficients from Table 8 were used to project the timing of Social Security receipt for a subset of the MINT I population that was screened for Old-Age, Survivors, and Disability Insurance (OASDI) eligibility, based upon the individual's and, where applicable, his or her spouse's quarters of Social Security coverage.⁹ The primary insurance amount (PIA) was used to screen for eligibility.¹⁰ If the PIA was above zero, then the individual was eligible for earned retired-worker benefits in his or her own right. If, where applicable, the individual's current or most recent spouse had a nonzero AIME, then the individual was qualified for spouse benefits. An individual's second-to-last spouse was also examined for a nonzero AIME if his or her second-to-last marriage ended in widowhood and he or she remarried after age 60.

⁹ Estimates are based on patterns of Social Security benefit timing for a sample whose normal retirement age (NRA) was 65. Under current law, the NRA is scheduled to gradually increase to 67 for future cohorts. For simplicity, Urban Institute researchers assumed that all future workers elected to receive benefits by or at age 67. Chapter 5 in Toder and others (1999) includes a detailed discussion of the related issues.

¹⁰ The PIA was computed using the average indexed monthly earnings (AIME). Researchers from the Urban Institute computed AIMEs using observed and projected annual earnings. Annual earnings through age 60 were indexed to the year the individual turned age 60. Annual earnings after age 60 were nominal earnings. The highest 35 years of earnings after age 21 were used in the AIME formula.

Table 8. Logistic Regression Coefficients			
Variable	Parameter Estimate (Standard Error)		
	Married Men	Married Women	Unmarried People
Constant	-2.3859** (0.5465)	-2.2842** (0.7561)	-0.5755** (0.1601)
Age 63	-1.1243** (0.1304)	-1.6080** (0.1448)	-1.3359** (0.1456)
Age 64	0.3890** (0.1235)	-0.5367** (0.1418)	-0.4756** (0.1322)
Age 65	1.4978** (0.1621)	-0.1443 (0.1763)	0.5582** (0.1508)
Age 66	-0.2732 (0.2488)	-1.2332** (0.2456)	-1.1512** (0.2243)
Age 67	-0.8792** (0.3017)	-1.2641** (0.2532)	-0.5951** (0.1961)
Education < 12	0.1340 (0.1250)	-0.1014 (0.1349)	0.2210 (0.1227)
Education > 12	-0.1702 (0.1174)	-0.0927 (0.1251)	-0.1775 (0.1201)
Non-Hispanic white	0.4141** (0.1491)	0.3157* (0.1473)	0.4088** (0.1180)
Have any pension?	0.3292** (0.1067)	-0.3119** (0.1109)	0.0299 (0.1087)
Earnings ages 35-55	1.6485** (0.1455)	1.4474** (0.2427)	0.9120** (0.1728)
Earnings ages 56-61	0.3440** (0.1277)	0.2942 (0.2464)	0.0851 (0.1736)
Earnings at t-1	-1.4664** (0.1004)	-1.7848** (0.2207)	-0.7840** (0.1380)
Spouse's age	0.0133 (0.00869)	0.0289** (0.0112)	
Spouse's education < 12	0.1104 (0.1255)	0.1953 (0.1379)	
Spouse's education > 12	-0.1012 (0.1162)	-0.1495 (0.1285)	
Spouse's earnings 35-55	0.3421* (0.1534)	1.0697** (0.1061)	
Spouse's earnings at t-1	-0.1654 (0.1121)	-0.5266** (0.0703)	
Spouse has a pension?		0.1381 (0.1149)	
Male			-0.1172 (0.1143)
Widowed			0.1542 (0.1015)
Log value of nonhousing wealth	-0.0155 (0.0524)		
Own a home			-0.0441 (0.1042)
Value of home	-0.0257 (0.0148)		
* indicates p < .05; ** indicates p < .01			

Income Projections

MINT I makes independent projections of each retiree’s income from Social Security benefits, pensions, assets, and earnings (for working beneficiaries) using microdata in four panels of the SIPP data for 1990 through 1993 matched with SSA administrative records for earnings in 1951 through 1996, benefits, and death.

Social Security Earnings

MINT I projects Social Security-covered earnings from 1997 to 2031. Predictions were derived using the 1990-1993 SIPP panels matched to Social Security SER data for individuals who were born between 1926 and 1965 and who had a positive panel weight for the second wave of the 1990-1993 SIPP panels.

Methodology: Social Security Earnings for Individuals in SIPP. Future earnings were predicted using an individual’s observed Social Security-covered earnings (from 1987 through 1996) and individual characteristics. Because the Social Security taxable maximum changes over time and because men (more than women) are likely to have their earnings capped at the taxable maximum, Brookings Institution researchers first created “less censored earnings”—estimates of potential earnings above the taxable maximum for all men with Social Security-covered earnings at the taxable maximum.¹¹

To estimate earnings above the taxable maximum, the researchers from the Brookings Institution divided earnings into three time periods: 1951-1977, when information was available on the quarter in which the individual’s earnings reached the

¹¹ Less censored earnings were found to produce predictions with smaller error and smaller bias (in the out-of-sample period) than Social Security-covered earnings.

taxable ceiling; 1978-1989, when the taxable earnings ceiling was still being raised relative to the average wage by legislative action; and the 1990s, when the ceiling was a stable ratio (2.46) to the average economy-wide earnings.

For the 1951-1977 period, information on the quarter in which an individual's earnings reached the taxable maximum was available. Using the quarter in which an individual reached the taxable maximum, one can infer what the individual's annual earnings would have been had they not been capped. Table 9 gives the range of potential annual earnings as a fraction of the taxable maximum. For example, if the taxable maximum was \$20,000 and in the first quarter of the year a given individual already has reached the taxable maximum, one can infer that his or her annual income, *ceteris paribus*, would be at least \$80,000 (4 times \$20,000). Therefore, the potential earnings of individuals who reached the taxable maximum in the first quarter of the year was at least 4 times the taxable maximum, which is what is reported in Table 9. The class means were derived from the distribution of earnings in the Current Population Survey (CPS) data for 1965, 1970, and 1975.

For the 1951-1977 period, earnings above the taxable maximum were estimated using the average values reported in Table 9. Those earnings were then truncated to 2.46 times the economy-wide average wage to make them consistent in their expected values with the reported data for 1990 to 1996.

For the 1978-1989 period, earnings above the taxable maximum were set at the CPS average of earnings above the taxable maximum for each year. The CPS average was computed separately for men and women. Those earnings were then truncated to 2.46

Table 9. Distribution of Class Means of Potential Earnings		
Quarter Reached Maximum	Range of Potential Earnings (Class)*	Mean of Class in CPS Data
4	$1 < w < 4/3$	1.14
3	$4/3 < w < 2$	1.53
2	$2 < w < 4$	2.36
1	$4 < w$	5.00
* w was the ratio of potential earnings to the taxable maximum.		

times the economy-wide average wage to make them consistent in their expected values with the reported data for 1990 to 1996.

For the 1990-1996 period, earnings above the taxable maximum were left at 2.46 times the economy-wide average.

Earnings were estimated using a fixed-effects specification where the dependent variable was “less censored earnings” for men and Social Security-covered earnings for women.¹² For both men and women, earnings in year t were divided by the economy-wide average earnings in year t . The model was:

$$y_{it} = \alpha + \beta_1 Age_{it} + \mu_i + \varepsilon_{it},$$

where y_{it} was earnings in year t for individual i , Age_{it} was a set of variables that represented individual i 's age in year t , and ε_{it} was a time-varying random error term that satisfied the standard assumptions. In a fixed-effects model, estimates of coefficients of variables that do not vary over time for a single observation cannot be obtained. The effects of those variables are captured in a person-specific individual effect, also known as the fixed-effect. In the equation above, μ_i was the fixed-effect and represented the permanent average difference in earnings between individual i and other members of his or her cohort. The fixed-effect persists over individual i 's entire work career.

Table 10 lists the set of age-group variables included in the model. Earnings predictions were generated separately for eight groups of SIPP respondents – four subsamples of men divided by educational attainment and four subsamples of women

¹² Researchers from the Brookings Institution considered other models for estimating earnings. Chapter 2 in Toder and others (1999) describes these models and their positive and negative features.

Table 10. Age Group Categories	
Variable Name	Definition
Age24	Age 22-24
Age29	Age 25-29
Age34	Age 30-34
Left out category (Age39)	Age 35-39
Age44	Age 40-44
Age49	Age 45-49
Age54	Age 50-54
Age57	Age 55-57
Age59	Age 58-59
Age61	Age 60-61
Age62	Age 62
Age64	Age 63-64
Age65	Age 65
Age67	Age 66

divided by educational attainment (individuals with less than a high school diploma, those with a high school diploma and no further education, those with 1 to 3 years of college, and those with 4 or more years of college). In the highest educational attainment category (individuals with 4 or more years of college), the age variables were interacted with a variable indicating whether the individual had received schooling beyond 4 years of college. In theory, one equation could have been estimated for men and a second equation for women. In practice, that would be difficult to accomplish.¹³

Projections: Social Security Earnings for Individuals in SIPP. Predictions of earnings outside the estimation period were derived by adding an estimate of the person-specific fixed-effect (μ_i) to estimates of $\beta_1 Age_{it}$ to produce an estimate of individual i 's expected covered earnings in year t . A time-varying error term was also added to the prediction in order to generate predictions that had a variance similar to that of actual covered earnings. The error term was generated by forming estimates of each individual's time-varying error term (ε_{it}) for each year between 1987 and 1996 and then randomly selecting one of the 10 "observed" error terms.

Using the date of death projected by RAND researchers, predicted earnings were zeroed out for all years including the year a person was predicted to die and the years after death. RAND researchers also estimated the probability that an individual would experience a health problem that limited the kind or amount of work that he or she could perform. Using that variable, researchers from the Brookings Institution estimated the

¹³ Estimating a panel model of earnings for men and women within these four separate groups was equivalent to allowing a full interaction between the effects of age and educational attainment in a fixed-effects specification. One advantage to estimating the model in this way was that large sample sizes could be managed more easily. A second advantage of estimating four equations was that it allowed the variance of the time-varying error term to differ by educational group.

probability of Disability Insurance (DI) onset and zeroed out earnings for years starting in the calendar year a person was predicted to begin receiving DI.

The MBR contains data on the receipt of DI benefits from 1957 through 1998. Researchers from the Brookings Institution derived predicted probabilities of DI onset for future years using MBR data for 1987 through 1994 matched with the SIPP and SER data files. The sample included all persons predicted to be alive and under age 65 in a given year. The probability of DI onset was estimated using a probit specification:

$$d_{it}^* = \alpha + \beta X_{it} + \varepsilon_{it},$$

$$d_{it} = 1 \quad \text{if } d_{it}^* > 0, \text{ and } 0 \text{ otherwise.}$$

In the above equation, d_{it} indicated whether individual i began receiving DI benefits in year t , X_{it} represented the characteristics of individual i in year t , and ε_{it} was a time-varying random error term. The regressors in the equation included race, disability status, age, education, and average Social Security-covered earnings in the 10-year period ending in the calendar year before year t .

Ten-year average earnings were divided into six to eight categories to reflect the nonlinear impact that past earnings were expected to have on the probability of DI onset. The first category represented a very low level of 10-year average earnings (15 percent or less of economy-wide average earnings), while higher categories reflected successively higher levels of 10-year average earnings. Researchers from the Brookings Institution used actual earnings to compute estimates of 10-year average earnings in 1977 through 1996, and they predicted Social Security-covered earnings to derive estimates of 10-year average earnings from 1997 on. Table 11 lists the set of 10-year average earnings variables included in the model.

Table 11. Categories of 10-Year Average Earnings as a Percentage of Economy-Wide Earnings		
Variable Name	Women	Men
Avgearn1	0-15 percent	0-15 percent
Avgearn2	15-30	15-30
Left out category (Avgearn3)	30-70	30-70
Avgearn4	70-100	70-100
Avgearn5	100-130	100-130
Avgearn6	130 percent or more	130-180
Avgearn7	-----	180-210
Avgearn8	-----	210 percent or more

For 1995 through 1997, the predicted probability of DI onset was adjusted to reflect information in the MBR on actual receipt of DI benefits. For 1998 and on, DI onset was predicted solely on the probabilities predicted by the coefficient estimates in the equation described above.

DI status was assigned by comparing the person's predicted probability of DI onset to a randomly generated number from a uniform (0,1) distribution. Persons with random numbers below their predicted probabilities were assigned to DI. Those with random numbers above their predicted probabilities were identified as non-DI recipients for the year. After DI status was determined, earnings were zeroed out for years starting in the calendar year a person was predicted to begin receiving DI.

Social Security Earnings for Individuals Not in SIPP. An individual's level of Social Security retirement benefits depends not only on his or her earnings records, but also, to a large extent, on his or her marital history. Individuals can receive spouse, divorced spouse, surviving divorced spouse, widow, or own retirement benefits. An individual's receipt and level of those benefit types depend on his or her own and (ex)-spouse's earnings records. In some cases, earnings records for multiple spouses are needed to determine an individual's Social Security retirement benefit. In the analysis described above, Social Security earnings were projected only for individuals who were surveyed in the SIPP. Without additional information, Social Security retirement benefits cannot be accurately measured in cases where individuals' (ex)-spouses are not in the SIPP survey. RAND researchers projected that 86,456 individuals in the MINT I data system had 119,290 marriages between them. In approximately 57 percent of marriages (67,480), the real spouse was not observed in the data.

To accurately estimate Social Security retirement benefits for those 67,480 marriages, a spouse was imputed from a pool of eligible donors. Once a match was made between a donor spouse and an imputed spouse, the earnings of the donor spouse were assigned to the imputed spouse. Urban Institute researchers imputed spouses using the methodology described in section 4.0. Yearly earnings of imputed spouses were aligned to correspond to yearly earnings of the donor spouse at each age. For example, the 1985 earnings of an imputed spouse were for a 20-year-old, while the 1985 earnings of the donor spouse were for a 15-year-old. After the alignment, the 1985 earnings of the imputed spouse corresponded to the 1990 earnings of the donor spouse (the donor spouse was 20 years old in 1990). Instead of being assigned the earnings of a 15-year-old, the 20-year-old imputed spouse was assigned the earnings of a 20-year-old donor spouse.

Partial Retirement Earnings

Methodology. To project the earnings of working Social Security beneficiaries, Urban Institute researchers first predicted which earnings groups beneficiaries would fall into and then randomly assigned earnings to those beneficiaries based on their predicted earnings groups.¹⁴ To increase sample sizes and to capture potential differences in patterns of work after retirement, the equations described below were estimated on one group of 62- and 63-year-olds and a second group aged 65, 66, 67, and 68. For Social Security beneficiaries in both age groups, the equations were estimated using the 1990-1992 SIPP data matched to SSA administrative records. Additionally, the equations were estimated on beneficiaries

¹⁴ Chapter 6 of Toder and others (1999) describes other methods that were tested to estimate partial retirement earnings.

aged 65-68 in the 1984 SIPP data because those individuals could be observed in a time period during which the Social Security exempt amount changed for persons in that age group. Between 1977 and 1983, the exempt amount for 65- to 68-year-olds more than doubled, increasing from \$3,000 in 1977 to \$6,600 in 1983.

First, Urban Institute researchers predicted the earnings groups of 62- to 63-year-old Social Security beneficiaries using the 1990-1992 SIPP data. Beneficiaries were divided into four earnings groups. Those in group 0 had no earnings, those in group 1 had earnings greater than 0 but less than 85 percent of the exempt amount, those in group 2 had earnings between 85 percent and 115 percent of the exempt amount, and those in group 3 had earnings greater than 115 percent of the exempt amount. The model was estimated on a sample of 1,533 respondents, where 1,094 individuals were in earnings group 0, 300 individuals were in earnings group 1, 94 individuals were in earnings group 2, and 45 individuals were in earnings group 3. The ordered probit equation was estimated as:

$$S_{it}^* = \beta X_{it} + \varepsilon_{it},$$

where S_{it}^* indicated which earnings group individual i fell into in year t , X_{it} represented the characteristics of individual i in year t , and ε_{it} was a random error term. The regressors in the equation included race, education, age, sex, marital status, family wealth, family pension income, family retirement account balances (IRA, 401k, and Keogh), average earnings, and spouse's age. The results of the order probit estimation are presented in Table 12.

Next, the earnings groups of Social Security beneficiaries aged 65-68 were predicted using the 1990-1992 SIPP data. The model was estimated on a sample of 6,138 respondents. Because the sample size of that group was considerably larger than the 62- to

**Table 12. Ordered Probit Results for Social Security
Beneficiaries Aged 62 to 63, 1990-1992 SIPP**

Variable	Parameter Estimate	Standard Error
Non-Hispanic White	0.1549	0.1015
Less than HS Education	-0.3413**	0.0955
HS Education	-0.1011	0.0843
Age 63	-0.0657	0.0682
Married Female	-0.4693**	0.1183
Married Female*Spouse Age	-0.0334**	0.0159
Family Wealth	-0.0023	0.0053
Family Pension Benefits	-0.7562**	0.1348
Family Pension*Married Female	0.4393**	0.1904
Family Retirement Balances	-0.0209	0.0391
Earnings ages 35-55	-0.1772**	0.0897
Earnings ages 56-61	0.4913**	0.0886
Earnings ages 56-61*Married Female	0.5721**	0.1898
Pseudo R-Squared	0.0580	
Chi-Squared	148.36 (df=13)	
Cutoff Value 1	0.2870	0.1346
Cutoff Value 2	1.1194	0.1371
Cutoff Value 3	1.7111	0.1450
** indicates p < .05		

63-year-old group, this model was estimated separately for unmarried individuals, married males, and married females. The ordered probit equation was estimated as:

$$S_{it}^* = \beta X_{it} + \varepsilon_{it},$$

where S_{it}^* indicated which earnings group individual i fell into in year t , X_{it} represented the characteristics of individual i in year t , and ε_{it} was a random error term. The regressors in the equation included education, spouse's age, respondent's age, family wealth, family pension income, and average earnings. The results of the order probit estimation are presented in Table 13.

Next, Urban Institute researchers predicted the earnings groups of Social Security beneficiaries aged 65-68 using the 1984 SIPP data in order to estimate the effect of an increase in the exempt amount on their earnings. The model was estimated on a sample of 5,942 respondents, where 4,553 individuals were in earnings group 0, 871 individuals were in earnings group 1, 290 individuals were in earnings group 2, and 228 individuals were in earnings group 3. First an ordered probit equation was estimated as:

$$S_{it}^* = \beta X_{it} + \varepsilon_{it},$$

where S_{it}^* indicated which earnings group individual i fell into in year t , X_{it} represented the characteristics of individual i in year t , and ε_{it} was a random error term. The regressors in the equation included the 1984 Social Security exempt amount, race, education, sex, age, and average earnings. The results of the ordered probit estimation are presented in Table 14. The coefficient on the Social Security exempt amount was not statistically significant. Although this result suggests that the Social Security exempt amount had no

**Table 13. Ordered Probit Results for Social Security
Beneficiaries Aged 65 to 68, 1990-1992 SIPP**

Variable	Unmarried Individuals		Married Males		Married Females	
	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
Less than HS Education	-0.3766**	0.0885	-0.2947**	0.0661		
HS Education	-0.1127	0.0842				
Spouse's Age			-0.0171**	0.0052	-0.0126	0.0081
Age 66	-0.1770*	0.0955				
Age 67	-0.2572**	0.0940	-0.1800**	0.0717	-0.1516	0.0881
Age 68	-0.3177**	0.0942	-0.3171**	0.0729	-0.2936	0.0959
Family Wealth	-0.0002	0.0094	-0.0007	0.0046		
Family Pension Benefits	-1.2946**	0.1836	-0.6591**	0.0807	-0.3996	0.0964
Earnings ages 35-59	-0.1109	0.0957	-0.5460**	0.0769		
Earnings ages 60-64	1.0328**	0.0885	0.7097**	0.0531	1.4470	0.0920
Pseudo R-Squared	0.0913		0.0845		0.1283	
Chi-Squared	229.98 (df=9)		278.02 (df=8)		257.74 (df=5)	
Cutoff Value 1	0.5757	0.1042	-1.0196	0.3316	0.3831	0.5565
Cutoff Value 2	1.4070	0.1104	-0.0222	0.3315	1.3188	0.5586
Cutoff Value 3	1.7707	0.1180	0.3688	0.3326	1.7084	0.5616
** indicates p <.05						

Table 14. Order Probit Model of Social Security Beneficiaries Aged 65 to 68, 1984 SIPP		
	Order Probit	
Variable	Parameter Estimate	Standard Error
Social Security Exempt Amount	-0.5619	0.4061
Non-Hispanic White	-0.1222**	0.0547
Less than HS Education	-0.1540**	0.0476
HS Education	0.0066	0.0511
Male	0.2251**	0.0440
Age 66	-0.1107**	0.0482
Age 67	-0.1970**	0.0492
Age 68	-0.3030**	0.0507
Earnings ages 45-59	-0.2067**	0.0663
Earnings ages 60-64	0.7134**	0.0567
Pseudo R-Squared	0.0533	
Chi-Squared	480.51 (df=10)	
Cutoff Value 1	0.5710	0.1705
Cutoff Value 2	1.2527	0.1711
Cutoff Value 3	1.7028	0.1722
** indicates p <.05		

impact on beneficiaries' decisions to work, other analyses suggest that the Social Security exempt amount does affect beneficiaries' decisions about how much to earn.¹⁵

As such, three earnings equations were estimated using ordinary least squares (OLS) to measure the impact of the Social Security exempt amount on how much beneficiaries earn. Each earnings equation represented the distribution of earnings within each group in order to capture differential effects of an increase in the exempt amount. The earnings equations were estimated as:

$$E^1 = \alpha^1 + \beta_1^1 SSEXempt + \beta_2^1 Age66 + \beta_3^1 Age67 + \beta_4^1 Age68 + \beta_5^1 AvgEarn_{(60-64)} + \lambda^1 + \varepsilon^1,$$

$$E^2 = \alpha^2 + \beta_1^2 SSEXempt + \beta_2^2 Age66 + \beta_3^2 Age67 + \beta_4^2 Age68 + \beta_5^2 AvgEarn_{(60-64)} + \lambda^2 + \varepsilon^2,$$

$$E^3 = \alpha^3 + \beta_1^3 SSEXempt + \beta_2^3 Age66 + \beta_3^3 Age67 + \beta_4^3 Age68 + \beta_5^3 AvgEarn_{(60-64)} + \lambda^3 + \varepsilon^3,$$

where E^1 represented a ratio of individual earnings to national average earnings between 0 and 0.85, E^2 represented a ratio of individual earnings to national average earnings between 0.85 and 1.15, and E^3 represented a ratio of individual earnings to national average earnings greater than 1.15. In the earnings equations, λ represented the inverse Mills ratio—a correction term that was calculated from the coefficients in the ordered probit model. The results of the earnings equations are presented in Table 15.

Projections. The general method for projecting partial retirement earnings involved a two-step approach. The first step was to project into which of the four earnings groups

¹⁵ In a separate analysis, a two-stage earnings equation was estimated. The first stage was a probit model of the decision to work, and the second stage was an earnings equation for those who worked. The coefficient on the Social Security exempt amount in the probit model was not statistically significant; however, the coefficient on the Social Security exempt amount in the earnings equation was statistically significant. Those results suggested that the Social Security exempt amount had no impact on beneficiaries' decisions about whether to work, but did affect their decisions about how much to work.

Table 15. Earnings Equations for Working Beneficiaries, by Earnings Group for Social Security Beneficiaries Aged 65 to 68, 1984 SIPP

Variable	Earnings Group 1		Earnings Group 2		Earnings Group 3	
	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
Social Security Exempt Amount	0.2532**	0.0702	0.9424**	0.0308	1.4448**	0.4959
Age 66	0.0139	0.0095	-0.0037	0.0044	-0.0517	0.0609
Age 67	0.0016	0.0108	0.0036	0.0050	0.0499	0.0786
Age 68	0.0232*	0.0129	0.0011	0.0061	0.1764*	0.0910
Earnings ages 60-64	0.0139	0.0212	-0.0011	0.0102	0.2146	0.1409
Constant	0.0473	0.0405	0.0324	0.0262	-0.1648	0.1956
Inverse Mill's Ratio	-0.0097	0.0292	-0.0157	0.0133	2.2050**	0.9017
R-Squared	0.0288		0.7256		0.3152	
F-Statistic	5.06 (df=6,864)		165.62 (df=6,283)		11.9 (df=6,221)	

** indicates p < .05

* indicates p < .10

individuals fell. The second step was to project the level of earnings for working beneficiaries. To project into which earnings group individuals fell, the predicted probabilities for each earnings group, based on the normal distribution, were computed as follows:

Probability Earnings Group 0:	$\Phi(\text{cutoff value } 1 - X\beta)$
Probability Earnings Group 1:	$\Phi(\text{cutoff value } 2 - X\beta) - \Phi(\text{cutoff value } 1 - X\beta)$
Probability Earnings Group 2:	$\Phi(\text{cutoff value } 3 - X\beta) - \Phi(\text{cutoff value } 2 - X\beta)$
Probability Earnings Group 3:	$1 - \Phi(\text{cutoff value } 3 - X\beta)$

These probabilities were computed using the coefficients and the cutoff values (from Tables 12 and 13) estimated with the 1990-1992 SIPP data, as well as the coefficients on the Social Security exempt variable (from Table 15) estimated with the 1984 SIPP data.

Assignment of earnings groups involved comparing the newly calculated predicted probability with a randomly drawn probability, where the randomly drawn probability was drawn from a uniform distribution (0,1). Suppose, for example, that the predicted probabilities for an individual were as follows:

Predicted Probability of Group 0:	0.532	(cumulative probability was 0.532)
Predicted Probability of Group 1:	0.251	(cumulative probability was 0.783)
Predicted Probability of Group 2:	0.148	(cumulative probability was 0.931)
Predicted Probability of Group 3:	0.069	(cumulative probability was 1.000)

If the randomly drawn probability was 0.456, then the individual was placed in group 0 because the randomly drawn probability was less than 0.532. If the randomly drawn probability was 0.900, then the individual was placed in group 2 because the randomly drawn probability was greater than 0.783, but less than 0.931.

Next, the projection of individuals' level of earnings involved a second randomly drawn probability and some distribution of earnings. The distribution of earnings came

from Social Security beneficiaries in the 1990-1992 SIPP data who were divided into three subsamples, on the basis of their earnings. Subsample 1 included 1990-1992 SIPP Social Security beneficiaries whose earnings fell into earnings group 1. Social Security beneficiaries in the 1990-1992 SIPP whose earnings fell into earnings group 2 were included in subsample 2, while those whose earnings fell into earnings group 3 were included in subsample 3. The level of earnings in the three subsamples were sorted from low to high. Based on a randomly generated number from a uniform (0,1) distribution, individuals projected to fall into earnings group 1 were assigned a level of earnings from the distribution of earnings of subsample 1. Individuals with a low-probability draw were assigned low earnings (relative to the group), and individuals with a high-probability draw were assigned high earnings (relative to the group). Likewise, individuals placed into earnings group 2(3) were randomly assigned a level of earnings from the distribution of earnings of subsample 2(3). Again, individuals with a low-probability draw were assigned low earnings (relative to the group), and individuals with a high-probability draw were assigned high earnings (relative to the group).

It is important to note that individuals were assigned the full earnings estimated in the section on “Projecting Social Security Earnings” until Social Security take-up. Only after each individual in the MINT I sample attained age 62 and began to collect Social Security benefits were his or her earnings levels set to the partial retirement earnings.

Pension Income

Pension income in the MINT I data system includes income from private, federal employee, military, and state and local pensions. Pension benefits were estimated from

information reported by each adult in the Retirement Expectations and Pension Plan Coverage, Annual Income and Retirement Accounts, and Assets and Liabilities topical modules of the SIPP.¹⁶ The SIPP provides descriptive data on the type of pension; years of pension participation; the annual contributions to and the balances of 401(k)s, IRAs, and Keogh accounts; and the contribution rates of employees.

Information on pension coverage from the respondent's primary and secondary job was used to project pension benefits. Once individuals were identified as participants in a retirement plan, Urban Institute researchers projected their pension benefits. Some workers were not covered by a pension plan at the time of the SIPP survey. To account for workers who would have future coverage, participation rates for defined benefit (DB) and defined contribution (DC) plans were projected. The projections were based on the pension participation rates in the 1990-1993 SIPP, where future pension coverage was randomly assigned on the basis of a worker's age, sex, and earnings quintile. For reasons described in Chapter 3 of Toder and others (1999), the earliest that retirees are projected to start collecting pension benefits is age 62.

Benefit Projections: Defined Pensions. Benefits from defined benefit plans were projected using Bureau of Labor Statistics (BLS) replacement rates that varied by age at retirement, years of service, occupation, sector of employment, and final salary.¹⁷ Urban Institute researchers adjusted DB benefits to account for future job changes, cost-of-living

¹⁶ The pension coverage of the labor force measured in SIPP was very similar to the coverage measured in the pension supplement to the 1993 CPS (Iams 1995). That suggests that SIPP provides a reliable estimate of pension coverage because the CPS has been the standard data source for pension coverage (Woods 1994). The CPS pension supplement was discontinued after 1993 and replaced by the SIPP pension topical module.

¹⁷ The BLS stopped determining occupation-specific replacement rates after 1989. Urban Institute researchers created occupation-specific replacement rates after 1989 by adjusting the BLS replacement rates reported in 1993 to occupation-specific replacement rates in 1989.

adjustments (COLAs), and labor force departures before retirement.¹⁸ Benefits expected from pensions on previous jobs (assumed to be from DB plans) and pre-retirement survivor benefits were also included in the DB pension estimates.¹⁹

Years of service on the job were computed from the SIPP work history topical module. Occupation and sector of employment were reported in the core SIPP panel and were based on Census Detailed Occupation and Industry codes. The final salary estimates varied by sector of employment. Private-sector estimates used the highest 5 consecutive years of earnings (actual or projected) out of 10 years immediately before retirement. Public-sector estimates were the average of the highest 3 consecutive years (estimates for military personnel with service dates before 1980 included the last year of earnings only).

Researchers from the Urban Institute also took into account the frequency of job change. First, they determined which workers were more prone to making job changes and how often those changes occurred. They assumed that each year 5 percent of workers in jobs with DB pension coverage would change jobs, regardless of worker or employment characteristics.²⁰ After determining who changed jobs, the researchers then determined a worker's tenure on each job. The assumption used was that each job was twice as long as the job that preceded it. Second, they determined how much pension benefits should be reduced, if at all, for making job changes.

¹⁸ See Chapter 3 of Toder and others (1999) for a detailed discussion of the COLA assumptions used to project DB pension benefits.

¹⁹ All married workers were assumed to take a joint and survivor benefit that paid survivors 50 percent of the benefit the couple would have received had both spouses lived.

²⁰ There were two exceptions. Workers aged 50 or older and workers with fewer than 10 years of service remaining until retirement were assumed to remain on their current job until retirement.

Projections: Defined Contribution Pensions and IRAs. Urban Institute researchers projected account balances for 401(k) plans, non-401(k) defined contribution plans, Keoghs, and IRAs based on account balances and contributions from both employees and employers. For plans where balances were known, the balance was projected to retirement and monthly contributions were accumulated from the time of the survey until retirement. For plans where balances were not known, the monthly contributions were accumulated over the entire period of plan participation. As with DB plans, pre-retirement survivor benefits were also included in the DC pension estimates.²¹

Employee contribution rates were based on self-reported SIPP data. The percentage contributed to each plan was assumed to remain constant until retirement. Because the SIPP data do not report employer match rates, that information was estimated using the Survey of Consumer Finance (SCF). Match rates were estimated separately for 401(k) and non-401(k) DC plans and were correlated with employee contribution rates. In cases in which the employer contributions were unknown, a randomly selected rate was assigned. For workers who did not contribute toward their defined contribution plan, Urban Institute researchers assumed an employer contribution of 5 percent of salary for 401(k) plans and 4.5 percent of salary for non-401(k) plans.

Account balances and new contributions were invested in stocks (50 percent) with a real rate of return of 6.98 percent and bonds (50 percent) with a real rate of return of 3.0

²¹ All married workers were assumed to take a joint and survivor benefit that paid survivors 50 percent of the benefit the couple would have received had both spouses lived. When a spouse died, 50 percent of his or her account balance was transferred to the surviving spouse. The account balance continued to accrue interest until the surviving spouse retired.

percent (the consumer price index (CPI) was assumed to be 3.5 percent). Investment experiences were varied by individual and year by setting the rates stochastically based on a normal distribution with a standard deviation of 17.28 percent for stocks and 2.14 percent for bonds. Currently, the MINT I model assumes a 1 percent administrative fee for each stock and bond account.

The final step in estimating defined contribution pensions was to annuitize the account balance into annual pension income. Account balances were annuitized in two ways. The first method used unisex mortality assumptions based on the 1989-91 Vital Statistics Decennial Life Tables. The second method used mortality assumptions that varied by sex, date of birth, race, and education, based on PSID survey data corrected for differences between the PSID and United States Vital Statistics data. Finally, the annualized benefits were reduced by 20 percent to reflect that people may be risk averse and not spend all their money.

Housing and Nonhousing Wealth

Researchers from the Urban Institute used longitudinal data from the PSID to estimate the age-wealth profile of families.²² Those estimates were then used to project housing and nonhousing wealth for SIPP respondents in the 1990-93 panels. The estimation procedure involved a two-step process. The first step was to estimate the probability of having positive housing and nonhousing wealth. The second step was to impute a positive amount conditional on having that wealth.

²² Urban Institute researchers could observe the wealth-generating process for individuals over a longer period (approximately 10 years) and could isolate individual-specific effects better with the PSID data than with the SIPP data.

Methodology: Housing Wealth. Urban Institute researchers first estimated the probability of positive housing wealth using PSID data in a random effects probit model:

$$d_{it}^* = \alpha + \beta X_{it} + \varepsilon_{it},$$

$$d_{it} = 1 \quad \text{if } d_{it}^* > 0, \text{ and } 0 \text{ otherwise.}$$

In the above equation, d_{it} indicated whether family i had positive housing wealth in year t , X_{it} represented the characteristics of family i in year t , and ε_{it} was a random error term.

The regressors in the equation included the respondent's age, marital status, race, an indicator of whether both spouses worked, family size, and average earnings of both spouses in the prior 5 years. Table 16 describes the results.

Conditional on having positive housing wealth, the value of housing wealth was estimated using PSID data in a random effects model:

$$W_{it} = \beta X_{it} + \mu_i + \gamma_t + C_a + \varepsilon_{it},$$

where W_{it} was the natural logarithm of housing wealth of family i in year t divided by the economy-wide average wage in the year the wealth was reported, X_{it} was a set of income and demographic variables, μ_i represented an individual-specific effect to control for heterogeneity, γ_t isolated certain time-specific effects for 1984 and 1989 to make parameter estimates relative to 1994, and C_a represented cohort-specific effects for each 5-year birth cohort beginning with the cohort born in 1930.

Demographic and income-related variables included in X_{it} were typical of those used in models predicting age-wealth profiles. The demographic variables included the respondent's age, age of his or her youngest child, marital status, sex, race, health status, and family size. The income-related variables included the difference between current earnings and average earnings, the fraction of current wealth held in equities, average

Table 16. Estimates of the Probability of Home Ownership	
Constant	-2.3479 (.0420)
Age of head	.0318** (.0007)
Head is married	.6684** (.0293)
Head is white	.2710** (.0194)
Two earners	.0741** (.0302)
Family size	.0819** (.0075)
Average earnings of both spouses	.4199** (.0133)
Chi-Square	4865.71

** indicates $p < .05$

earnings of both spouses in the prior 5 years, an indicator of whether both spouses worked, and an indicator of whether the family had any pension income. Table 17 describes the results of the estimation.

Methodology: Nonhousing Wealth. Next, Urban Institute researchers estimated the probability of having positive nonhousing wealth using PSID data in a random effects probit model:

$$d_{it}^* = \alpha + \beta X_{it} + \varepsilon_{it},$$

$$d_{it} = 1 \quad \text{if } d_{it}^* > 0, \text{ and } 0 \text{ otherwise.}$$

In the above equation, d_{it} indicated whether family i had positive nonhousing wealth in year t , X_{it} represented the characteristics of individual i in year t , and ε_{it} was a random error term. The regressors in the equation included the respondent's age, marital status, race, an indicator of whether both spouses worked, family size, and average earnings of both spouses in the prior 5 years. Table 18 describes the results.

Conditional on having positive nonhousing wealth, the value of nonhousing wealth was estimated using PSID data in a random effects model. The equation estimated is:

$$W_{it} = \beta X_{it} + \mu_i + \gamma_t + C_a + \varepsilon_{it},$$

where W_{it} was the natural logarithm of nonhousing wealth of family i in year t divided by the economy-wide average wage in the year the wealth was reported, X_{it} was a set of income and demographic variables, μ_i represented an individual-specific effect to control for heterogeneity, γ_t isolated certain time-specific effects for 1984 and 1989 to make

Table 17. Estimates of Housing Wealth	
Constant	-2.3116 (.1206)
Age of head	.0711** (.0049)
Age of head squared	-.0004** (.0000)
Age of youngest child	-.0011 (.0019)
Head is married	-.0653 (.0524)
Head is divorced	-.2435** (.0528)
Head is widowed	-.1274** (.0592)
Head is white	.2582** (.0268)
Head is male	.0772* (.0417)
Family size	.0306** (.0081)
Health status of head	.1057** (.0217)
C01	.1794** (.0544)
C02	.1284** (.0577)
C03	.1268** (.0529)
C04	.1760** (.0446)
C05	.0831** (.0424)
C06	-.1154** (.0430)
Y1984	-.4926** (.0288)
Y1989	-.3572** (.0268)
Average earnings of both spouses	.2126** (.0125)
Current earnings-average earnings	.0953** (.0138)
Two earners	-.0895** (.0261)
Family has pension income	.2161** (.0250)
% Wealth in equities	.0707** (.0164)
Chi-Square	2884.60

** indicates $p < .05$

* indicates $p < .90$

Table 18. Estimates of the Probability of Nonhousing Wealth	
Constant	-.3151 (.0379)
Age of head	.0113** (.0006)
Head is married	.5362** (.0310)
Head is white	.4979** (.0226)
Two earners	-.0700** (.0337)
Family size	-.0451** (.0075)
Average earnings of both spouses	.4157** (.0161)
Chi-Square	2582.23

** indicates $p < .05$

parameter estimates relative to 1994, and C_a represented cohort-specific effects for each 5-year birth cohort beginning with the cohort born in 1930.

In addition to the demographic and income-related variables used in the housing wealth equation, X_{it} includes an indicator of whether the family owned a home. Table 19 describes the results of the estimation.

Projections: Housing and Nonhousing Wealth. Once parameter estimates were obtained using PSID data, they were used to project family wealth for families in the SIPP data who reported positive wealth. Research on wealth suggests that SIPP underrepresents property wealth of the wealthy but adequately represents the wealth of the remaining population (Curtin, Juster, and Morgan 1989). To account for this, an individual-specific “residual” was computed as the difference between the actual wealth reported in the SIPP data and predicted wealth from the estimated equation. That residual was used to center the predicted wealth estimates to match reported wealth on the SIPP and was then added to projected wealth to obtain final estimates of wealth. In cases in which an individual had a change in marital status, the average of the residuals of both spouses (if present) was used in the projections. Use of a residual wealth component in this manner helped to ensure that the projections were tied to reported wealth in the base SIPP files.

For projections of housing wealth, Urban Institute researchers assumed that households in the SIPP 1990-1993 data who reported owning their own homes remained homeowners throughout the projection period. Under that assumption, if a married couple who owned their home in the SIPP 1990-1993 data subsequently divorce, each spouse

Table 19. Estimates of Nonhousing Wealth	
Constant	-3.0430 (.1379)
Age of head	.0517** (.0065)
Age of head squared	-.0003** (.0001)
Age of youngest child	-.0096** (.0030)
Head is married	.2155** (.0548)
Head is divorced	-.3043** (.0520)
Head is widowed	-.1957** (.0716)
Head is white	.5549** (.0347)
Head is male	.3088** (.0462)
Family size	-.0593** (.0109)
Health status of head	.1842** (.0303)
C01	-.1115 (.0775)
C02	.0480 (.0816)
C03	.0803 (.0726)
C04	.0738 (.0606)
C05	-.0681 (.0532)
C06	-.1244** (.0488)
Y1984	-.9285** (.0385)
Y1989	-.7323** (.0362)
Average earnings of both spouses	.5721** (.0184)
Current earnings-average earnings	.3420** (.0203)
Two earners	-.1160** (.0369)
Family has pension income	.4029** (.0383)
% Wealth in equities	.0780** (.0289)
Family owns home	.5158** (.0296)
Chi-Square	5902.01

** indicates $p < .05$

would continue to be a homeowner. For individuals who were not homeowners in the SIPP 1990-1993 panels, the probit value from the housing wealth equation was compared with a uniformly distributed random number to determine if the family would receive an imputed value for housing wealth. After projecting whether housing wealth was positive, the researchers then projected the level of housing wealth for those with positive wealth using the coefficients from the housing wealth regression. Individual-specific residuals were computed and used to calibrate the housing wealth projections against the SIPP data.

For projections of nonhousing wealth, Urban Institute researchers calculated the probability of positive nonhousing wealth for SIPP 1990-1993 respondents using the coefficients from the nonhousing wealth probit equation and compared that value with a uniform random number to determine if the family should be assigned nonhousing wealth. After projecting whether nonhousing wealth was positive, the researchers then projected the level of nonhousing wealth for those with positive wealth using the coefficients from the nonhousing wealth regression. Individual-specific residuals were computed and used to calibrate the nonhousing wealth projections against the SIPP data.

Future Total Retirement Income

All components of income and assets at the age of initial Social Security receipt were aggregated to represent total family income and were projected through 2031 or the date of death, whichever came first. Total income in any year was the sum of projected income from financial assets, Social Security benefit income, earned income, defined benefit pension income, and imputed rental income for those who were homeowners. This

section discusses how changes in postretirement financial assets were estimated and how total income was projected through 2031 or the date of death.

Postretirement financial assets were decremented using a reduced form model of the rate of decline of financial assets for older Americans. The model estimated the curvature of the age-wealth profile and should capture the rate of saving or dissaving of retired individuals. The model was estimated for a sample of couples and individuals from the 1984 and 1990-1993 SIPP panels who were born in or before 1922.²³ Financial assets were computed from the SIPP data using family IRAs, Keoghs, 401(k) balances, the equity (value-debt) of vehicles, other real estate, businesses and farms, and the balance of stocks, mutual funds, bond values, checking accounts, savings accounts, certificates of deposit, and money market accounts. For married couples, all spouse information was retained regardless of the spouse's age. The model was estimated separately for married couples and single individuals. The log of financial assets at time t was estimated using OLS. That equation predicted the level of financial wealth at each age for families with positive financial assets at retirement. The equation estimated was:

$$\ln(W_{it}) = \alpha + \beta X_{it} + \varepsilon_{it},$$

where W_{it} represented financial assets of family i in year t , X_{it} represented the characteristics of family i in year t , and ε_{it} was a random error term. The regressors in the equation included family Social Security benefits, an indicator of home ownership, housing wealth, an indicator of whether the head or wife died within 27 months after the age of initial entitlement, an indicator of whether there was a high earner in the family, and an indicator of pension income. It also included the head's sex, marital status, race, and

²³ Two families were omitted from the sample because they had exceptionally high values of financial assets (over 100 times the national average wage).

birth year, as well as the head's and spouse's DB pension income, average indexed earnings from age 50 to age 60, education, and age at initial entitlement. Table 20 shows the results of the regression.

Using the parameter estimates reported in Table 20, Urban Institute researchers updated the level of financial assets at the age of initial Social Security receipt for each postretirement year. Upon the death of a married spouse, the survivor was assumed to inherit all of the financial assets and the home equity and half of the deceased spouse's DB pension. The widow's financial assets then declined at a rate computed using the coefficients in the model of single individuals. Upon divorce, each partner was assumed to retain half of the financial assets, home, and combined DB pension income. Each partner's financial assets then declined at a rate computed using the coefficients in the model of single individuals. Upon (re)marriage, partners combined their assets as of the new marriage date, and their financial assets declined at a rate computed using the coefficients in the model of married couples. Finally, financial assets were assumed to decline an additional amount in the last year of life.²⁴

Urban Institute researchers computed total income in each year after retirement as the sum of projected income from financial assets (including DC pensions), Social Security benefit income, earned income, and defined benefit pension income. Total income for homeowners also included an imputed rental income based on a 3 percent per year real rate of return on home equity. Income from financial assets in a given year was based upon the

²⁴ Married couples had two additional reductions in financial assets as each partner died.

Table 20. OLS Regression Results by Marital Status		
Variable	Parameter Estimate	
	Married Couples	Single Individuals
Constant	-2.523**	-1.790**
Social Security Benefit	1.130**	2.524**
Ln(Home Wealth)	0.405**	0.342**
Head DB Pension	0.726**	1.374**
Wife DB Pension	0.376	
Head Average Earn Age 50-60	0.400**	
Wife Average Earn Age 50-60	0.026	
Male		0.063
Divorced		-0.366**
Widowed		1.924**
Black		-0.904**
Hispanic		-0.317
Head High School	0.438**	0.498**
Wife High School	0.426**	
Head College	0.658**	0.937**
Wife College	0.450**	
Head Age Initial Entitlement	0.011	0.016*
Wife Age Initial Entitlement	0.019**	
1906<=Head Born<=1918	-0.106	-0.277**
1911<=Head Born<=1916	-0.290**	-0.405**
1917<=Head Born	-0.437**	-0.588**
Head/Wife Dies Within 27 Months	-0.207**	-0.256**
Head Age*HomeOwner	-0.019**	-0.016**
Head Age*Black	-0.004	
Head Age*White	0.009**	0.007*
Head Age*Family Get Pension	-0.003**	
Head Age*High Earnings		0.004**
Head Age*Widowed		-0.027**
* indicates p <.05; ** indicates p <.01		

stock of financial assets left over after taking into account the previous year's spending. Income from the stock of financial assets was computed by determining the amount the family could buy if it annuitized 80 percent of its financial assets. Two different scenarios were used to annuitize financial assets. In the first, the annuity calculation was based on age, sex, education, and race. In the second, the annuity calculation was based on age only. The function calculates the present value of one dollar paid annually until death discounted by 3.5 percent per year CPI and 3.0 percent real rate of return.

Individual Income Taxes

RAND researchers developed a model for estimating individual federal taxes, state taxes (including sales and local taxes), and FICA taxes for each year from 1990 to 2031. The tax model was based on 1998 tax laws. The model included the following assumptions:

- Respondents who were unmarried as of the end of the year filed a single tax return. Married respondents filed a joint tax return. Individuals who became widowed during the reference year and did not remarry in that year filed as married.
- Dependent children could not be claimed as an exemption.
- There was no income from unemployment compensation.
- No deductible IRA contributions were made.
- No student loan interest deduction was made.
- All respondents took the standard deduction, where the standard deduction took account of the respondent's age but assumed that the individual was not blind.
- Respondents were eligible for elderly tax credits only. They were not eligible for tax credits due to disability, child care, education, adoption, foreign tax payments, or other factors.
- The model assumed no self-employment.

For computing federal and state tax burdens, RAND researchers assumed that 100 percent of income from assets was taxable since the MINT I model lacks sufficient data to determine exactly how much income from assets was taxable.²⁵ Income flows from assets were based on annuitization of 80 percent of assets.

State tax regimes varied widely in tax base and tax rates; however, there was significantly less variation when state taxes and local taxes were combined.²⁶ Therefore, the researchers approximated state and local total tax burdens as a constant fraction of federal income tax liability. Because the MINT I data system does not project the future state of residence, California was used as the basis for computing that fraction (California had the median state and local tax burden). The ratio of California taxes (including local taxes and fees) to federal taxes for the elderly population was 83.5 percent. That fraction is a parameter that can be easily modified in the model. The fiscal amounts (thresholds, standard deductions, and exemption amounts) were adjusted according to the CPI for urban wage earners and clerical workers. The Social Security contribution base, above which no OASDI contributions were made, was assumed to increase in proportion with projections in the Social Security average wage index. Taxable Social Security benefits were not indexed and remained the same throughout future years.

Taxable income was computed as AGI minus the value of exemptions and deductions. Federal taxes were then computed according to the instructions provided in

²⁵ That fraction is a parameter that can be easily modified.

²⁶ Nine states did not levy personal income tax at all; 25 states and the District of Columbia based state income tax on federal adjusted gross income (AGI on the 1040 tax forms); eight states based tax liabilities on federal taxable income; two states based state income tax on federal income tax liability; and the remainder specified their own tax bases.

the 1998 1040 tax form. State taxes were computed as 0.835 of the federal income tax liability. Finally, FICA taxes (the sum of OASDI and Hospital Insurance taxes) were computed using earned income as the tax base. These three components could be summed together to get an estimate of individual total income tax liability.

References

- Curtin, Richard; F. Thomas Juster; and James N. Morgan. 1989. "Survey Estimates of Wealth: An Assessment of Quality." In *Studies in Income and Wealth*. Vol. 52, edited by Robert E. Lipsey and Helen Stone Tice. Chicago: University of Chicago Press, pp. 473-548.
- Iams, Howard M. 1995. "The 1993 SIPP and CPS Pension Surveys." *Social Security Bulletin*. 58(4):125-130.
- Panis, Constantijn, and Lee Lillard. 1999. "Near Term Model Development." Draft Final Report, SSA Contract No: 600-96-27335. Santa Monica, Calif.: RAND.
- Toder, Eric, and others. 1999. "Modeling Income in the Near Term – Projections of Retirement Income Through 2020 for the 1931-1960 Birth Cohorts." Draft Final Report, SSA Contract No: 600-96-27332. Washington, D.C.: Urban Institute.
- Woods, John R. 1994. "Pension Coverage Among Baby Boomers: Initial Findings from a 1993 Survey." *Social Security Bulletin* 57(3):12-26.