Projecting Social Security Earnings: Past Is Prologue

*by Howard M. Iams and Steven H. Sandell**

Accurate projections of lifetime earnings are useful in projecting Social Security benefits, trust fund balances, and economic resources of the elderly and the effects of changes in Social Security policy. This article projects lifetime Social Security earnings until retirement using data from the Survey of Income and Program Participation (SIPP) matched to Social Security records of annual earnings from 1951 through 1993.

We first develop, estimate, and test gender-specific multiple regression models of 10-year earnings intervals using the matched 1984 SIPP panel. We find strong relationships predicting the mean indexed monthly earnings level in the 10-year period of 1984-93. We then use the models to project (unobserved) Social Security earnings from 1994 through retirement for persons born between 1931 and 1955. By adding projected earnings to observed annual earnings to date, we forecast lifetime Social Security earnings for persons retiring early in the 21st century.

*The authors are with the Division of Policy Evaluation, Office of Research, Evaluation and Statistics, Social Security Administration. The views in this article are the authors' only and do not necessarily reflect the position of any government agency or research organization. The authors had equal responsibility for this work.

Acknowledgments: We wish to thank the following persons for comments on earlier drafts: participants at seminars at the Center for Economic Research of Tilburg University, at the Department of Economics, University of Leiden, the Netherlands, and at the Public Policy Institute of the American Association of Retired Persons, along with Harriet Duleep, Susan Grad, Alan Gustman, Kajal Lahiri, Lee Lillard, Michael Leonesio, Olivia Mitchell, Robert Myers, Cheryl Neslusan, David Pattison, Kalman Rupp, Patrick Scrivin, Ralph Smith, Kelvin Utendorf, Denton Vaughan, and Peter Wheeler—all with the Social Security Administration. The article is based on a presentation given at the U.S. Bureau of the Census 1996 Annual Research Conference.

I. Introduction

Social Security benefits, a key component of both economic well-being in old age and Federal Government expenditures, are based on lifetime earnings. Although there is strong interest in Social Security benefits over the next 25 years as the baby boom population begins retiring, we do not know what these future retirees will have earned over their lifetimes. Because the work lives of current retirees differ markedly from those of future retirees born during World War II and the baby boom years, we cannot assume that the experience of current retirees applies to future retirees. Current retirees experienced a different economy, society, and Social Security system during much of their work lives than future retirees will have experienced. Thus, projecting lifetime earnings will be important for public policy purposes.

This article develops a technique to project future earnings. It describes simulation models using survey data matched to administrative records. The models, under development in the Office of Research, Evaluation and Statistics, Social Security Administration (SSA), use the Survey of Income and Program Participation (SIPP) panels matched to SSA's record of Social Security covered earnings from 1951 through 1993. In combination with earnings to date, models are used to project future earnings for the population that will be aged 65 or older in 2020, improving on current techniques that use cross-sectional survey data without a history of Social Security earnings. A clear advantage of our approach over conventional methods is a more accurate estimate of lifetime earnings, reflecting individual variation and growth in accumulated covered earnings.

Some researchers have made longitudinal inferences using a series of cross-sectional surveys taken at different points over time of earnings and/or labor force participation rates for gender-specific, age groups (for an example, see Bianchi 1995, pp. 117-122; Smith and Ward 1984). The difficulty with this approach, particularly for women, is that individuals enter and leave the labor force, so differences between periods reflect different samples in addition to differences in earnings over time. The average Social Security taxable earnings for earners in 1980 and for earners in 1990 is presented in table 1.[1] It also presents the average for persons with earnings in both 1980 and 1990 and for those with earnings in only one of those years. Among women, 1980 earners averaged $13,971 and 1990 earners averaged $17,406, a difference of $3,435. However, some earners in 1980 were not earners in 1990, and vice versa. Women with earnings in both periods increased their earnings by $4,986, from $14,394 to $19,380. Women with only 1980 earnings averaged $11,633, and women with only 1990 earnings averaged $11,588. Thus, a series of cross-sections misrepresent the changing earnings of a cohort over time. (Among men, the longitudinal change was only slightly higher than the cross-sectional change indicated.)

A second approach for estimating lifetime earnings uses observed age-specific earnings to estimate age-earnings profiles. In this approach, the age-earnings profile developed from data representing persons of different ages in a cross-sectional data set reflects earnings across the lifecourse of a "synthetic" cohort (for an example, see Lumsdaine, Stock, and Wise 1994; 1995). The age-earnings profile is interpreted as the earnings history of a synthetic or hypothetical group, as if that group experienced earnings observed at each age in the cross section. Of course, no group actually experiences these earnings because they are the survey-year earnings for different cohorts. Unfortunately, because the earnings structure has changed dramatically, the recent cohort's experience is a misrepresentation of an older cohort's experience when that group was younger, and vice-versa. Increased wage rates and labor supply for more recent birth cohorts of women are not reflected in the observed earnings of older cohorts of women. Thus, it would be misleading to project the earnings of women who are currently young based on the earnings of women who are currently old.

A single year's earnings are inadequate as a predictor of past lifetime earnings, as illustrated by matched SSA earnings data from the 1990 longitudinal SIPP panel.[2] Table 2 shows how misleading, in general, a single year's (1993) earnings are as an indicator of the trend for lifetime earnings. The last single year of earnings is on average much higher than the average of earnings to date for each cohort, except for the 1931-35 cohort, which is beginning retirement. If we use these earnings to represent the history of earnings for retirees in 1993, the mean monthly earnings would be $1,846 for men and $1,077 for women. This compares to the actual average observed to date of $1,598 and $617, respectively, for the 1931-35 cohort. The average based on the last year's earnings, as one would use in a cross-sectional analysis, was much higher than the average observed to date. These differences reflect the increase over time in women's labor force activity and the increase in the maximum earnings taxable for Social Security purposes (earnings are truncated at higher levels in later years).

Table 2 also presents the coefficient of determination (that is, the R-square) between each individual's average monthly earnings at age 22 through 1993 (wage indexed to adjust for economy-wide growth in earnings over time) and monthly earnings in 1993.[3] The explained variation between 1993 (current) and observed earnings to date ranged from 0.37 and 0.29, for men and women born in the 1931-35 cohort, to about 0.61 for those in the 1956-58 cohort. This indicates that single-year earnings 'explained' about 29 percent to 61 percent of the variation in observed earnings to date, depending upon sex and cohort. The explained variation was lower among older cohorts than for more recent cohorts.[4]

Among the many changes in the past decades, two major changes affect Social Security earnings histories. The first major change was the increased labor force activity of women, particularly mothers. Because mothers born in the depression, particularly the better educated, were more likely than women born later to drop out of the labor force to care for young children, they had more years without earnings than later cohorts (Iams and Sandell 1994). Thus, a synthetic cohort approach to estimating future earnings underestimates the lifetime earnings of younger cohorts.[5]

The second major change was the increased amount of total annual earnings taxed by Social Security with the rise in the maximum earnings subject to Social Security taxes, especially in the 1970s. The taxable maximum increased about twelvefold, from $4,800 in 1965 to $57,600 in 1993, while average earnings in the national economy increased only about fivefold, from $4,659 to $23,133 (see table 3). The low maximum in earlier years leads to greater truncation of earnings subject to tax and affects a greater proportion of earners. For example, about 49 percent of the men and 13 percent of women workers had total earnings above the maximum in 1965, compared to 9 percent and 2 percent in

Table 1.—Average covered earnings[1] in 1980 and in 1990, by sex and earnings patterns

Sex and earnings pattern	1980	1990	Implied earning growth
Women			
Earners in—			
Year............	$13,971	$17,406	$3,435 cross section
Both years...	14,394	19,380	4,986 longitudinal
Year only.....	11,633	11,588	...
Men			
Earners in—			
Year............	23,911	28,799	4,709 cross section
Both years...	24,693	29,485	4,792 longitudinal
Year only.....	16,513	20,718	...

[1] Earnings are in 1990 dollars after wage indexing with the SSA benefit method and are calculated in annual amounts. Weighted with Wave 4 Topical Module weight.

Source: 1990 SIPP matched to SSA records.

1992 (Social Security Administration 1993; Social Security Administration 1995, table 4.B.4). The rising maximum increased the amount of annual earnings taxed for Social Security purposes and included in retirement benefit calculations. This increased the average Social Security earnings to date of recent cohorts more than for cohorts born earlier, and for men more than for women.

Better estimates of lifetime earnings could improve the accuracy of benefit estimates for future retirees. These improved estimates could be used to evaluate proposals to change the structure of Social Security benefits. They also could be used to improve actuarial projections. To provide better estimates of Social Security benefits of future retirees, this article projects lifetime Social Security earnings from age 22 until retirement age, using data from SIPP matched to Social Security records of annual earnings from 1951 through 1993. We first develop, estimate, and test multiple regression models of 10-year earnings intervals using the matched 1984 SIPP panel.[6] For the majority of persons aged 29-53 in 1984, the 10 years include earnings from the period in which they are well-established in their careers. The models are used to project Social Security earnings in 10-year intervals from 1994 to retirement. The earnings observed for individuals through 1993 are combined with earnings projected until retirement

Table 2.—Coefficient of determination for correlation of 1993 Social Security earnings to lifetime earnings to date, by sex of retired workers born between 1931-58[1]

Birth cohort	Men			Women		
	Coefficient of determination-earnings in 1993 to lifetime	Lifetime earnings per month	1993 earnings per month	Coefficient of determination-earnings in 1993 to lifetime	Lifetime earnings per month	1993 earnings per month
1931-58	0.53	$1,657	$1,902	0.53	$816	$1,127
1931-35	.37	1,598	1,326	.29	617	698
1936-40	.48	1,718	1,719	.42	719	1,005
1941-45	.56	1,784	2,053	.49	759	1,133
1946-50	.56	1,730	2,032	.56	863	1,259
1951-55	.61	1,636	2,023	.58	906	1,233
1956-58	.62	1,478	1,977	.61	888	1,168

[1] Earnings are in 1988 dollars after wage indexing with the SSA benefit method and are calculated in monthly amounts. Lifetime measure averages earnings from age 22 through 1993. Social Security earnings are from SSA records of taxable annual earnings in covered jobs from 1951-93.
Weighted with Wave 4 Topical Module. Retired workers defined as respondents with at least 25 percent of years with earnings in years age 22 through 1993.
Source: 1990 SIPP matched to SSA records.

Table 3.—National earnings as a percentage of the Social Security taxable maximum and percentage reaching the maximum, by selected years.

Year	National average earnings	SSA taxable maximum	Percent reaching taxable maximum	Percentage under taxable maximum	
				Men	Women
1951	$2,799	$3,600	78	65	97
1955	3,301	4,200	79	63	96
1960	4,007	4,800	83	61	94
1965	4,659	4,800	97	51	87
1970	6,186	7,800	79	62	94
1975	8,631	14,100	61	76	98
1980	12,513	25,900	48	86	99
1985	16,883	39,600	43	89	99
1990	21,028	51,300	41	91	98
1992	22,935	55,500	41	91	98
1993	23,133	57,600	40	...	...
1994	24,090	60,600	40	...	...

Source: Social Security Administration 1995, tables 2.A.9 and 4.B.4. 1994 estimate from the Board of Trustees of the Federal Old-Age and Survivors Insurance and the Federal Disability Insurance Trust Funds 1994, table III.B1.

age. We make separate projections of lifetime earnings by gender for persons born from 1931 through 1955.

Strong relationships predicting the mean indexed monthly earnings level (MIME) in the 10-year period 1984-93 are found. We estimate a model using only socioeconomic and demographic survey information to predict the next 10 years of earnings. Predictions are improved by using the record of past earnings from the SSA administrative record. One set of models relies exclusively on the information in the Social Security administrative record. To improve the predictive power of the models marginally, we add basic socio-demographic survey data in the 1984 SIPP panel. To a large extent, past is prologue: the last 10 years of earnings are good predictors of the next 10 years of earnings. The relationships from our preferred models are used to project earnings until retirement using data from the 1990 SIPP panel matched with SSA earnings records.

II. Modeling Lifetime Earnings

Modeling earnings requires consideration of both theoretical and statistical issues.

A Framework

Conceptually, the determinants of lifetime earnings are (definitionally) the determinants of wage rate (the price for each unit of work) and lifetime hours of work. To develop empirical models, attention must be paid to individual characteristics and external factors that reflect the circumstances under which the person works.

Individual factors include characteristics specific to the individual that affect productivity. Broadly speaking, these can be considered "human capital" and can be represented empirically by education and labor market experience. Other individual characteristics, such as sex and age, may reflect human capital or labor market discrimination. Some characteristics reflecting individual productivity, such as motivation and intelligence, are important but difficult to measure.

Although we will not develop a full model of earnings for this article, the economic and sociological literature suggest several variables that influence labor supply, for example, wage rate, gender, and marital status.

Estimating the model involves several complexities. First, there are important person-specific characteristics that do not vary year-to-year, while others do. For example, some of the determinants of wage rates such as education will not vary from year-to-year for individuals. Other determinants, like work experience, which reflect labor supply decisions, may change from year-to-year. Second, many important variables are often not measured in surveys, including motivation, intelligence, and attitudes toward work. Third, what you did last year affects what you do this year, which could be important even after accounting for independent factors (Heckman and Willis 1977).

We developed a reduced form model where earnings observed in earlier years are used to predict future earnings.[7] Earnings in earlier years reflect the effect of the variations in human capital, labor supply, and wage rates. Other selected variables are added to determine whether there is additional explanatory power in these variables beyond what is reflected in the earnings history.

Empirical Specification

We use four approaches to predict average monthly earnings for the 1984-93 decade. First, the survey-data-only approach predicts average monthly earnings based on survey information available in 1984, using a typical cross-sectional model. Second, the current-earnings approach projects the average monthly earnings over the decade based on 1983 average monthly earnings from administrative records, the most recent year before the 1984-93 prediction period. This is similar to using last year's earnings, often the only available earnings measure in surveys. Third, the administrative data (AD) approach predicts monthly earnings for 1984-93 based on 1974-83 mean indexed monthly earnings (MIMEs) from administrative records. Fourth, the total approach bases predictions on regressions using survey information as well as the administrative records of earnings from the previous decade. We wage index earnings to 1988 following the procedure used for calculating Social Security benefits (Bondar 1995).[8]

The prediction model with only survey variables includes several basic characteristics available from the matched SIPP survey. These include marital status (married with spouse present, the "other" category is omitted); minority status (black, white Hispanic, the "other than black or white Hispanic" category is omitted); years of education; employment in manufacturing (other industries make up the omitted category); employment in major occupations (managerial, professional and technical, sales, clerical, and administrative support, the other occupation is the omitted category), and job experience (reported years working in the type of work in the 1984 current or most recent job).[9] This model also includes dummy variables for birth cohorts that estimate the difference in mean earnings in the next 10-year period between the specified cohort and the reference 1931-35 cohort. As discussed by Glenn (1976), differences among birth cohorts may reflect aging effects (life-cycle stages), cohort effects (differences unique to the life experiences of one cohort compared with another cohort), and period effects (differences between the 1974-83 decade and the 1984-93 decade).

The predictors in the AD models include earnings in 1983, the number of years with maximum taxable earnings, and birth cohort. The models use the 1974-83 MIME as a predictor. The models do not rely on the match to the SIPP, and, therefore, could be estimated using SSA administrative data alone.

The AD and total approaches use lagged earnings from SSA administrative records. Earnings in 1983 are the most recent before the prediction period. The AD and total models use two lagged-earnings variables as independent variables to predict future taxable earnings: indexed 1983 monthly earnings, and indexed earnings per month in the 1974-83 period.

These variables are included because we expect persons with higher average earnings in the past to have greater earnings in the future. These variables could represent the observed continuity in the labor force behavior of women (for example, see Shaw 1994; Nakamura and Nakamura 1994; Duleep and Sanders 1994).[10]

We use information from previous earnings, reflecting measurable and unmeasurable factors, to predict future earnings. Because the equations are used primarily for predictive purposes, it is not important that some of the potential explanatory ability of the nonearnings variables is dissipated by the inclusion of the lagged-earnings measures.

Years of earnings at the taxable maximum between 1974 and 1983 measure the effect of this constraint on measured earnings in the previous decade on the prediction of future taxable earnings. Persons with earnings reaching the taxable maximum have earnings greater than or equal to Social Security taxable earnings. This means that their taxable maximum rose when SSA rules increased the taxable earnings during the 1970s. Thus, because persons whose earnings are at or above the maximum are likely to have higher earnings in future years, we expect a positive regression coefficient on the number of years reaching the maximum.

The statistical models are ordinary least square multiple regression equations for the survey data only, AD, and total approaches. We estimate gender-specific models because of known gender differences in the structure of employment and earnings. The equations use sample weights from the longitudinal panel, normalized by the ratio of the weighted universe to the sample size. The normalization gives a better indication of statistically significant relationships, while taking into account the representative sampling.[11]

Data

We use the 1984 panel of SIPP matched to the SSA administrative record of earnings from 1951 through 1993 to estimate our earnings models. The 1984 panel consists of about 20,000 households drawn from the noninstitutionalized population. The SIPP, developed by the Census Bureau to better measure sources of income and participation in government income security programs, collects core information every 4 months for the 32 months in the panel's life. This includes labor force activity, and income and program participation of persons in the household. Periodically, the SIPP also collects information on special topics, such as assets and pension coverage, work histories, marital histories, and fertility histories.

In cooperation with the Census Bureau, the Social Security Administration has matched SSA administrative records to the 1984,1990, and 1991 SIPP panels (Iams 1993; Iams and Sandell 1995). Individual's survey responses were exactly matched to the administrative record of their own annual Social Security taxable earnings for 1951 through 1993. The match rate was approximately 95 percent in the 1984 panel, and at least 90 percent in the 1990 and 1991 panels.

The sample for this analysis, men and women born in 1931 through 1955, represents the population aged 65 or older in 2020. This group reaches Social Security early retirement age in 1993-2017. The 1931-40 cohort was born in the depression; the 1941-45 cohort was born during World War II; and the 1946-55 cohort was born in the first half of the baby boom. The population were aged 29 to 53 in 1984 and aged 39 to 63 in 1994 (table 4). For most of the sample, actual earnings are available through middle age. To have complete earnings until around retirement age (through age 60), we need to project 22 years of future earnings for the youngest persons in the sample, and do not need to project any earnings for the oldest persons (that is, we can use actual earnings).

To test our model predictions, we partitioned the 1984 SIPP panel into two subgroups: one subgroup was used for estimating the model, and the other to test the predictions.[12] There is no single criterion for selecting the best model for projecting earnings. Comparing the level of explained variation and comparing predictions with actual experience on an out-of-sample group are useful with microdata such as the SIPP (Nakamura, Nakamura, and Duleep 1990).

III. Estimation Results

Table 5 presents the means and standard deviations of variables for the modelling sample by sex. Men have higher Social Security earnings than women but similar age distributions. Major sex differences exist in job characteristics. While more women worked in clerical/administrative-support occupations, more men worked in manufacturing and in managerial occupations, and worked more years in their 1984 line of work.

The explained variation in the four approaches is presented in table 6. First, the model using only the survey variables explained about 25 percent of the variation in future earnings for men and women. This is typical of regression equation models based on socioeconomic and demographic characteristics available in cross-sectional surveys (McNeil and Lamas 1987; Jacobsen and Levin 1995). Second, 1983 earnings explained about 61 percent of the variation in 1984-93 earn-

Table 4.—Birth cohort ages in 1974, 1984, 1994, 2004, and 2020

Cohort birth years	1974	1984	1994	2004	2020
	Age in years				
1931-35 (early depression).....	39-43	49-53	59-63	69-73	85-89
1936-40 (late depression)......	34-38	44-48	54-58	64-68	80-84
1941-45 (World War II).........	29-33	39-43	49-53	59-63	75-79
1946-50 (early baby boom I)..	24-28	34-38	44-48	54-53	70-74
1951-55 (early baby boom II)	19-23	29-33	39-43	49-53	65-69

ings for women and 63 percent for men. Third, using only administrative variables including 10 years of lagged earnings (the AD model) explained 63 percent of the variance for women and 67 percent for men. Adding 10 years made only a marginal improvement. Fourth, adding SIPP variables (the total model) marginally increased the explained variance percentage from 63 to 64 percent for women and from 67 to 69 percent for men. Compared with the survey-variables-only model, models that included lagged earnings explained much more variation in future earnings.[13]

Most variables in table 6 were statistically significant in the expected direction. The estimated models are consistent with the human capital framework (the effects of education and labor market experience) and with research on earnings by sociologists and demographers. Both the coefficients of the earnings in 1974-83 and the earnings in 1983 are significant at the 0.01 level.

When earnings in 1974-83 are independent variables in the equation, the regression coefficients of the other independent variables can be interpreted as the effect of these variables on the change in earnings between the two periods. Having previous period earnings in the equation controls for that period's earnings. Therefore the regression coefficients of the other variables can be interpreted as the effects of these variables on the growth of earnings between the 10-year periods. For example, married-with-spouse-present women's earnings are expected to be $239 less per month than those of other women (column 2 of table 6). Using the information in column 4 of table 6, we can compare the 1984-93 earnings of married-with-spouse-present women with nonmarried women who have identical 1974-83 earnings. The married-with-spouse-present women will have earnings in the 1984-93 period of $47 per month less than otherwise identical nonmarried women. Thus, it can be concluded that married-with-spouse-present women are estimated to have an average $47 slower earnings growth than other women (column 4 of table 6). The negative job experience coefficient for men in the equations including previous earnings indicates that men with more job experience (at the same level of past earnings) had significantly lower growth in earnings.

We applied the prediction approaches from our study sample to a test sample to predict mean indexed monthly earnings levels (MIMEs) in 1984-93. Because the test sample was not used in model estimation, it is statistically valid as a test sample. Using this sample, we compared the effectiveness of the approaches. Table 7 presents the actual average, the predicted average, and the absolute level of prediction error for mean monthly earnings in 1984-93.

The total model with 1974-83 earnings gives the best prediction of earnings in the next decade. The actual average observed to date was closest to the predicted average with the models that included MIME—men, $1,757 vs. $1,760-$1,762; women, $813 vs. $834-$837. The SIPP model predicted higher earnings than observed. The current earnings approach predicted lower earnings than observed for women but higher earnings for men.

Table 5.—Means and standard deviations of variables in multiple regression equations for persons born in 1931-55 in the modeling sample, by sex

Variables	Women	Men
Earnings record variables:		
MIME 1984-93	$845.48	$1,837.38
	(884.90)	(1,196.49)
Proportion of years with earnings in 1984-93	0.67	0.82
	(.38)	(.30)
MIME 1974-83	$628.29	$1,678.08
	(661.07)	(922.47)
Proportion of years with earnings in 1984-93	.62	.87
	(.37)	(.26)
1983 earnings	$9,667.72	$24,076.42
	(11,660.64)	(16,543.33)
Number of years with taxable maximum	.03	.55
	(0.26)	(1.10)
1936-40 cohort	.17	.16
	(.37)	(.35)
1941-45 cohort	.18	.18
	(.39)	(.36)
1946-50 cohort	.24	.25
	(.42)	(.41)
1951-55 cohort	.26	.26
	(.44)	(.42)
Survey variables:		
Black	.11	.10
	(.32)	(.28)
White Hispanic	.05	.06
	(.23)	(.22)
Education (in years)	12.7	13.2
	(2.79)	(2.96)
Married, spouse present	.72	.77
	(.45)	(.40)
Manufacturing industry	.17	.26
	(.37)	(.42)
Manager	.08	.14
	(.27)	(.34)
Professional-technical	.15	.15
	(.36)	(.34)
Sales	.11	.10
	(.31)	(.29)
Clerical	.27	.05
	(.44)	(.20)
Job experience	5.75	10.3
	(6.71)	(7.89)
Normed sample size	5,500	4,716

[1] Standard deviations in parentheses. Sample size reflects adjustment for longitudinal panel weights normed to reflect the universe/sample size ratio. MIME is mean indexed monthly earnings over the reference period. Earnings are wage indexed to 1988 using SSA benefit procedure.

Source: 1984 SIPP matched to SSA earnings records.

Table 6.—Multiple regression equations predicting mean monthly indexed earnings (MIME) in 1984-93, by sex [1]

Variables	Women				Men			
	Current earnings model	Survey (SIPP) model	AD model (SER variables)	Total model (SIPP-SER variables)	Current earnings model	Survey (SIPP) model	AD model (SER variables)	Total model (SIPP-SER variables)
Intercept	270.59*	-256.09*	81.22*	-154.56*	453.32*	-497.21*	-87.81*	-539.04*
	(28.26)	(-4.12)	(4.07)	(-3.58)	(24.95)	(-5.41)	(-2.63)	(-8.90)
Earnings record variables:								
MIME 1974-83	...	...	.19*	.17*	...	...	.35*	.36*
		...	(9.39)	(8.10)	...	...	(16.63)	(17.43)
1983 earnings	0.06*	...	.05*	.05*	.06*	...	.04*	.04*
	(94.20)	...	(44.31)	(42.88)	(93.60)	...	(37.77)	(34.16)
Number of years at taxable maximum	...	...	23.46	11.3	...	...	35.87*	24.89**
	...	...	(.80)	(.40)	...	...	(3.44)	(2.45)
1936-40 cohort	...	148.20*	107.81*	88.39*	...	210.87*	187.73*	173.67*
	...	(4.02)	(4.18)	(3.47)	...	(3.90)	(5.25)	(4.99)
1941-45 cohort	...	277.58*	191.92*	167.39*	...	271.69*	276.64*	229.90*
	...	(7.66)	(7.58)	(6.66)	...	(5.00)	(7.87)	(6.58)
1946-50 cohort	...	332.03*	247.69*	209.97*	...	413.69*	433.04*	372.84*
	...	(9.50)	(10.33)	(8.68)	...	(7.91)	(13.15)	(11.04)
1951-55 cohort	...	314.27*	182.93*	146.97*	...	376.93*	556.43*	498.43*
	...	(9.06)	(7.74)	(6.09)	...	(7.00)	(16.62)	(14.14)
Survey variables:								
Black	...	-48.28	...	-29.75	...	-339.26*	...	-89.61*
	...	(-1.42)	...	(-1.27)	...	(-6.41)	...	(-2.62)
White Hispanic	...	-62.40	...	-48.88	...	-117.51	...	-16.54
	...	(-1.32)	...	(-1.49)	...	(-1.76)	...	(-.38)
Education (in years)	...	45.70*	...	21.22*	...	89.19*	...	31.41*
	...	(10.06)	...	(6.72)	...	(14.60)	...	(7.87)
Married, spouse present	...	-238.93*	...	-47.06*	...	408.95*	...	122.37*
	...	(-10.00)	...	(-2.18)	...	(10.94)	...	(5.04)
Manufacturing industry	...	331.59*	...	26.70	...	709.37*	...	121.10*
	...	(11.71)	...	(1.33)	...	(20.08)	...	(5.08)
Manager	...	579.71*	...	99.77*	...	521.84*	...	162.86*
	...	(13.56)	...	(3.30)	...	(10.62)	...	(5.10)
Professional-technical	...	428.52*	...	162.85*	...	421.05*	...	303.61*
	...	(11.50)	...	(6.26)	...	(8.15)	...	(9.09)
Sales	...	131.56*	...	11.04	...	469.61*	...	74.10**
	...	(3.56)	...	(.43)	...	(8.71)	...	(2.12)
Clerical	...	230.07*	...	85.91*	...	-133.73	...	22.3
	...	(8.53)	...	(4.58)	...	(-1.84)	...	-.48
Job experience	...	38.37*	...	.43	...	22.97*	...	-2.89*
	...	(23.89)	...	(.35)	...	(11.01)	...	(-2.10)
Adjusted explained variance	.61	.25	.63	.64	.62	.24	.67	.69

[1] Survey model contains only SIPP variables; AD model contains only administrative record variables; total model contains both SIPP and record variables.

* Indicates significance at the 0.01 level of confidence.

** Indicates significance at the 0.05 level of confidence.

Note: T test ratios are in parentheses below the coefficient. Multiple regression equations are weighted with normalized weights from the longitudinal panel.

Source: 1984 SIPP matched to Social Security Administration summary earnings records.

The mean absolute error level for the total model (with survey and administrative variables) was lower than the other models by a small margin. The mean absolute prediction error was $362 for women and $501 for men (table 7). The least successful approach clearly was the survey-variables-only model, typical of cross-sectional data: the mean error was $581 and $897 respectively for women and men. Predictions based only on 1983 (indexed) earnings also were higher for men than the total approach with 1974-83 earnings.

We also estimated the average absolute predicton error for selected subgroups defined by sex and marital status, education, minority status, and birth cohort (table 8). The average error was much higher for the survey only than other models within all subgroups. The average error only slightly varied across the other models within subgroups, but it was usually lower for the total model or the current earnings model. Using the total model, average error was lowest among divorced women and men, never-married women, and married men. The total method also averaged the lowest errors among women with 12 to 15 years of education and all education groups of men other than those with 13-15 years. The total model also averaged the lowest error among white Hispanic women, women born in 1951-55, and men born in 1941-55. Errors were slightly higher for the 1931-35 cohorts of men and women for the total model than for the current earnings and administrative earnings models. Thus, the total model seems the most successful of the four models tested in terms of predicting earnings for each subgroup.

We prefer predicting the next 10 years of earnings with total models using 1974-83 earnings from the previous decade, although our analysis to date does not definitively establish it as the best projection approach. Among the approaches considered, the level of prediction error is among the lowest in important subgroups of women and overall. The total approach incorporates survey characteristics that are useful if socio-demographic characteristics change in later time periods, and the better information is provided by the SSA earnings histories. The average earnings levels and prediction errors for models with 1974-83 earnings were better than both the survey-only approach and the current earnings-only approach with men. Although using 1983 (indexed) earnings with men provides decent projections for the next decade, the low correlation between current and previous lifetime earnings previously discussed and shown in table 2 merits caution in using this approach.

The birth cohort parameters exhibit an important pattern: earnings growth between the 10-year periods were generally much higher among more recent cohorts. The ages of the cohorts in the various calendar years are shown in table 4. Persons earlier in their work lives increased their earnings more than persons observed later in their work lives. This is expected if less experienced workers have greater human capital increases, which will be reflected by higher wage rate increases (holding hours of work constant). In addition, some groups changed their labor supply. The 1951-55 cohort of women was in the prime child-rearing years in the 1974-83 period. Their reduced labor supply in this period probably influenced their earnings more in that 10-year period than in subsequent periods (Iams and Sandell 1994).

IV. Projections[14]

We use the total model to estimate earnings in the future. We determine the pattern of earnings over the known 10 years using the same pattern to estimate earnings until age 62. As in any projections, this assumes that the relationships that we observed during the period we estimated the model, 1974 to 1993, will hold in the future.[15] These relationships include estimates of life-cycle earnings patterns, as well as effects of education and other independent variables on earnings. Although we cannot test these assumptions, they remain our best guesses about the future.

To project the MIME until retirement we use the models described in table 6, estimated using the full 1984 sample.[16] The estimated coefficients are combined with individuals' survey responses in the 1990 panel of the SIPP and their administrative earnings from 1984-93 to predict earnings in the 10-year period from 1994-2003. For the youngest cohorts, earnings are also projected from 2004 until 2017 using the applicable dummy variable coefficients.[17] Average monthly earnings for the 10-year interval are used for either the full 10-year periods 1994-2003 and 2004-13 or the partial period, as applicable. These projected earnings are added to the

Table 7.—Predicted mean indexed monthly earnings and absolute error in 1984-93, using a test sample, by sex[1]

Sex and earnings variable	Predicted mean earnings	Mean absolute error
Women:		
Actual mean	$813	...
Current 1983 earnings	790	$371
Survey (SIPP) variables[2]	842	581
AD model (earnings record variables)[2]	837	370
Total model (SIPP-earnings record	834	362
Men:		
Actual mean	1,757	...
Current 1983 earnings	1,890	580
Survey (SIPP) variables[2]	1,824	897
AD model (earnings record variables)[2]	1,762	516
Total model (SIPP-earnings record	1,760	501

[1] Means estimated with normalized sample weights from the longitudinal panel using the first balanced replicate. Earnings wage indexed to 1988.

[2] Based on predictions from regression equations in table 5. Survey model contains only SIPP variables; AD model contains only administrative record variables; total model contains both SIPP and record variables.

Source: 1984 SIPP matched to SSA earnings records from 1951-93.

observed earnings from age 22 through 1993 to obtain the total lifetime earnings and to calculate the MIME. The projection does not incorporate any future (expected or unexpected) changes in real wages.[18]

We interpret the estimated coefficients for the cohort dummy variables as reflecting stage-of-work-life (aging) differences in earnings. The changes between two periods are assumed to be similar for identical age groups.[19] For example, the change between time one and time two for persons aged 41-45 at time one is used to forecast the change between time two and time three for persons aged 41-45 at time two. This "aging approach" applies the 1984 model's cohort dummy variable parameters for identical age groups in 1994. (See table 2 for other comparisons.) Because the oldest cohort was

Table 8.—Mean absolute prediction error in 1984-93 by selected characteristics using a test sample[1]

Characteristics	Current 1983 earnings model	Survey variables model	AD model (earnings record variables)	Total model (SIPP-earnings record variables)
Women	$371	$581	$370	$362
Marital status:				
Married	353	578	370	362
Divorced	377	629	380	367
Never married	384	583	383	376
Education (in years):				
0-11	236	402	284	248
12	331	503	332	326
13-15	398	601	392	389
16 or more	498	895	514	518
Black	290	576	326	313
White Hispanic	355	464	368	345
Birth cohort:				
1931-35	254	558	281	292
1936-40	298	597	336	331
1940-45	350	623	372	357
1946-50	383	602	409	392
1951-55	432	534	402	393
Men	580	897	516	501
Marital status:				
Married	545	920	533	518
Divorced	417	829	422	414
Never married	428	799	475	454
Education (in years):				
0-11	431	706	413	400
12	497	849	486	476
13-15	513	925	505	514
16 or more	609	1,058	629	589
Black	379	801	414	401
White Hispanic	460	855	505	490
Birth cohort:				
1931-35	569	973	495	509
1936-40	509	926	509	512
1940-45	499	910	504	497
1946-50	507	849	513	485
1951-55	525	829	545	511

[1]Means estimated with normalized sample weights from the longitudinal panel using the first balanced replicate. Earnings wage indexed to 1988. Based on predictions from regression equations in table 5. Survey model contains only SIPP variables; AD model contains only administrative record variables; total model contains both SIPP and record variables.

Source: 1984 SIPP matched to SSA earnings records for 1951-93.

approaching retirement in the 1984 model, the best estimate for cohorts approaching retirement would be the parameter for the 1931-35 cohort, persons aged 49-53 in 1984 and 59-63 in 1994. The 1951-55 cohort does not reach retirement age at the end of the second 10-year period. For these persons, we project average earnings from 2014-17 for a third 10-year period and use the average for those extra years. Thus, we project the last earnings of the 1951-55 cohort for the period 2014-17 using the dummy variable effect for the 1931-35 cohort who were near retirement.

The aging approach suffers from some of the problems of using cross-sectional, age-earnings estimates, which we criticized in the introduction. But there are important differences in our projection technique. First, we use the individual's own (observed) earnings from age 22 through 1993, and project only earnings from 1994 until retirement. Of the 38 years used in the final estimate, we project as many as 22 years for the youngest individuals and do not project any years for the oldest persons. Second, our method uses the previous 10 years of earnings (as well as other information) to estimate future earnings. Most cross-sectionally based estimates do not use prior earnings history information for individuals or estimates of cohort differences in earnings. Because the individual lifetime projections include about 20 to 30 years of actual earnings, these projected values include the diversity of individuals' lifetime experience.

Table 9 presents the observed MIMEs from ages 22 through the year 1993 and the projected MIMEs through age 60. Overall, the average of projected MIMEs to age 60 are slightly higher than the average of observed MIMEs through the year 1993. For men, the average of observed MIMEs was $1,700, compared with $1,721 projected with Summary Earnings Records (SER) variables and $1,730 projected with SIPP-SER variables. For women, the observed MIME was $814 compared with projected MIMEs to age 60 of $890 and $895, respectively. Thus, the inclusion of projected future earnings has a much larger effect for women than for men. Our projection of unobserved, future earnings until retirement, compared with earnings to date, increases the estimates of lifetime earnings. The average MIME increases about $21-$30, or about 1 percent of the observed $1,700 for men, and by $76-$81, or about 10 percent for women. Differences for men are largest for the baby boom cohorts. For women, the baby boom cohorts and women born during World War II have higher predicted MIMEs than reflected by their earnings observed through 1993. The average of observed and projected MIMEs are quite similar among the depression cohorts because few projected years are added to observed earnings.

VI. Conclusion

A recent report of the Committee on National Statistics (1995, p. 3), calls for development of a simulation model "to take account of all, or at least the main, sources of retirement income support." It also calls for greater use of administrative data matched to survey data for simulation. This article begins this process using the Survey of Income and Program Participation panels matched to Social Security Administration records of covered earnings from the SER.

Comparison of our modeling effort to the common social science approach of applying age-earnings profiles from a cross-sectional data base or estimating lifetime earnings as a mathematical function of current earnings is informative. As shown in table 2, current-year earnings yield inaccurate estimates of lifetime Social Security earnings. This partly

Table 9.—Mean indexed monthly earnings (MIME) at age 22 through 1993 and projected to age 60 for retired workers, by sex and birth cohort [1]

Sex	Birth cohort	Actual observed MIME to 1993	MIME projected to age 60: With administrative variables	MIME projected to age 60: With administrative and survey variables
Women, total	1931-55	$814	$874	$879
	1931-35	614	616	616
	1936-40	726	745	746
	1941-45	763	820	827
	1946-50	883	963	959
	1951-55	924	1,009	1,024
Men, total	1931-55	1,700	1,725	1,733
	1931-35	1,579	1,576	1,576
	1936-40	1,720	1,702	1,708
	1941-45	1,785	1,778	1,789
	1946-50	1,748	1,792	1,804
	1951-55	1,651	1,716	1,724

[1]Retired workers defined as respondents, with at least 25 percent of years with taxable earnings in years ages 22–60.

Source: Social Security Administration (SSA) calculations using the 1990 longitudinal panel of the SIPP matched to SSA earnings records. See text for methodology.

reflects changes in women's employment and changes in the extent of covered earnings taxed by Social Security.

Our main objective was a reasonable approximation of lifetime earnings taxable for Social Security purposes for persons retiring over the next 25 years. Using the 1984 SIPP panel, we estimated regression models of Social Security earnings for a 10-year period from 1984-93 based on variables from the SER and SIPP. Our models explain a high proportion of variation for the 10-year interval. We tested our four models' predictions on a test subsample from the 1984 SIPP. Based on the test sample predictions, we conclude that using 10 years of administrative earnings data is preferred to projections based only on current earnings or regression models using only survey variables. We next projected Social Security earnings until retirement using the characteristics and administrative earnings from the 1990 SIPP panel and the models based on the 1984 SIPP panel. We combined projected future earnings with actual earnings from age 22 through the year 1993. From this, we derived the lifetime mean indexed monthly earnings level. Projected average earnings are about 1 percent higher than the average observed through 1993 for men, and about 10 percent higher for women. This reflects the increase in maximum taxable earnings and increased earnings among women.

The addition of SIPP survey variables to models with administrative earnings variables only marginally improves the fit of the models and the accuracy of predictions from the models, at least for our sample that uses earnings from the prime work years. This suggests that Social Security benefits can be projected with administrative data alone.

Additional refinements could improve the estimation. Whereas this article collapsed the detailed earnings information available in SSA's administrative records into 10-year averages, individual years of earnings and patterns of earnings could be used. This article's projections do not incorporate projections of future changes in real wages, labor supply, or other important factors. Different functional forms could be incorporated and more sophisticated econometric techniques could also improve the estimation. As future earnings become available, we can test how well the current 10-year period predicts future 10-year periods.

Notes

[1] Earnings in this article always refer to earnings taxed for Social Security purposes unless otherwise specified. We use the Social Security record of annual, taxable earnings from 1951-93 because it provides a longitudinal record of covered earnings over this period. The data are calculated from Social Security earnings records matched to the 1990 panel of the Survey of Income and Program Participation. In table 1, reported 1980 earnings were adjusted to reflect the increase in the Social Security wage index between 1980 and 1990.

[2] The data in table 2 represent persons born in 1931 through 1958 who had covered earnings in at least one-quarter of the years elapsed after age 22 (see note 13) and who are expected to receive Social Security retired-worker benefits when they become age eligible.

[3] The estimate of covered earnings to date for each respondent was constructed using the respondent's own record from SSA of annual, taxable earnings in covered employment for the 1951-93 period. The annual earnings were wage indexed to 1988 dollars using the procedure for calculating a Social Security benefit (Bondar 1995; Social Security Administration 1993). Average earnings per month in the interval was estimated as total indexed earnings taxed by Social Security divided by the months elapsed from the year an individual was age 22 through 1993.

[4] The higher correlation for recent cohorts reflects in part the fact that 1993 earnings is a larger portion of earnings from age 22 through 1993 for more recent than earlier cohorts. For example, a person age 32 has 10 work years and 1993 earnings are one-tenth of the total. Similarly, a person age 62 has 40 work years and 1993 earnings are one-fortieth of the total.

[5] For a discussion of the rising labor force activity of more recent cohorts, see Hill and O'Neill (1990), Ferber (1994), or Goldin (1990).

[6] Although the choice of 10 work years as an interval is somewhat arbitrary, it matches the interval for Social Security covered earnings used in the Health and Retirement Survey. Ten-year administrative earnings summary data have been appended to the Health and Retirement Survey restricted public-use data (Mitchell, Olson, and Steinmeier 1996).

[7] A reduced-form model is one that does not develop all of the intermediate relationships, such as the separate effect of education on the wage rate and on annual hours of work. Further research could improve the model.

[8] The year of indexation is arbitrary and does not affect the results. We index to 1988—the index year in the 1990 Primary Insurance Amount calculation—the basic benefit for new retirees in 1990.

[9] We do not explicitly include measures of the presence of young children. See Iams and Sandell (1994) for a discussion of trends in full-time caregiving of young children. Because the lagged earnings variable measures earnings over a decade for persons at least in middle age, the variable incorporates effects from raising young children in the 1974-83 period.

[10] Annual earnings measures could be used to refine and potentially improve the estimates. These refinements will be considered in extensions of this research.

[11] Significance tests with larger samples can detect small differences as unlikely by chance (Hays 1981, chapter 10). We apply weights reflecting the observed sample size and the proportions in the U.S. population.

[12] The study sample consisted of 5,119 men and 5,548 women. Among the men, 16 percent were in the 1931-35 cohort; 16 percent were in the 1936-40 cohort; 18 percent were in the 1941-45 cohort; 25 percent were in the 1946-50 cohort; and 26 percent were in the 1951-55 cohort. The respective percentages for women were 15 percent, 17 percent, 18 percent, 24 percent, and 26 percent. The prediction sample consisted of 1,735 men and 1,908 women.

We exclude persons we know have died subsequent to the 1984 survey to avoid estimating their lifetime earnings. Unfortunately, we can only exclude persons who received Social Security benefits and subsequently died.

[13] Other approaches might be considered, such as a specific-effects model by constructing a dummy variable for each individual.

However, person-specific dummy variables cannot be used to improve projections in other samples with different persons.

[14] We project earnings only for persons expected to be fully insured for retired-worker benefits—those with nonzero earnings in one-fourth of the elapsed years. Because Social Security provides retirement benefits to fully insured workers, we also projected retired-worker status. We identified retired workers based on years with earnings observed from age 22 through 1993 and years of earnings projected from 1994 through age 60 (Iams and Sandell 1996). Retired workers have at least 25 percent of years with earnings from age 22 through age 60 (Iams and Sandell 1995).

[15] The appropriateness of this assumption cannot be tested until data on future earnings are available for our sample.

[16] The models estimated with the full sample are almost identical to the total model in table 6. The estimates are available upon request.

[17] The MIME from ages 22-60 is similar to the Average Indexed Monthly Earnings (AIME), part of the Social Security retired-worker benefit calculation. The AIME for benefits is the 35 years of highest earnings. Annual taxable earnings from ages 21-60 are wage indexed before calculating the average per month in the required 35-year computation period for benefits. (See Bondar 1995; SSA 1993.) However, higher earnings after age 60 can be substituted for lower earnings during that period.

In order to estimate the total earnings until retirement, we added the indexed earnings observed through 1993 to the indexed earnings projected in future years until the year of reaching age 61. We projected earnings through 2003 using SIPP and SER variables, presented in table 6. In order to make more precise estimates, we calculated the model using the entire 1984 sample by merging the modeling and test subsamples (equations available upon request). The full-sample model coefficients were very similar to those presented in table 6.

[18] Of course the results could be adjusted to reflect future changes in real wages, labor supply, or other important factors. These adjustments are beyond the scope of this article.

[19] A cohort-specific interpretation is an alternative. We believe that this approach would overstate growth in earnings in early and late stages of the work life. The cohort approach assumes that birth cohort effects on earnings growth observed between the 1974-83 and 1984-93 periods will be duplicated for the cohorts in the period between 1984-93 and 1994-2003. In other words, the effects "belong" to the birth cohort. But we believe the large cohort coefficient for persons in their 20s and 30s reflects their movement into "career" jobs and the sharply increasing human capital among inexperienced workers (Bianchi 1995, table 3.9; Iams 1993, tables 2 and 3). Such dramatic increases are unlikely to continue in middle age, when jobs and human capital are more stable. The cohort approach also would overestimate earnings change at older age by assuming the growth effects for middle aged persons apply to cohorts moving into their retirement years when labor supply and earnings are reduced.

The third classic interpretation for changes associated with time and age group is a period effect, reflecting different time periods. We believe that period differences are indirectly taken into account by other variables specified in the model, such as job characteristics and wage indexing of covered earnings.

Bibliography

Bianchi, Suzanne M. 1995. "Changing Economic Roles of Women and Men." *State of the Union: America in the 1990s. Volume I, Economic Trends.* Reynolds Farley (ed.). New York: Russell Sage Foundation, pp. 107-154.

Bondar, Joseph. 1995. "How to Compute a Retired-Worker Benefit." *Social Security Bulletin,* Vol. 58, No. 1 (Spring), pp. 76-86.

The Board of Trustees, Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds. 1994. *1994 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds.* Washington, DC: U.S. Government Printing Office.

Bye, Barry V. and Salvatore J. Gallicchio. 1988. "A Note on Sampling Variance Estimates for Social Security Program Recipients From the Survey of Income and Program Participation." *Social Security Bulletin,* Vol. 51, No. 10 (October), pp. 4-21.

Committee on National Statistics. 1995. "Toward Improved Modeling of Retirement Income Policies: Interim Report." Commission on Behavioral and Social Sciences and Education. National Research Council. Washington, DC.

Congress of the United States Congressional Budget Office. 1993. *Baby Boomers in Retirement: An Early Perspective.* Washington, DC: U.S. Government Printing Office.

Duleep, Harriet Orcutt and Seth Sanders. 1994. "Empirical Regularities Across Cultures." *The Journal of Human Resources,* Vol. 29, No. 2 (Spring), pp. S328-S347.

Eller, T.J. and Wallace Fraser. 1995. "Asset Ownership of Households: 1993." *Current Population Reports. Household Economic Studies.* Washington, DC: U.S. Bureau of the Census. Report No. P70-47.

Ferber, Marianne A. 1993. "Women's Employment and the Social Security System." *Social Security Bulletin,* Vol. 56, No. 3 (Fall), pp. 33-55.

Glenn, Norval D. 1976. "Cohort Analysts' Futile Quest: Statistical Attempts to Separate Age, Period, and Cohort Effects." *American Sociological Review,* Vol. 44, pp. 900-904.

Goldin, Claudia. 1990. *Understanding the Gender Gap: An Economic History of American Women.* New York: Oxford University Press.

Groves, Robert M. 1989. *Survey Errors and Survey Costs.* New York: *John Wiley and Sons.*

Hayes, William L. 1981. *Statistics.* Third Edition. *New York: CBS College Publishing.*

Heckman, James J. and Robert J. Willis. 1977. "A Beta-logistic Model for the Analysis of Sequential Labor Force Participation by Married Women." *Journal of Political Economy,* Vol. 85, No. 1, pp. 27-58.

Hill, Ann M. and June E. O'Neill. 1990. "A Study of Intercohort Change in Women's Work Patterns and Earnings." New York: Center for the Study of Business and Government, Baruch College.

Iams, Howard M. 1993. "Earnings of Couples: A Cohort Analysis." *Social Security Bulletin,* Vol. 56, No. 3 (Fall), pp. 22-31.

____. 1995. "A Comparison of SIPP and CPS Pension Measures." *Social Security Bulletin,* Vol. 58, No. 4 (Winter), pp. 125-130.

Iams, Howard M. and Steven H. Sandell. 1994. "Changing Social Security Benefits to Reflect Child-Care Years: A Policy Proposal Whose Time Has Passed?" *Social Security Bulletin*, Vol. 57, No. 4 (Winter), pp. 10-22.

____. 1996. "Past is Prologue: Simulating Lifetime Social Security Earnings for the Twenty First Century." *Proceedings from the 1996 Annual Research Conference.* Washington, DC: U.S. Bureau of the Census.

Jacobson, Joyce P. and Lawrence M. Levine. 1995. "Effects of intermittent labor force attachment on women's earnings." *Monthly Labor Review*, Vol. 118, No. 9 (September), pp. 14-19.

Lumsdaine, Robin L., James H. Stock, and David A. Wise. 1995. "Why Are Retirement Rates So High at Age 65?" Working Article No. 5190. Working Paper Series. Cambridge, MA: National Bureau of Economic Research, Inc.

____. 1994. "Pension Plan Provisions and Retirement: Men and Women, Medicare, and Models." *Studies in the Economics of Aging.* David A. Wise (ed.). Chicago: The University of Chicago Press, pp. 183-212.

McNeil, John M. and Enrique J. Lamas. 1987. "Male-Female Differences in Work Experience, Occupation, and Earnings." *Current Population Reports.* Series P-70, No. 10. Washington, DC: Bureau of the Census.

Mitchell, Olivia; Jan Olson; and Thomas Steinmeier. 1996. "HRS/AHEAD Documentation Report: Construction of the Earnings and Benefits File (EBF) for Use with the Health and Retirement Survey." Ann Arbor, MI: The University of Michigan, Institute for Social Research.

Nakamura, Alice Orcutt and Masao Nakamura. 1994. "Predicting Female Labor Supply: Effects of Children and Recent Work Experience." *The Journal of Human Resources.* Vol. 29, No. 2 (Spring), pp. 304-327.

Nakamura, Alice Orcutt; Masao Nakamura; and Harriet Orcutt Duleep. 1990. "Alternative Approaches to Model Choice." *Journal of Economic Behavior and Organization.* Vol. 14, pp. 97-125.

Sandell, Steven H. and Howard M. Iams. 1997. "Reducing Women's Poverty by Shifting Social Security Benefits from Retired Couples to Widows." *Journal of Policy Analysis & Management.* Forthcoming in Vol. 16, No. 2 (Spring), pp. 281-297.

Shaw, Kathyrn. 1994. "The Persistence of Female Labor Supply: Empirical Evidence and Implications." *The Journal of Human Resources*, Vol. 29, No. 2 (Spring), pp. 348-378.

Smith, James P. and Michael P. Ward. 1984. *Women's Wages and Work in the Twentieth Century.* Santa Monica, CA: Rand Corporation.

Social Security Administration. 1995. *Annual Statistical Supplement, 1995* to the *Social Security Bulletin.* Washington, DC: U.S. Government Printing Office.

____. 1993 *Social Security Handbook: 1993.* 11th Edition. Washington, DC: U.S. Government Printing Office.

U.S. Bureau of the Census. *Statistical Abstract: 1994.* Washington, DC: U.S. Government Printing Office.

Appendix: Test Sample Predictions

To use an independent subsample for testing, we identified the cases in the first primary sampling unit of the first balanced replicate created by the Census Bureau for variance estimation—approximately one-quarter of the sample (Bye and Gallechio 1988). The models were estimated using the remaining three-quarters of the sample cases (henceforth termed the "modeling sample"). We use the independent subsample for testing the model predictions from the modeling sample equations.

Prediction approaches were applied to the random replicate test sample to predict mean indexed monthly earnings levels (MIMEs) in 1984-93. Because the test sample was not used in model estimation, it is statistically valid as a test sample. Using this sample, we compared the effectiveness of the approaches. Table 7 presents the actual average, the predicted average, and the absolute level of prediction error in the mean monthly earnings in 1984-93. To estimate the absolute value of prediction error, we subtracted the predicted value from the actual value and ignored the sign of the difference. We estimate predicted earnings using current 1983 earnings, the survey variables-only model (table 6, columns 2 and 6), the administrative data (AD) model (table 6, columns 3 and 7), and the total model with baseline earnings (table 6, columns 4 and 8).

The total model with baseline earnings gives the best prediction of earnings in the next decade. The actual average observed to date was closest to the predicted average with the MIME model—men, $1,757 vs. $1,760-$1,762; women, $813 vs. $834-$837. The Survey of Income and Program Participation (SIPP) model predicted higher earnings than observed, and the current earnings approach predicted lower earnings than observed for women. The current earnings approach for men slightly underpredicted observed earnings.

The mean absolute error level was lower than the other models by a small margin using the total model with survey and administrative variables. The least successful approach clearly was the survey variables-only model, typical of cross-sectional data: The mean error was $581 and $897, respectively, for women and men. Predictions based only on 1983 (indexed) earnings also were higher for men than the total approach with baseline earnings.

We prefer predicting the next 10 years of earnings with total models using baseline earnings from the previous decade. The average earnings levels and prediction errors for models with baseline earnings were better than both the survey-only approach and the current earnings-only approach with men. Although using 1983 (indexed) earnings with men provides decent projections for the next decade, the low correlation between current and previous lifetime earnings previously discussed and shown in table 2 causes caution in using this approach. In addition, current earnings would overestimate future earnings of younger cohorts as they approach retirement with reduced labor supply and earnings.