THE SURVEY OF INCOMIE AND
PROGRAM PARTICIPATION

PAST IS PROLOGLUE: SIMULATING LIFETIME SOCIAL SECURITY EARNINGS FOR THE TWENTY-FIRST CENTURY

No. 214

Howard M. Iams & Steven H. Sandell
Office of Research & Statistics, Social
Security Administration

March 1996

U.S. Department of Commerce BUREAU OF THE CENSUS

PAST IS PROLOGUE:
Simulating Lifetime Social Security Earnings
for the Twenty-First Century
by

Howard M. lams and Steven H. Sandell
Office of Research and Statistics
Social Security Administration
March 1996

Abstract:

This paper 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 ten-year earnings intervals using the matched 1984 SIPP panel. We find strong relationships predicting the mean indexed monthly earnings level in the ten 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 twenty-first century.

One set of models relies exclusively on the information in the Social Security administrative records. We improve the predictive power of the models slightly by using basic socio-demographic characteristics from the 1984 SIPP.

KEYWORDS

Social Security benefits, lifetime earnings, projection, cohort

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 twenty-five years, we do not know what future retirees will have earned over their lifetimes. We cannot assume that the experience of current retirees applies to future retirees, because the work lives of current retirees differ markedly from those of future retirees born during World War II and the baby boom. 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 paper 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 and Statistics in the 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 aged 65 and older in

*ACKNOWLEDGMENTS

The views in this paper 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. We wish to thank the following persons for comments on earlier drafts: Harriet Duleep, Susan Grad, Alan Gustman, Kajal Lahiri, Michael Leonesio, Olivia Mitchell, Cheryl Neslusan, David Paftison, Ralph Smith, Kelvin Utendorf, Denton Vaughan, and Peter Wheeler.

2020, improving on current techniques using cross-sectional data. A clear advantage of our approach over conventional methods is a more accurate estimate of lifetime earnings reflecting the accumulated covered earnings
by each individual.

Some researchers have made longitudinal inferences using a series of cross-sectional surveys taken at different points of time with earnings and/or labor force participation rates for gender-specific age-groups (see, for example, Bianchi, 1995 pp. 117-122, and Smith and Ward, 1984). The difficulty with this approach, particularly for women, is that individuals enter and leave the labor force. As an illustration, Table 1 presents the average Social Security earnings for earners in 1980 and for earners in 1990. It also presents the average for people with earnings in both 1980 and 1990 and 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 $1 1,588. Thus, a series of cross-sections misrepresents the changing earnings of a cohort overtime. (Among men, the longitudinal change was only slightly higher than the cross-sectional change indicated.)

A second approach uses observed age-specific earnings to estimate age-earnings profiles. In this approach, earnings for the different age groups in a cross-sectional data set are used to represent earnings across the lifecourse of a "synthetic" cohort (see for example, Lumsdaine, Stock and Wise 1994 and 1995). The age-earnings profile is interpreted as earnings histories 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 current earnings for different cohorts. Unfortunately, if the earnings structure changes, the recent cohort's experience is a misrepresentation of an older cohort's experience when it was younger and vice-versa. Increased wage rates and labor supply for more recent birth cohorts of women, for example, 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 data from the 1990 longitudinal SIPP panel. Table 2 (columns 3 and 4 for men, and columns 6 and 7 for women) shows how misleading current (1993) earnings are as an indicator of the trend for lifetime earnings. The current year average is usually much higher than the lifetime average within each cohort, except for the 1931-35 cohort which is beginning retirement. If we use these current 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 1931-36 cohort. The lifetime average based on the cross-section of current earnings 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 between each individual's average lifetime monthly earnings to date (adjusted for economy-wide growth in earnings over their lifetimes) and monthly earnings in 1993. 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 current earnings 'explained' about 29 percent to 61 percent of the variation in observed lifetime earnings to date, depending upon gender and cohort. The explained variation was lower among older cohorts than for more recent cohorts.

A 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 (lams and Sandell 1995). Thus, a "synthetic" cohort approach to estimating future earnings overestimates the lifetime earnings of older cohorts. (For a discussion of the rising labor force activity of more recent cohorts see Hill and O'Neill (1990), Ferber (1994), or Goldin (1990)).

The amount of total annual earnings taxed by Social Security changed with the increase 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 Appendix A). The low maximum truncated the taxable earnings of higher earners. For example, about 51 percent of male and 87 percent of female workers had total earnings below the maximum in 1965, compared to 91 percent and 98 percent in 1992 (Social Security Administration 1993-. Social Security Administration 1995, Table 4.B.4). The rising maximum increased the amount of annual earnings taxed by SSA and included in retirement benefit calculations. This increased the average lifetime earnings of recent cohorts more than for cohorts born earlier. It affected the calculation of average earnings for men more than women.

This paper presents a simulation of lifetime Social Security earnings until retirement 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 ten-year earnings intervals using the matched 1984 SIPP panel. For the majority of people, aged 29-53 in 1984, this represents earnings in their career job. We then use the models to project (unobserved) Social Security earnings from 1994 through retirement. We combine the earnings we can observe for individuals through 1993 with earnings we project until retirement age. We make separate projections by gender for persons born from 1931 through 1955.

We find strong relationships predicting the mean indexed monthly earnings level in the ten-year period 1984-93. One set of models relies exclusively on the information in the Social Security administrative record. We improve the predictive power of the models marginally by using basic socio-demographic survey data in the 1984 SIPP panel. Apparently, past is prologue. 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

Conceptual Framework:

Modeling lifetime earnings requires consideration of both theoretical and statistical issues. Conceptually the determinants of lifetime earnings are (definitionally) the determinants of the wage rate (the price for each unit of work) and lifetime hours of work (labor supply). To develop empirical models, attention must be paid to individual behavior and characteristics, secular trends, and year-specific events as well as statistical properties. We can start with the identity:

1) Lifetime Earnings = years of work life (wage rate * hours per year)

With this in mind we can divide our conceptual analysis into the determinants of the wage rate and the determinants of labor supply. The determinants of the wage rate, in turn, include individual and external factors.

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 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.

External factors reflect the circumstances under which the person works. These include other inputs of production such as the capital stock and technology. In a competitive labor market economy wide productivity affects individuals'wages. This includes secular trends, such as the increase in productivity over time as well as cyclical circumstances, such as recessions. An individual will tend to have lower wages in poor economic environments than in more prosperous times. Obviously, the individual and external factors can interact.

The various dimensions of labor supply are an integral part of lifetime earnings. Hours per week, weeks per year and years over the lifetime are three dimensions that could be considered separately. However, for the purposes of the current analysis, the same factors affect all three.

Although it is not useful to develop a full model of labor force participation for this paper, the economic and sociological literature suggest several variables that influence labor supply. These include: wage rate, gender, marital status and presence of young children (especially for women), and societal trends.

Estimating the model involves several conundrums. First, there are important person-specific, year-to year correlations among many of the variables. For example, some of the determinants of wage rates will not vary from year to year for individuals. Other determinants, like experience, are a function of the labor supply decision. Second, many important variables are often not measured in surveys. These include motivation, intelligence and attitudes toward work. Third, "state dependence," the correlation between labor supply with earnings between periods that remains after taking into account independent factors could be important (Heckman and Willis 1977).

We develop a reduced form model where earnings in earlier years are used to predict earnings in later years. Earnings in earlier years reflect the effect of the factors discussed above. Selected other 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 the next decade of average monthly earnings, from 1984 through 1993. First, the current earnings approach projects the average monthly earnings over the next decade as equal to 1983 average monthly earnings, the most recent administrative earnings. Second, the survey data only approach predicts average monthly earnings based on survey information available in 1984, using a typical cross-sectional model. Third, the MIME approach predicts monthly earnings based on 1974-83 Mean Indexed Monthly Earnings. This bases predictions on regressions using survey information and the administrative records of earnings from the previous decade. Finally, based on 1974-83 earnings, the SIME approach separately models two components: the level of earnings [Selected Indexed Monthly Earnings including only years with earnings (SIME)] and the level of labor supply [the percentage of years with earnings]. In the SIME approach, the next decade of average earnings is estimated by the predicted annualized monthly earnings level times the predicted years of earnings. We wage index earnings to 1988 using the procedure for calculating Social Security benefits (see Bondar 1995).

The statistical models are ordinary least square multiple regression equations for the survey data only, MIME and SIME 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.

The survey data only prediction model includes several basic characteristics available from the matched SIPP survey. These include marital status (married spouse present, other omitted), minority status (black, white hispanic, omitted is other than black or hispanic), years of education, employment in manufacturing (other industries omitted), employment in major occupations (managerial, professional and technical, sales, clerical and administrative support, other omifted),and job experience (reported years working in the type of work in the 1984 current or most recent job). This model also includes dummy variables for birth cohort which estimate the difference in mean earnings in the next ten-year period between the cohort and the reference cohort (1931-35). As discussed by Glenn (1 976), differences between 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 current earnings, MIME, and SIME approaches used lagged earnings from SSA administrative records. Indexed 1983 monthly earnings is the most recent earnings level before the prediction period. As well as representing earnings level, this may reflect recent health conditions or work effort. The MIME and SIME models use two lagged earnings variables as independent variables to predict future earnings: indexed 1983 monthly earnings and earnings per month in 1974-83. They are included because we expect persons with higher average earnings to have greater earnings in the future. This is consistent with observed continuity in the labor force behavior of women (see, for example, Shaw 1994; Nakamura and Nakamura 1994; Duleep and Sanders 1994). Because the equations will be used primarily for predictive purposes, we do not care if some of the potential explanatory ability of the non-earnings variables is dissipated by the inclusion of the lagged earnings measures. We are able then to use information from previous earnings, reflecting measurable and unmeasurable factors, to predict future earnings. When earnings in the previous ten years are controlled, the regression coefficients can be interpreted as change in earnings between the two periods.

Years of earnings at the taxable maximum in 1974-83 measures the effect of the constraint in the previous decade of taxable earnings on the future earnings (with increased taxable maximums in the next decade). Persons with earnings reaching the taxable maximum have total earnings greater than taxable earnings. This means that their taxable earnings would rise when SSA rules increased the taxable maximum during the 1970s. Thus, we expect a positive signed regression coefficient on the number of years reaching the maximum.

The MIME and SIME regression approaches initially use administrative record variables and then add survey variables. The independent predictors in the administrative models include earnings in 1983, the number of years with maximum taxable earnings, and birth cohort. The single-equation version of the model uses the 1974-1983 MIME as a predictor; the components version of the model uses the 1974-83 mean level of Selected Indexed Monthly Earnings in years with earnings (SIME) and the 1974-83 proportion of years with earnings. These models do not rely on the match to the SIPP, and therefore could be estimated using SSA administrative data alone.

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 program participation in government income security programs, collects core information every four months for each of thirty-two months in the panel's life. This includes labor force activity, income and program participation of persons in the household. Periodically, the panel included special topics such as assets and pension coverage, work histories, marital histories, and fertility histories. The design of questions in the 1984 panel became the basis for new SIPP panel questions into the 1990's. The Census Bureau plans to extend the SIPP to 48 months with a
larger sample than previous panels.

In cooperation with the Census Bureau, the Social Security Administration has matched SSA

administrative records to the 1984, 1990, and 1991 SIPP panels (see lams 1993: lams and Sandell 1995). SIPP respondents' records were exactly matched to the administrative record of their own annual Social Security taxable earnings in 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 paper focuses upon men and women born in the period 1931-55. This sample represents the population aged 65 and over in 2020, first reaching Social Security retirement age in 1993 through 2017. They were aged 29 to 53 in 1984 and aged 39 to 63 in 1994 (Table 3). Actual observed earnings are available through middle age for most of the sample. In order to have complete earnings until around retirement age (through age 60), we need to project 22 years of future earnings for the youngest persons, and we need no projection for the oldest persons. The 1931-40 cohort was born in the depression: the 1941-45 cohort was born in World War II; and the 1946-55 cohort was born in the first half of the baby boom.

There are no clear criteria for selecting the better approach. Comparing the level of explained variation and comparing predictions on an out-of-sample group are useful with micro data such as the SIPP (Nakamura, Nakamura, and Duleep 1990). To test our model predictions, we partitioned the 1984 SIPP panel into two subgroups: one subgroup was used for estimating the model, and another subgroup was used to test the predictions. Because we preferred 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 (see Bye and Gallechio 1988). We estimated the models using the remaining three-quarters of the sample cases (henceforth termed the modeling sample). We use the first primary sampling unit, first balanced replicate subsample for testing the model predictions from the modeling sample.

III. Estimation Results

Table 4 presents the variable means and standard deviations of characteristics in the modeling sample separately by gender. The earnings record variables indicate that men have higher Social Security earnings than women but similar age distributions. Major gender differences exist in job characteristics: while more women worked in clerical-administrative support occupations, more men worked in manufacturing, worked in managerial occupations, and worked more years in their 1984 line of work.

The explained variation in three models can be compared (Table 5). First, the correlation of 1983 earnings with 1984-93 earnings was 0.70 for women and 0.79 for men suggesting that current 1983 earnings explained about 49 percent of the variation in future earnings for women and 62 percent for men. Second, the model using only the survey variables explained about 25 percent of the variation in future earnings. This is typical of regression equation models based on socioeconomic and demographic characteristics available in cross-sectional surveys (see for example, McNeil and Lamas (1 987) or Jacobsen and Levin (1 995)). Third, using only administrative variables including lagged earnings (the MIME model) explained 63 percent of the variance for women and 67 percent for men. Adding SIPP variables to the MIME models 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 only model, models with lagged earnings explained much more variation in future earnings.

The pattern of significant variables was very similar for the MIME models in Table 5 and the SIME models in Table 6A. Most variables were statistically significant in the expected direction. The estimated models are consistent with the human capital framework discussed above and with research on earnings by sociologists and demographers. Both the coefficients of the ten-year earnings and the 1983 earnings are highly significant. However, the coefficient for the 1983 earnings seems to have greater explanatory power. Perhaps this is predictive of future health conditions or work effort.

The interpretations of coefficients for the demographic variables are different depending on whether reviousearnings are used as an explanatory variable in the equation. When the previous earnings variables are included, the coefficients of the regressors are the effects of the variables after previous earnings are controlled. So they can be loosely interpreted as the effects of these variables on the growth of earnings between the ten-year periods. Married spouse present women, for example, exhibit on average $47.06 slower earnings growth than other women (column 3 of Table 5), and their earnings are expected to be $238.93 less per month than other women (column 1 of Table 5). The negative job experience coefficient for men in the equation with previous earnings means that men with more job experience had significantly lower growth in earnings. In contrast, without controlling for previous earnings, both men and women had significantly higher earnings with higher job experience.

The birth cohort parameters form an important pattern: earnings between the ten-year periods were generally much higher among more recent cohorts. The ages of the cohorts in the various calendar years are shown in Table 3. Persons earlier in their work life increased their earnings in the next decade more than persons later in their work life. 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 1951-1955 cohort of women was in the prime child-bearing years in the 1974-83 period. Their reduced labor supply in this period probably affected their earnings more than in subsequent periods (lams and Sandell 1994).

Predicting the proportion of years with earnings (Table 6B) was less successful than predicting earnings. If the proportion of years with earnings is bimodal (as is likely for women because some worked ten years and other worked no years in 1974-83), then we could not expect a high level of explanation. Our models explained about two-thirds of the variation in earnings, but only about two-fifths and half of the variation in proportions of years with earnings for women and men, respectively.

IV. Test Sample Results

We then applied the prediction approaches to the random replicate test sample to predict Mean Indexed Monthly Earnings (MIME) in 1984-1993. 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 single-equation and components of earnings 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 5, columns I and 4), the one equation MIME model with baseline earnings (Table 5. columns 3 and 6), and the two equation SIME model with baseline earnings (Table 6).

The one-equation MIME 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 (SIPP) model predicted higher earnings than observed, while the two-equation SIME model predicted much lower earnings than observed. The current earnings approach predicted much lower earnings than observed for women. The current earnings approach for men slightly underpredicted observed earnings.

The mean absolute error level was lowered by a small margin using the two-equation SIME model with baseline earnings. 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 MIME-SIME approaches with baseline earnings. The level of error seems similar among the current earnings approachfor women, MIME, and SIME although the two equation SIME model was slightly lower.

We prefer predicting the next ten years of earnings with MIME 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 lifetime earnings previously discussed and shown in Table 2 suggests 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.

Among the two approaches with baseline earnings, we prefer the one-equation MIME estimate. First, the models with the MIME approach more accurately predict the future earnings levels for both sexes. Secondly, the MIME approach gives similar levels of absolute error for individuals to the two-equation SIME approach. Perhaps a simultaneous equation approach would be a better implementation of the components model, but it would be difficult to implement for projections. In the components approach, the two equations were treated as if they were independent. That is faulty because the correlation between the proportion of years worked and SIME is positive. This positive correlation explains why the cohort-specific estimates of MIME are lower using the two equation method. Persons with a below average proportion of years worked are likely to have lower SIMEs and vice versa. The estimates do not reflect this positive correlation.

V. Projections

We interpret the estimated coefficients for the cohort dummy variables as reflecting the stage of work life (aging) differences in earnings. The changes between two periods are assumed to be similar for identical age groups. For example, the change for middle-aged persons between time one and time two is used to forecast the change for middle-aged persons between time two and time three. 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 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 "aging approach" suffers from some of the problems of using cross-sectional age-earnings estimates which we criticized in the introduction. There are important differences. First, we use the individual's own (observed) earnings from age 22 through 1993, and project only earnings from 1994 until retirement. Of the 39 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 the aging effect coefficient) to estimate future earnings. The cross-sectionally based estimates do not use prior information about individuals or about cohort differences in earnings.

Table 8 presents the observed Mean Indexed Monthly Earnings (MIME) from ages 22 through 1993 and the projected MIME through age 60. The projection does not incorporate any future(expected or unexpected) changes in wages. Overall, the average of projected MIMEs are slightly higher than the average of observed MIMEs with earnings through 1993. For men, the average of observed MIMEs was $1,700 compared with $1,725 projected with SER variables and $1,733 projected with SIPP-SER variables. For women, the respective MIMEs were $814 compared with $874 and $879. Thus, the inclusion of projected future earnings has a larger effect for women than for men. This indicates that our projection of unobserved, future earnings until retirement increases the estimates of lifetime earnings (compared to earnings to date). The average MIME increases about $25-33 or about 2 percent of the observed $1700 for men, and by $60-64 or about 7 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 (1 995, 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 paper begins this process using the Survey of Income and Program Participation (SIPP) panels matched to Social Security Administration records of covered
earnings from the Summary Earnings Record (SER).

Using the 1984 SIPP panel, we estimated regression models of Social Security earnings in 1984-93 based on variables from the SER and SIPP. Our models explain a remarkably high proportion of variation in ten-year earnings. We tested our model predictions against several alternatives on a test subsample from the 1984 SIPP. Based on the test sample predictions, we conclude that a single equation model predicting the Mean Indexed Monthly Earnings (MIME) in the next decade is preferred to projections with current earnings, with a survey variables only based regression model, and with a components approach predicting level of earnings and years with earnings separately. Additional research, such as experimenting with different functional forms or incorporating more sophisticated econometric techniques, could improve the models.

We next projected Social Security earnings until retirement applying the 1984 projection models with baseline earnings to the 1990 SIPP panel. Projected average earnings is slightly higher than the average observed through 1993, particularly for women. This is consistent with 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 predictions from the models, at least for our sample where earnings from the prime work-years are used for prediction. This suggests that lifetime earnings of individuals and their Social Security benefits could be projected adequately with administrative data alone by the Office of the Actuary.

Identification of husbands and wives (and divorcees and widows) is critically important to accurately model retirement benefits. In Social Security, auxiliary spouse and survivor benefits depend on the specific earnings of couples. Social Security records alone cannot identify marital status unless auxiliary benefits are received, usually in retirement. Thus, actuarial models should make use of SIPP data matched to administrative records.

Our modeling effort has implications for 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. Current earnings yield less accurate estimates of lifetime Social Security earnings. This partly reflects changes in women's employment and changes in the extent of covered earnings taxed by Social Security. We believe that simulations would be improved greatly by using actual earnings.

The SSA administrative earnings data cannot be released to the research community because of restrictions under privacy and tax legislation. However, the SIPP data would be enhanced by releasing several measures of lifetime earnings derived SSA earnings records. Our modeling effort suggests that the mean indexed monthly earnings in a decade, the proportion of years with earnings in a decade, and the lifetime average earnings through a recent year would enhance analysis of SIPP data.

We recommend that the Census Bureau and the Social Security Administration begin a process that would make the data publicly available to researchers. Recent efforts which consider appending aggregate earnings measures based on Social Security Administrative records to the Health and Retirement Survey could be expanded by appending such earnings to SIPP. Confidentiality protection is essential, but this could be possible using highly aggregated information.

Bibliography

Bianchi, Suzanne M. 1995. "Changing Economic Roles of Women and Men". pp. 107-154 in State of the Union: America in the 1990s. Volume 1 Economic Trends. edited by Reynolds Farley. NY,NY: Russell Sage Foundation.

Bondar, Joseph. 1995. "How to Compute a Retired-Worker Benefits". Social SecuritV 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: 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, D.C.

CONGRESS OF THE UNITED STATES CONGRESSIONAL BUDGET OFFICE. 1993. BabV Boomers in Retirement: An Early Perspective. Washington, D.C.

Duleep, Harriet Orcuft 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". U.S. Bureau of the Census. Current Population Reports. Household Economic Studies. 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. Volume 44, pp. 900-904.

Goldin, Claudia. 1990. Understanding the Gender Gap: An Economic HistorV 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. NY, NY: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.

. 1996. "A tale of two pension surveys: the SIPP and the CPS". Social Security Bulletin. forthcoming.

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.

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 Paper 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". in Studies in the Economics of Agin-g. edited by David A. Wise. 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, D.C.: Bureau of the Census

Nakamura, Alice and Masao Nakamura. 1994. "Predicting Female Labor Supply: Effects of Children and Recent Work Experience". The Journal of Human Resources. Vol. XXIX, Number 2 (Spring), pp. 304-327.

Nakamura, Alice Orcuft, Masao Nakamura, and Harriet Orcuft Duleep. 1990. "ALTERNATIVE APPROACHES TO MODEL CHOICE". Journal of Economic Behavior and Or_qanization. Vol. 14, pp. 97-125.

Sandell, Steven H. and Howard M. lams. Forthcoming. "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).

Shaw, Kathyrn. 1994. "The Persistence of Female Labor Supply: Empirical Evidence and Implications". The Journal of Human Resources. Vol. XXIX, Number 2 (Spring), pp. 348-378.

Smith, James P. and Michael P. Ward. 1984. Women's Wages and Work in the Twentieth Centu@. Santa Monica, Ca: Rand Corporation.

Social Security Advisory Council. 1995. Report. Washington, D.C.: Social Security Administration.

Social Security Administration. 1995. Annual Statistical Supplement, 1995 to the Social SecuritV Bulletin. Washington, D.C.

. 1993 Social Security Handbook: 1993. 1 lth Edition. Washington D.C.: U.S. Government Printing Office.

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

Appendix A: Social Security Taxable Maximum Changes

The Social Security maximum on taxable covered earnings truncates the covered earnings each year. In relation to average real earnings, this maximum reached a low point about 1965. This can be seen by dividing the National Average Earnings by the Taxable Maximum (see Table Al). The Taxable Maximum was approximately equal to the National Average Earnings in 1965, and was more than twice that after 1980. The covered earnings base for benefit purposes is covered earnings (truncated by) the Taxable Maximum. The Taxable Maximum was set by law prior to 1975 (SSA 1995, Table 2.A3). Thereafter, the maximum was subject to automatic adjustment.

Received: from hhesoas by smtp-gw2.census.gov (SMTPLINK V2.10.05) ;
Tue, 25 Jun 96 08:25:02 EST

Return-Path: <Howard.M.lams@s3abac8.ssa.gov>
Received: from it-relayl.census.gov by hhesoas (5.x/SMI-SVR4)
id AA23924; Tue, 25 Jun 1996 08:21:04 -0400

Received:from gate.census.gov (internet.census.gov [148.129.65.2]) by it-relayl.census.gov (8.7.5/8.7.3/vl.9) with SMTP id IAA22664 for <enriquel@hhesoas.hhes.census.gov>; Tue, 25 Jun 1996 08:21:39 -0400 (EDT)
From:Howard.M.lams@s3abac8.ssa.gov

Received:from info.census.gov by gate.census.gov with SMTP id AA03938 (InterLock SMTP Gateway 3.0 for <enriquel@hhesoas.hhes.census.gov>); Tue, 25 Jun 1996 08:21:38 -0400

Received:from s3abab9.ssa.gov (s3abab9.ssa.gov [199.173.224.2]) by info.census.gov (8.7.5/8.7.3) with SMTPid QAA29392 for <elamas@Census.gov>; Mon, 24 Jun 1996 16:18:27 -0400 (EDT)

Received:from s3abac8.ssa.gov by s3abab9.ssa.gov (AIX 4.1/UCB 5.64/4.03)
id AAl 5936; Mon, 24 Jun 1996 16:02:57 -0400

Received: from ccsmtpgw2.ssa.gov by ssa.gov (5.x/SMI-SVR4)
id AAl 7360; Mon, 24 Jun 1996 15:54:49 -0400

Received: from cemail by ccsmtpgw2.ssa.gov (SMTPLINK V2.1 1.01)
id AA835657568; Mon, 24 Jun 96 10:19:14 EST

Date:Mon, 24 Jun 96 10:19:14 EST
Message-Id:<9605248356.AA835657568@ccsmtpgw2.ssa.gov>
To:elamas@census.gov
Return-Receipt-To:Howard.M.lams@s3abac8.ssa.gov
Subject: Sipp Working Paper