Projecting Retirement Income of Future Retirees with Panel Data: Results from the Modeling Income in the Near Term (MINT) Project

*by Barbara A. Butrica and Howard M. Iams**

This article describes a model that projects the retirement income of Social Security beneficiaries from 1997 through 2031 using a number of panel data sources. With these data, we examine the composition of retirement income for future retirees in various birth cohorts, racial groups, marital states, and educational categories.

*The authors are with the Division of Policy Evaluation, Office of Research, Evaluation, and Statistics, Office of Policy, Social Security Administration. They thank Susan Grad, James Nunns, John Phillips, Karen Smith, Paul Van de Water, Peter Wheeler, and the MINT team members of Lee Cohen, James Moore, and Mikki Waid for comments and suggestions. This article is a revised version of an earlier paper presented at the National Tax Association Annual Meeting held in Atlanta in October 1999.

Summary

Cross-sectional data capture only a point in time and miss individual changes in earnings, labor force participation, marriage, fertility, and health. Because panel data follow individuals over time, they do not have this problem. The problems or concerns with cross-sectional data may be compounded when these data are used to make projections. Iams and Sandell (1997) found that using panel data on earnings explained much more variation in future earnings than using cross-sectional survey data.

Panel data are also needed to estimate Social Security benefits, especially for women. Because of auxiliary benefits paid to spouses, ex-spouses, and widow(er)s of entitled workers, an individual's Social Security retirement benefit depends not only on his or her earnings history, but also on his or her marital history and the earnings histories of current and previous spouses.

When we compare projected unreduced Social Security benefits with what they would be if we didn't have marital history or earnings history data for men, we find that:

- Benefits computed using only earnings histories are not very different from benefits computed using both earnings and marital histories.
- Benefits computed using only current earnings and marital histories underestimate benefits for those in earlier birth cohorts and overestimate benefits for those in the most recent birth cohort.
- Benefits computed without either marital or earnings histories underestimate benefits for all birth cohorts, but by much more for earlier cohorts than for more recent cohorts.

For women we find that benefits computed without marital or earnings histories underestimate benefits in all birth cohorts. The largest differences are for women in earlier birth cohorts.

Using both marital and earnings histories to estimate unreduced Social Security benefits, we find that men are projected to continue receiving higher benefits than women, although the gap is expected to narrow as the baby boomers near retirement age.

We also look at the composition of projected total income available at retirement for those with incomes in the 45th–55th percentiles of the income distribution and find that:

- Total income at retirement is projected to be larger for men than for women in every birth cohort.
- Women are projected to receive the largest share of their total income from Social Security benefits.
- Men are projected to receive the largest share of their total income from other income sources, although this share declines as the baby boomers near retirement age.

Advantages Of Panel Data

Cross-sectional data capture only a point in time—a period that might not be representative of an individual's overall life experiences—and miss individual changes in health, labor force participation, marriage, and fertility. For this reason, the incidence of poverty and the population in poverty, for example, look very different when one uses cross-sectional data instead of longitudinal data. Bane and Ellwood (1986) found that although most poverty spells are relatively short, poverty spells can appear very long when using cross-sectional data because the chronically poor are oversampled. Iams and Sandell (1997) compared a single year of earnings, available in cross-sectional data, with lifetime earnings, available only in panel data. They found that the last single year of earnings was on average much higher than the average of earnings to date for most age groups.

The problems with cross-sectional data are compounded when these data are used to make projections. Iams and Sandell (1997) predicted average monthly earnings for 1984 through 1993 using several different models. One model included only demographic variables based on 1984 cross-sectional survey data. A second model included only mean indexed monthly earnings computed using 1974–83 panel data from Social Security Administration (SSA) records. The model using only cross-sectional survey variables explained about 25 percent of the variation in future earnings for men and women; at the same time the model using panel data on earnings explained 63 percent of the variance for women and 67 percent for men. The model that used panel data on earnings explained much more variation in future earnings than did the model that used cross-sectional survey data.[1]

Panel data are also needed to estimate Social Security benefits. An individual's Social Security retirement benefit depends not only on his or her earnings history, but also, to a large extent, on his or her marital history and the earnings histories of the current and previous spouse(s). Because of auxiliary benefits paid to spouses, ex-spouses, and widow(er)s of retired workers, the level of Social Security benefits can reflect the relative lifetime earnings of both spouses as a couple.[2] Butrica, Iams, and Sandell (1999) found that married women's projected Social Security benefits were substantially underestimated when benefits were not calculated using the lifetime earnings of both husbands and wives. Information on marital or earnings histories, however, is only available in panel data.

Modeling Income in the Near Term (MINT)

SSA's Office of Research, Evaluation, and Statistics, with substantial assistance from the Brookings Institution, the RAND Corporation, and the Urban Institute, is modeling retirement income in the near term (MINT).[3] The MINT model overcomes the issues raised in the previous section by using panel data to project the retirement income of individuals at the age of first Social Security benefit receipt. With these data, the MINT model directly measures the experiences of individuals up to the early 1990s and statistically projects their future retirement income (Social Security benefits, pension income, asset income, and earnings of working beneficiaries), marital changes, and mortality. The MINT model is described in greater detail in Butrica and Iams (forthcoming), Panis and Lillard (1999), and Toder and others (1999).

Data Sources

Because no single data source contains both marital and earnings histories, the MINT model matches demographic information and marital histories from the Census Bureau's Survey of Income and Program Participation (SIPP) with SSA administrative records for earnings from 1951 through 1996, for benefits, and for date of death. Additional variables of interest, such as wealth, are projected using other sources of panel data.

The SIPP is a nationally representative survey designed as a continuous series of panels extending over 32 to 40 months. Each new panel begins in February and includes approximately 14,000 to 36,700 interviewed households. All U.S. civilian noninstitutionalized household members who are age 15 or older are interviewed. Every 4 months the SIPP asks a respondent about his or her labor force participation, program participation, and income for the current month and the previous 4 months.

The MINT projections on marital history rely on data extending across the entire period of the full SIPP panel. SIPP gathered information on marital histories through three marriages in the second-month interview in each SIPP panel. The SIPP marital history is similar to the special marital history module in the Census Bureau's Current Population Survey (CPS).

The MINT model also relies heavily on the pension coverage information reported in Spring 1991 in the 1990 panel, Spring 1993 in the 1991 and 1992 panels, and Winter 1995 in the 1993 panel. The SIPP questionnaire on pension coverage is similar to the CPS questionnaire on pension coverage and gives similar results (Iams 1995).

Earnings histories are found in SSA's Summary Earnings Record (SER)—an administrative data file containing nearly 400 million earnings records from 1951 to the present. Each record, established when a Social Security number is assigned, contains the earnings history of a worker. The earnings included on the SER file are those that are subject to Social Security taxes. Therefore, only workers in covered employment are included in this file. Additionally, earnings above the Social Security taxable maximum are not reported in this file, as these earnings are not subject to Social Security taxes.

Benefit information is found in the SSA's Master Beneficiary Record (MBR)—an administrative data file containing 161 million records that is used to generate Social Security benefit checks under the OASDI program. The data file includes information on date of application, entitlement (for example, retired-worker beneficiary, auxiliary beneficiary, disability beneficiary), and payment amounts.

Finally, the MINT model uses date of death information that

is found in the Social Security Numident file—an administrative data file containing one record for every Social Security number issued. Each record includes information contained in the application for a Social Security number (for example, date of birth, sex, race, mother's maiden name, father's name, and place of birth), as well as the date of death. Where this information was not available, mortality was estimated using the 1968–94 Panel Study of Income Dynamics (PSID) for individuals aged 30 or older.

After all these files are merged, the final data file contains approximately 113,000 SIPP respondents born between 1926 and 1965. For most analyses, the policy universe for retirement income estimates is the surviving population born from 1931 through 1960 that is expected to reach retirement age and to receive Social Security retirement and survivor benefits. In addition to these sample criteria, the policy universe of the analyses described in this article excludes disabled persons.

Uses of MINT Data

Because many of the parameters in the MINT data system can be altered, the MINT model has numerous possible uses in policy evaluation. Currently the MINT model is being used to examine cohort differences in the sources of retirement income and to assess the impact of Social Security benefit reforms on the level of benefits, expected retirement income, and expected poverty level of future retirees. With its detailed demographic information, the MINT model is also being used to examine economic well-being in retirement and differences in Social Security benefits and retirement income by gender, race, education, marital status, and birth cohort.

Currently, the MINT project does not model behavioral responses to policy changes. The MINT model directly measures the experiences of survey respondents up to the early 1990s and statistically projects their characteristics into the future, adjusting for expected demographic and socioeconomic changes. The MINT model implicitly assumes that future populations will behave the same way as past populations with regard to such choices as educational attainment, marriage partners, job types, and the decision to work. Furthermore, the MINT model assumes that the interdependence of the outcomes from such choices, like education and earnings, will remain unchanged in the near future.[4]

Projecting Retirement Income of Future Retirees

Table 1 describes the projected marital status of men and women at retirement by birth cohort.[5] The share of individuals who have still not married at retirement is projected to be larger for children of the baby boom than for their parents (compare 4 percent of never married men and 3 percent of never married women in the 1931–35 birth cohort with 6 percent of never married men and 7 percent of never married women in the 1956–60 cohort). At the same time, a large share of all birth cohorts is comprised of men and women who have married at least once in their lifetimes. To project Social Security benefits accurately for these individuals, one needs to know their marital histories. Cross-sectional data do not provide this degree of detail.

Social Security Benefits

Table 2 makes this point clearer by comparing monthly Social Security benefits to what they would be if we did not have marital history data or earnings history data.[6] The first column represents projected unreduced Social Security benefits that take into account an individual's marital and earnings history, as well as the earnings histories of the current and previous spouse(s). These benefits are computed using algorithms based on SSA's actual computation formula and are equivalent to those represented in the second column plus any auxiliary benefits to which the individual is entitled. Henceforth, we will refer to the numbers in column 1 as benchmark benefits. The second column represents what unreduced Social Security benefits are projected to be if they are estimated with only earnings history data. These benefits are the individual's primary insurance amount (PIA) computed without using any marital history data. One can find this type of information in SSA's administrative earnings records. The third column represents what unreduced Social Security benefits are projected to be if they are estimated with only marital history data. These

Table 1.—Projected marital status of men and women at retirement, by birth cohort

[In percent]

	Men					Women				
Birth cohort	Total[1]	Never married	Married	Widowed	Divorced	Total[1]	Never married	Married	Widowed	Divorced
1931–35	100	4	84	5	8	100	3	67	19	11
1936–40	100	4	81	4	11	100	5	61	19	16
1941–45	100	5	78	4	13	100	5	62	17	17
1946–50	100	4	77	5	13	100	6	60	15	19
1951–55	100	6	75	5	14	100	7	60	15	19
1956–60	100	6	74	5	15	100	7	59	15	19

[1] Due to rounding, totals may not sum to exactly 100 percent.

Source: Authors' calculations using MINT data.

benefits are computed using only 1996 earnings and the individual's marriage history. One can find this type of information in panel data such as the SIPP and the PSID that ask respondents to recall their marital histories. The fourth column represents what unreduced Social Security benefits are projected to be if they are estimated without either marital or earnings history information. These benefits are the individual's PIA computed using only the individual's 1996 earnings. One can find this type of cross-sectional information in the CPS. The first four columns correspond to men and the last four columns, which are identical to each of the first four columns, correspond to women.

We first consider men. For all birth cohorts, benefits computed using only earnings histories (column 2) are not very different from the benchmark benefits. This result is not surprising since benefits for most men depend only on their own earnings histories. Benefits computed using only marital histories (column 3) underestimate the benchmark benefits for those in earlier birth cohorts. For those in the earliest birth cohort, who were 61–65 years old in 1996, benefits based on 1996 earnings most likely underrepresent average lifetime earnings because most individuals in this cohort would have already left the labor force or reduced their work hours as they approached retirement. Benefits computed without earnings histories overestimate the benchmark benefits for those in the most recent birth cohort. For these individuals, who were 36–40 years old in 1996, benefits based on 1996 earnings most likely overrepresent average lifetime earnings because they were still in the labor force in 1996 with earnings that were just beginning to peak. Finally, benefits computed without either marital or earnings histories (column 4) underestimate the benchmark benefits for all birth cohorts, but by much more for earlier cohorts than for more recent cohorts.

We next consider women. For all birth cohorts, benefits computed without marital or earnings histories underestimate the benchmark benefits. Benefits computed using only earnings histories (column 6) underestimate the benchmark benefits (column 5) much more for earlier birth cohorts than for more recent birth cohorts. This is because women in the 1931–35 birth cohort were not as likely as women in the 1956–60 birth cohort to have their own earnings. Most women in the 1931–35 birth cohort receive auxiliary benefits based on their husbands' earnings; however, husbands' earnings are impossible to access without marital histories that link couples together. Benefits computed using only marital histories (column 7) also underestimate the benchmark benefits much more for earlier birth cohorts than for more recent birth cohorts. Although these benefits take into account husbands' earnings by way of marital histories, the earnings information is based on cross-sectional 1996 earnings and not on earnings histories. For the reasons already mentioned, 1996 earnings will underestimate average lifetime earnings for earlier cohorts and overestimate average lifetime earnings for more recent cohorts. This means that benefits for earlier birth cohorts will be greatly underestimated and benefits for more recent birth cohorts will not be very different from the benchmark benefits. Finally, as was the case for men, benefits computed without either marital or earnings histories (column 8) underestimate the benchmark benefits for all birth cohorts, but by much more for earlier birth cohorts than for more recent cohorts.

Table 3 shows the projected mean monthly Social Security benefit at retirement for various subgroups by birth cohort. These benefits are computed using both marital and earnings histories and are comparable to those presented in columns 1 and 5 of table 2. Consistent with earnings patterns, we find that unreduced Social Security benefits among men are gener-

Table 2.—Projected mean monthly unreduced Social Security benefit at retirement computed with and without marital and earnings histories

[1998 dollars]

	Men				Women			
Birth cohort	Benchmark[1] (1)	Only earnings history[2] (2)	Only marital history[3] (3)	No marital or earnings history[4] (4)	Benchmark[1] (5)	Only earnings history[2] (6)	Only marital history[3] (7)	No marital or earnings history[4] (8)
1931–35	$997	$982	$412	$344	$702	$438	$262	$194
1936–40	1,032	1,017	865	815	756	512	576	453
1941–45	1,045	1,027	970	924	800	561	722	549
1946–50	1,047	1,021	1,032	978	824	637	802	618
1951–55	1,015	990	1,013	964	832	678	813	625
1956–60	986	963	1,000	954	819	690	792	595

[1] Represents an individual's unreduced benefits computed using algorithms based on SSA's actual computation formula that uses marital and earnings histories.

[2] Represents an individual's PIA computed using his/her earnings history, but no marital history.

[3] Represents an individual's unreduced benefit computed using his/her 1996 earnings and the 1996 earnings of his/her spouse.

[4] Represents an individual's PIA computed using his/her 1996 earnings and no marital history.

Source: Authors' calculations using MINT data.

ally highest for those who are white, married, or who have education beyond high school. Unreduced Social Security benefits among women are highest for those who are white, widowed, or who have education beyond high school. Men in all birth cohorts are projected to receive higher benefits than women, although the gap is expected to narrow as the baby boomers near retirement age.

Total Retirement Income

Table 4 looks at the composition of projected annual income available at retirement for those with incomes in the 45th–55th percentiles of the income distribution. All statistics in the table represent the mean value for respondents whose total incomes fall between the 45th and 55th percentiles. The percentages in the table represent the mean share of total income from Social Security benefits, other income (assets, pensions, and imputed rental income), and earnings. As expected, total income at retirement is projected to be larger for men than for women in every birth cohort. Women are projected to receive the largest share of their total income from Social Security benefits; in contrast, men are projected to receive the largest share of their total income from other income sources. Also, the share of total income from earnings is projected to be larger for men than for women. Social Security benefits are projected to increase from 37 percent of total income for men in the 1931–35 birth cohort to 41 percent of total income for men in the 1956–60 birth cohort. Income from other sources is projected to decline from 46 percent of total income for men in the 1931–35 birth cohort to 41 percent of total income for men in the 1956–60 birth cohort.

Both men and women's total income is projected to rise through the 1946–50 birth cohort. This trend may be explained by increases in educational attainment that resulted in higher average wages (relative to national average earnings) for those born in the 1946–50 birth cohort compared with those born in earlier birth cohorts. The increase in average wages might increase average lifetime earnings and, in turn, increase Social Security benefits and retirement income (Toder and others 1999). Total income for men and women is projected to decline substantially after the 1946–50 birth cohort. This trend may be explained by the relatively poor job market opportunities available for later birth cohorts compared with the 1946–50 birth cohort (Toder and others 1999). Additionally, from about 1947 through 1973, average wages (adjusted for inflation) grew at 2–3 percent per year. After 1973, average wage growth slowed dramatically (Levy 1998). The decrease in average wages might decrease average lifetime earnings, and, in turn, decrease Social Security benefits and retirement income.

In addition to factors described above, variability in women's total income across birth cohorts is influenced by increased labor force participation and earnings, an increase in the number of unmarried women projected at retirement, and an increase in the percentage of women who will not be eligible for auxiliary benefits.

Table 3.—Projected mean monthly unreduced Social Security benefit at retirement, by birth cohort

[1998 dollars]

Birth cohort	1931–35	1936–40	1941–45	1946–50	1951–55	1956–60
	Men					
Total	$997	$1,032	$1,045	$1,047	$1,015	$986
Race:						
White	1,027	1,058	1,065	1,072	1,044	1,014
Black	784	834	893	831	797	808
Other	748	807	860	936	942	849
Marital status:						
Never married	819	769	797	831	812	812
Married	1,017	1,050	1,064	1,065	1,035	1,003
Widowed	956	1,033	1,081	1,072	1,051	998
Divorced	907	987	1,007	1,004	982	972
Education:						
Less than high school	874	875	827	773	701	685
High school	1,021	1,043	1,035	1,010	966	930
College graduate	1,045	1,089	1,121	1,131	1,141	1,158
	Women					
Total	$702	$756	$800	$824	$832	$819
Race:						
White	709	765	808	835	844	842
Black	677	714	749	749	749	688
Other	598	640	774	766	791	745
Marital status:						
Never married	593	645	648	786	811	750
Married	612	654	703	742	768	770
Widowed	999	1,032	1,059	1,087	1,067	1,049
Divorced	774	853	932	887	852	820
Education:						
Less than high school	627	649	635	549	528	482
High school	706	758	795	805	807	782
College graduate	797	844	906	945	956	997

Source: Authors' calculations using MINT data.

Conclusion

Panel data play a critical role in the estimation of Social Security benefits because the formula for computing benefits requires average lifetime earnings and because auxil-

Table 4.—Composition of projected annual total income at retirement for those with incomes in the 45th-55th percentiles, by birth cohort[1]

[1998 dollars]

	Men					Women				
		Mean percentage share of total income					Mean percentage share of total income			
Birth cohort	Mean total income	Total[2]	Social Security	Other	Earnings	Mean total income	Total[2]	Social Security	Other	Earnings
1931–35.....	$21,850	100	37	46	18	$18,723	100	48	40	12
1936–40.....	22,055	100	36	48	17	18,646	100	46	41	12
1941–45.....	22,338	100	36	46	18	18,761	100	45	42	13
1946–50.....	22,583	100	37	47	17	19,624	100	43	43	14
1951–55.....	20,839	100	41	44	16	18,987	100	44	42	14
1956–60.....	19,291	100	41	41	18	17,258	100	48	38	14

[1] The sample excludes disabled individuals and individuals without any total income. Total income is reported at age 62 for those who first receive Social Security benefits at or before age 62, and at age 67 for those who first receive Social Security benefits after age 62. Total income for married individuals is divided by 2 to represent per capita total income. All reported statistics represent the mean value for respondents between the 45th and 55th percentiles of total income.

[2] Due to rounding, totals may not sum to exactly 100 percent.

Source: Authors' calculations using MINT data.

iary benefits are still commonly paid to spouses, divorced spouses, and widow(er)s of entitled workers. The MINT model estimates the retirement income of future Social Security beneficiaries and the beneficiary population in 2020 using panel data to measure the effects directly into middle age from such choices as educational attainment, marriage partners, job types, and the decision to work. SSA will use the MINT model to assess the impact of Social Security benefit reforms on the level of benefits, expected retirement income, and expected poverty level of future retirees for various subgroups of the population

Notes

[1] A third model included only 1983 earnings data from SSA administrative records. This model explained about 61 percent of the variation in future earnings for women and 63 percent for men. Although it has been shown that lagged earnings are strong predictors of future earnings, there is no reason to believe that this relationship would continue 20 or 30 years into the future.

[2] For a more complete description of how Social Security benefits are computed, see Butrica, Iams, and Sandell (1999).

[3] The near term ranges from 1997 through 2031.

[4] Any alternative behaviors will have to be incorporated into the model by developing scenario-based analyses. For example, using MINT data to examine the impact of raising the early retirement age, we could assume that one behavioral response is to continue to work and another behavioral response is to file for disability benefits. We could then compare the economic well-being of individuals who continue to work with those who exit the labor force once they begin receiving disability benefits.

[5] Except where noted, retirement is defined as the age at which an individual first receives Social Security benefits.

[6] These benefits and those presented in table 3 are unreduced Social Security benefits. They represent the full benefit payable at the normal retirement age and are not reduced for early retirement.

References

Bane, Mary Jo and David T. Ellwood. 1986. "Slipping into and out of Poverty: The Dynamics of Spells." *Journal of Human Resources,* Vol. 21, No. 1, pp. 1–23.

Butrica, Barbara A., Howard M. Iams, and Steven H. Sandell. 1999. "Using Data for Couples to Project the Distributional Effects of Social Security Policy Changes." *Social Security Bulletin,* Vol. 62, No. 3, pp. 20–27.

Butrica, Barbara A. and Howard M. Iams. Forthcoming. "Using Panel Data to Project the Retirement Income of Future Retirees." National Tax Association Proceedings, 92nd Annual Conference, Atlanta, GA.

Iams, Howard M. 1995. "The 1993 SIPP and CPS Pension Surveys." *Social Security Bulletin*, Vol. 58, No. 4, pp. 125–130.

Iams, Howard M. and Steven H. Sandell. 1997. "Projecting Social Security Earnings: Past Is Prologue." *Social Security Bulletin,* Vol. 60, No. 2, pp. 3–16.

Levy, Frank. 1998. *The New Dollars and Dreams: American Incomes and Economic Change*. New York: The Russell Sage Foundation.

Panis, Constantijn and Lee Lillard. 1999. "Near Term Model Development." Final Report, SSA Contract No:600–96–27335. Santa Monica, CA: RAND Corporation.

Toder, Eric and others. 1999. "Modeling Income in the Near Term—Projections of Retirement Income Through 2020 for the 1931–1960 Birth Cohorts." Final Report, SSA Contract No: 600-96-27332. Washington, DC: The Urban Institute.