The Effects of Differential Income Replacement and Mortality on U.S. Social Security Redistributions

Li Tan
Cory Koedel

First Draft: January 2017
This Draft: February 2018

We study redistributions via the United States Social Security retirement system for cohorts of men born during the second half of the 20th century. Our focus is on redistributions across race and education groups. The cohorts we study are younger than cohorts studied in previous, similar research, and thus more exposed to recent increases in earnings inequality. All else equal, this should increase the degree of progressivity of Social Security redistributions due to the structure of the benefit formula, but we find that Social Security redistributions exhibit little progressivity for individuals born as late as 1980. Differential mortality rates across race and education groups are the primary explanation. While black-white mortality gaps have narrowed some in recent years, they remain large and dull progressivity. Mortality gaps by education level are also large and unlike the race gaps are widening, which puts additional regressive pressure on Social Security redistributions.

Acknowledgments
Tan is in the Department of Economics, and Koedel is in the Department of Economics and Truman School of Public Affairs, at the University of Missouri. We thank David Kaplan and Shawn Ni for useful comments and suggestions and Lori Beth Reeder for assistance with accessing and using the Synthetic SIPP. All errors are our own.

1. Introduction

The Social Security retirement system (Old-Age and Survivors Insurance, or OASI) is the largest social program in the United States, with benefit payments equal in value to roughly five percent of the United States gross domestic product (U.S. Social Security Administration, 2015). In addition to providing secure retirement income for older Americans, Social Security is also designed to transfer resources across workers. Some aspects of the system – most notably the benefit formula – redistribute income progressively. Other aspects exert regressive pressure on redistributions, including the higher return to longer lifespans and spousal and survivor benefits (Gustman and Steinmeier, 2001; Butrica and Smith, 2012). Previous research on the net progressivity of Social Security is mixed, although most studies conclude that the system is at least modestly progressive (Garrett, 1995; Gustman and Steinmeier, 2001; Liebman, 2002; Butrica, Iams and Smith, 2003; Smith, Toder and Iams, 2003; Fullerton and Mast 2005; Coronado, Fullerton and Glass, 2011; Goda, Shoven and Slavov, 2011; Gustman, Steinmeier and Tabatabai, 2013).

We provide new, prospective estimates of the redistributive impacts of Social Security for recent cohorts born during the second half of the 20th century. Our forward-looking analysis is facilitated by the application and extension of a novel method for forecasting lifetime earnings of young workers based on Gu and Koenker (2017; also see Koenker and Mizera, 2014). We focus on understanding Social Security redistributions across racial/ethnic and education groups and on men's own benefits. The latter allows us to isolate two mechanisms for redistribution: the progressive Social Security earnings-replacement formula and differential mortality.[1]

Studies by Gustman and Steinmeier (2001), Gustman, Steinmeier and Tabatabai (2013), and Smith, Toder and Iams (2003), using data from cohorts born during the early-to-mid-20th century, find

[1] Our focus on men's own benefits facilitates a clear, detailed investigation of these two specific channels, but our estimates ignore other sources of redistribution in the system. In the conclusion we briefly contextualize our findings within the broader literature that addresses other redistributive aspects of Social Security OASI.

that redistributions based on own contributions and benefits are progressive across income groups. Smith, Toder and Iams (2003) also find that redistributions are progressive across race and education groups; moreover, they find a positive trend in progressivity from 1931-60. All else equal, rising earnings inequality during the latter half the 20th century (Autor, Katz and Kearney, 2008) should put additional progressive pressure on Social Security redistributions for younger cohorts because of the progressive earnings-replacement formula. However, our findings for cohorts born between 1962-67 and 1975-80 indicate that redistributions between individuals who differ by race/ethnicity and education level exhibit little progressivity. Moreover, progressivity is improving only modestly between these cohorts.

A key factor driving our results is differential mortality. It is well understood that mortality is an important determinant of Social Security returns but surprisingly, the use of mortality data in Social Security research is inconsistent and often poorly-documented (Goda, Shoven and Slavov, 2011).[2] Our analysis suggests that estimates of Social Security returns are quite sensitive to how mortality rates are incorporated. With respect to tracking progressivity trends, an issue with previous research is that mortality-rate gaps across socioeconomic groups have been applied statically. That is, while the overall mortality trend is allowed to evolve over time, the adjustment factors that permit differential mortality across groups have been held fixed.[3] In reality, mortality gaps are changing (Auerbach et al., 2017; Bound et al., 2014; Waldron, 2007; Goldman and Orszag, 2014). For the recent cohorts we study, mortality rates for African Americans and Hispanics are improving relative to whites and mortality rates are diverging between individuals who differ by education level. We illuminate the role played by differential mortality in a decomposition exercise of the Social Security internal rate of return (IRR). Our decompositions show that while earnings dynamics have indeed increased the progressivity of

[2] Brown, Liebman and Pollet (2002) is a notable exception.
[3] For example, see Smith, Toder and Iams (2003) and Gustman, Steinmeier and Tabatabai (2013). We elaborate on the issue of mortality calculations below.

Social Security redistributions between the cohorts we study, as expected; diverging mortality rates have offset much of the gains in progressivity that would otherwise occur across education groups.

2. Background and Motivation

The Social Security retirement system is a pay-as-you-go defined benefit system. Because contributions and benefits are not linked at the individual level, the system permits resource transfers across individuals.[4] Liebman (2002) notes that the most significant source of transfers is the progressive Social Security benefit formula. In brief, individuals' earnings over the worklife are indexed to average wage growth (inclusive of inflation), summed over the career, and then the highest 35 years are used to compute each worker's AIME. The monthly benefit that the worker receives – the Primary Insurance Amount (PIA) – is based on a kinked formula that replaces a fraction of AIME, with lower AIME values being replaced at a relatively high rate and higher AIME values being replaced at a lower rate. Thus, higher-earning individuals, and individuals who are more attached to the workforce over their careers, have a lower IRR all else equal (we show the AIME and PIA formulas below; also see Fullerton and Mast, 2005).

Although the earnings-replacement formula is clearly progressive, other aspects of the Social Security benefit structure put regressive pressure on redistributions. For example, differences in mortality rates across workers offset some of the progressivity of the system because benefits are paid until death and higher earners have greater longevity (Auerbach et al., 2017; Goda, Shoven and Slavov, 2011; Waldron 2013; Goldman and Orszag, 2014). Spousal and survivor benefits also offset some of the progressivity of Social Security because high earners are more likely to have spouses and spouses who live longer (Butrica and Smith, 2012; Gustman and Steinmeier, 2001).

[4] Resource transfers can occur both within (Liebman, 2002) and between (Boll, Raffelhüschen and Walliser, 1994) cohorts. In this study, we focus on intra-cohort transfers.

A number of studies have used data from the SSA to perform retrospective analyses of the redistributive effects of Social Security. These studies focus on retirees or near-retirees for whom full-career work histories are available. An advantage of this work is the quality of the data; the research can provide clear insight into the redistributive effects of Social Security for earlier generations of workers. For example, Gustman, Steinmeier and Tabatabai (2013) use data from cohorts born between 1936-41 and 1948-53 to show that the Social Security system redistributed wealth from higher-earning individuals to lower-earning individuals in calculations based on own benefits and contributions. They also find that progressive redistributions weaken substantially when distributions between households are considered (although there is a positive progressivity trend owing to women's increased workforce participation).

It is also important to study the redistributive effects of Social Security prospectively. Constantly evolving demographic, health, and labor-market conditions are such that estimates from previous generations may have limited applicability for young workers (e.g., see Butrica, Iams and Smith, 2003; Goldman and Orszag, 2014; Jeong, Kim and Manovskii, 2015). Prospective studies to date have relied on large-scale microsimulation models. The Cornell Microsimulation Model (CORSIM) is a high-profile example. The foundation of CORSIM is a representative sample of Americans taken from the 1960 U.S. Census (Caldwell et. al., 1999). The simulation program grows the 1960 sample demographically and economically in one-year intervals through the year 2100.[5] Other microsimulation models that have been used to study Social Security prospectively include the Urban Institute's Dynamic Simulation of Income Model (DYNASIM; Favreault and Smith, 2004), the Social Security Administration's Modeling Income in the Near Term model (MINT; Smith and Favreault; 2013), the Congressional Budget Office's Long Term Model (CBOLT; Meyerson et. al., 2009), and

[5] CORSIM also served as the basis for developing Canadian and Chinese nationwide microsimulation models. In 2001, CORSIM was purchased by the SSA and used to develop the Policy Simulation Model (POLISIM).

the Future Elderly Model (FEM; Goldman et. al., 2015) developed by the USC Roybal Center for Health Policy Simulation. These models all fit under the broad "microsimulation" umbrella but use different foundational datasets and microsimulation processes (Li and O'Donoghue, 2013).

Evidence from microsimulation-based projections of Social Security redistributions is inconclusive about whether the system is becoming more progressive over time. Studies using the (current) MINT (Biggs, Sarney and Tamborini, 2009) model and the CORSIM model (Caldwell et al., 1999) show no progressivity trends for cohorts born after the 1940s; while studies using the CBOLT model (Congressional Budget Office, 2015) and an earlier version of MINT (Smith, Toder and Iams, 2003) indicate an increasingly progressive system. The most recent microsimulation-based estimates comparable to our own are reported in Smith, Toder and Iams (2003). These authors use the MINT model to examine net Social Security benefit gaps for cohorts born between 1931 and 1960 as measured across income, race and education groups. As noted above, they find progressive redistributions between race and education groups, and estimate a trend of increasing progressivity over time.

Limitations of the literature built on microsimulation evidence include a lack of transparency and replicability. Kashin, King and Soneji (2015) raise the issue of replicability with forecasts from the Office of the Chief Actuary of the Social Security Administration. There are also significant barriers to replicating other forward-looking Social Security estimates. One reason is that the microsimulation models used to populate the literature are complex and proprietary (Cumpston, 2011; Li and O'Donoghue, 2013). It is difficult to obtain the documentation necessary for replication and sensitivity testing (e.g., see National Academy of Sciences, 2015, p. 24) and because the models are expensive to develop, it may not be in developers' pecuniary interests to disclose modeling details (Cassells, Harding and Kelly, 2006).[6] In contrast, our methods, while not as comprehensive as microsimulation alternatives, offer the advantage of being transparent and accessible. They can be useful for opening up the study of some aspects of the Social Security retirement system to a broader group of scholarly contributors.

3. Data

3.1 Validation Sample

As noted in the introduction, we rely on forecasts of lifetime earnings for young workers in the cohorts we study to estimate the expected Social Security IRR, which we use to study redistributions. Before getting to our main analysis, we first validate our forecasting procedure using an older cohort for whom key Social Security values can be calculated using actual SSA data. Our validation cohort includes men from the 1996 Survey of Income Program Participation (SIPP) who were born between 1946-51. SSA data for these men are available through 2011 via the restricted-use SIPP.

The SIPP is a household survey designed as a continuous series of national panels. Each panel features a nationally representative sample of individuals interviewed three times each year over a several-year period. The 1996 SIPP has four waves with tri-annual earnings data. In addition to earnings, the SIPP data also include information about workers' race/ethnicity designations and education levels, which we use in our forecasts. Table 1 provides basic descriptive statistics for the full sample of men in the 1996 SIPP and the validation sample (with sample weights).

We use individuals born between 1946-51 for the validation exercise because they are the youngest individuals in the 1996 SIPP with reasonably complete Social Security earnings records

[6] In addition to the example from the National Academy of Sciences referenced in the text, a second example is from the Future Elderly Model (FEM). A description of the FEM on the Health and Retirement Study website states: "Because of the complexity of the simulation [of the FEM], support for these files outside a collaboration will not be provided." (see http://hrsonline.isr.umich.edu/index.php?p=shoavail&iyear=C9, retrieved on 06.09.2016).

available in SSA data. For the youngest of these individuals, born in 1951, the SSA data allow us to construct full earnings profiles through age-60 (i.e., 2011). We obtained access to the restricted-use SSA records via the Synthetic SIPP (U.S. Census Bureau, 2015), which is a tool provided by the United States Census Bureau to facilitate broader access to the restricted-use data, including the SSA supplement (Benedetto, Stinson and Abowd, 2013).

3.2 Forecasting Samples

After confirming that our forecasted AIME distribution based on the public-use SIPP matches the distribution calculated using SSA data for the validation cohort, as documented below, we apply our forecasting procedure to a younger cohort of workers from the 1996 SIPP born between 1962-67, as well as a young cohort from the 2008 SIPP born between 1975-80, using the public-use files. As noted above, the 1996 SIPP is a four-year panel with tri-annual earnings records; the 2008 SIPP is a five-year panel and also contains tri-annual earnings data. The validation work is based on the 1996 SIPP and thus uses four years of earnings data. For analytic consistency, all of our forecasts are based on four years of data (i.e., we ignore the fifth year of earnings data from the 2008 file when we forecast lifetime earnings for the cohort born between 1975-80).

We examine Social Security IRR differences across race/ethnicity and education groups. We divide individuals into one of four racial/ethnic groups: non-Hispanic black, non-Hispanic white, Hispanic, and other. For our analysis by education level we also use four categories: less than high school, high school, some college, and bachelor's degree or higher. Table 2 provides descriptive statistics for the 1996 and 2008 SIPP files and the forecasting subsamples (with sample weights); the first column of Table 2 replicates the first column of Table 1 because the validation cohort and the 1996 forecasting cohort are from the same SIPP file.

4. Methodology

4.1 Framework for Earnings Forecasts

To estimate forward-looking Social Security values for young workers we require lifetime earnings profiles, which we aim to forecast using a model based on SIPP earnings data. Gu and Koenker (2017) develop an empirical Bayes estimation framework that can be used to construct earnings expectations over the work life for individuals using a limited subsample of earnings data (also see Koenker and Gu, forthcoming). Combining the analytic insights of Gu and Koenker (2017) with evidence from Murphy and Welch (1990) on wage modeling, we start with the following linear regression model to forecast earnings:

$$Y_{it} = \mathbf{I}_{\mathbf{i}}^{\mathbf{re}} * (\boldsymbol{\beta}_{\mathbf{0}}^{\mathbf{re}} + \boldsymbol{\beta}_{\mathbf{1}}^{\mathbf{re}} * EXP_{it} + \boldsymbol{\beta}_{\mathbf{2}}^{\mathbf{re}} * EXP_{it}^2 + \boldsymbol{\beta}_{\mathbf{3}}^{\mathbf{re}} * EXP_{it}^3 + \boldsymbol{\beta}_{\mathbf{4}}^{\mathbf{re}} * EXP_{it}^4) + \boldsymbol{\lambda}_{\mathbf{t}} + \varepsilon_{it} \quad (1)$$

In equation (1), Y_{it} is earnings relevant to Social Security for individual *i* in year *t*.[7] $\mathbf{I}_{\mathbf{i}}^{\mathbf{re}}$ is a vector of indicator variables that uniquely identifies each individual's race-by-education (*re*) group, of which there are 16 possible groupings (4x4) per the preceding section.[8] EXP_{it} is workforce experience calculated as pseudo-experience and included as a quartic following Murphy and Welch (1990).[9] $\boldsymbol{\lambda}_{\mathbf{t}}$ is a vector of indicators for each tri-annual survey, included to control for changes in inflation and wage growth over the course of the data panel. The model allows for differential returns to experience over

[7] Specifically, Y_{it} is calculated as total earnings from employment plus 92.35 percent of earnings from self-employment, and then top-coded to 110 percent of the maximum taxable value under Social Security, which censors earnings for the top 4 percent of individuals.
[8] Ideally the educational groups would be based on education observations at a fixed age (e.g., education at age-25), but because the SIPP panel is short this is not possible. Instead, we group all individuals by their highest observed education level in the data. This introduces some measurement error into the forecasting process (e.g., consider an individual who is age-28 and will obtain a higher degree in his 30s; his 40-year old counterpart in our sample would be coded with the higher education level, but he would not); however, the degree of measurement error is limited by the fact that only a small fraction of workers cross significant educational-certification thresholds after their mid-to-late 20s (per tabulations from the United States Census, see here: https://www.census.gov/hhes/school/data/cps/2014/tables.html).
[9] Workforce experience is not directly available in the SIPP (and even if it were, reporting accuracy would be a concern). We overcome this data limitation by constructing a pseudo-experience variable for individuals in each education group. For the education groups (1) less than high school, (2) high school, (3) some college, and (4) bachelor's or more, we define pseudo experience as age minus (1) 18, (2) 19, (3) 20 and (4) 22, respectively. Pseudo experience has been shown to be an effective proxy in previous modeling of earnings profiles by Murphy and Welch (1990).

the work life across race and education groups as indicated by the *"re"* superscripts on the β parameters. This is important for our application because differential group-level earnings trajectories can result in differential Social Security contribution and benefit profiles. The inclusion of $\boldsymbol{\lambda}_\mathbf{t}$ in the model ensures that the β parameters capture real wage differences between groups and by experience. ε_{it} is the error term.

Accounting for earnings heterogeneity across individuals, and within-individual earnings volatility, is a significant analytic challenge. Following Gu and Koenker (2017), we use the residuals from equation (1) to model earnings dynamics in a random-effects framework as follows:

$$\varepsilon_{it} = \alpha_i + v_{it} \tag{2}$$

$$v_{it} = \rho v_{it-1} + \sqrt{\theta_i}\xi_{it}, \quad \xi_{it} \sim N(0,1) \tag{3}$$

The parameter ρ measures the serial correlation in the residuals, v_{it}, and can be estimated by profile likelihood. Letting φ_{it} denote $\varepsilon_{it} - \rho\varepsilon_{it-1}$, equations (2) and (3) can be combined as follows:

$$\varphi_{it} = (1-\rho)\alpha_i + \sqrt{\theta_i}\xi_{it}, \quad \xi_{it} \sim N(0,1) \tag{4}$$

Equation (4) can be viewed as a Gaussian location-scale mixture model, where $(1-\rho)\alpha_i$ and θ_i are the location and scale parameters. Assuming that α_i and θ_i are drawn iid from some distribution *F*, *F* can be derived consistently through non-parametric MLE (Gu and Koenker, 2017). With *F* serving as the prior distribution, and with likelihood functions based on earnings data $\varphi_i = \{\varphi_{i0}, \varphi_{i1}, ..., \varphi_{iT}\}$, it is straightforward to calculate the posterior distribution for each individual.[10]

[10] The likelihood functions are $\frac{1}{N_i}\sum_{t=1}^{N_i}\varphi_{it} \mid \alpha_i, \theta_i \sim N((1-\rho)\alpha_i, \theta_i / N_i)$ and $\sum_{t=1}^{N_i}(\varphi_{it} - \frac{1}{N_i}\sum_{j=1}^{N_i}\varphi_{ij})^2 / \theta_i \mid \theta_i \sim \chi^2(N_i - 1)$ where N_i is the number of φ_{it} for individual *i*.

We draw one set of $\{\hat{\alpha}_i, \hat{\theta}_i\}$ from each individual's own posterior distribution, which captures the individual's mean and variance of earnings. With the values $\{\hat{\alpha}_i, \hat{\theta}_i\}$ and parameter vector β, we can recover individual earnings profiles over the entire career.

Allowing for heterogeneity in earnings and earnings volatility is important for studying Social Security because the formula that converts AIME to PIA is nonlinear and favors individuals with more income volatility (see details below). Our estimation of the F distribution shows a positive correlation between α_i and θ_i, indicating that ignoring within-individual earnings volatility would introduce bias into estimates of Social Security values.[11]

In addition to the conceptual benefit of jointly modeling unobserved earnings heterogeneity across workers and within-worker earnings volatility, our approach is a methodological improvement over previous attempts to incorporate individual earnings heterogeneity in Social Security research. For example, the MINT model uses fixed effects to forecast individual earnings from age 55 to retirement (Smith and Favreault, 2013). However, the standard incidental parameters problem emerges; i.e. the individual heterogeneity parameters are not consistently estimated. The approach outlined above relies on information from the full sample to inform the prior on the incidental parameters. This requires random effects modeling, which usually assumes an arbitrary functional form on the distribution of incidental parameters, but this assumption can be replaced with the weaker assumption that the incidental parameters are iid distributed per Kiefer and Wolfowitz (1956). We estimate the distribution of incidental parameters non-parametrically under the iid assumption.[12]

---

[11] Gu and Koenker (2017) find a negative correlation. The reason for the difference is that we use linear earnings and they use log earnings. The magnitude of earnings volatility is weaker among high earners under the log transformation.

[12] Previous attempts to perform similar estimations have been hampered by computational demands. For example, the EM algorithm has a slow rate of convergence and is inefficient when applied to large datasets (Jiang and Zhang, 2009). However, Koenker and Mizera (2014) develop an alternative algorithm based on convex optimization that converges much faster and we apply their algorithm in our work.

Equations (1), (2) and (3) require panel data on individual earnings and are not suited to evaluate individuals who are not observed in the workforce, or who are in and out of work. In fact, in their original demonstration of the method with PSID data, Gu and Koenker (2017) drop all individuals without complete earnings profiles over the interval of earnings observations used for forecasting. In our investigation of Social Security redistributions, it is critical to account for everyone in a cohort so we must expand the approach.

As a first step in the process of obtaining career earnings profiles for all individuals, we define an analytic sample on which to estimate equation (1). This sample includes all men in the SIPP data with positive earnings in all periods of our data panel (as in Gu and Koenker, 2017), plus individuals with positive earnings in at least one period. To estimate equations (2) and (3), the prior *F*, and the serial correlation coefficient ρ; we restrict the sample to individuals with positive earnings in all periods because individuals who are in and out of the workforce and have no/fewer consecutive earnings records will attenuate our estimate of ρ. Figure 1 plots the estimated joint distribution of α and θ as well as the likelihood plot of ρ. Note that α and θ are positively correlated within individuals, indicating that higher income workers also have higher earnings volatility.[13]

To calculate the individual posterior we require the individual likelihood function. For each individual with a complete earnings record, we have a complete series of $\varphi_i = \{\varphi_{i0}, \varphi_{i1}, ..., \varphi_{iT}\}$, from which it is straightforward to derive the likelihood function. However, for individuals who are in and out of the workforce the series is incomplete. For those with partial work histories, we collapse their work-history gaps and order available earnings observations chronologically to calculate φ_i. As long

[13] As in Gu and Koenker (2017), we find relatively modest persistence in earnings as captured by ρ when implementing this richer procedure for accounting for earnings dynamics.

as an individual has at least two φ_{it} available (which requires at least three earnings records), the likelihood is based on the set of available φ_{it}.

We also require wage projections for individuals who are not observed in the workforce at all or have less than three earnings records during the SIPP sampling window. It is important to account for these individuals because their Social Security values will impact cohort-level outcomes. Approximately 22 percent of individuals born between 1946-51 in the 1996 SIPP have less than three earnings records. Individuals may have less than three earnings records for two reasons. 14 percent of individuals only participate in the survey for one or two waves, thus it is impossible for us to observe more than three earnings records for them. Eight percent of individuals participate in the survey for at least three waves but still have less than three earnings records because they do not have positive earnings in at least one wave.

Because we do not have sufficient earnings information on these individuals, we draw α and θ directly from the prior. That is, we treat these individuals as "average" in the sample. Treating them as average is probably inaccurate – it is reasonable to expect that they are negatively selected, in which case they would have below-average conditional earnings if they were observed working. This is particularly likely for the 8 percent of individuals who are included in 3 or more SIPP survey waves but have fewer than 3 earnings records. However, as a matter of practice, our solution of using the prior distribution to estimate earnings for non-working individuals is effective in the sense that our cohort-level forecasts of the AIME distribution for the validation sample matches the distribution using the real SSA data.[14] We also examine the sensitivity of our findings to alternative ways of dealing with these individuals that build in negative selection, and our results are not sensitive to reasonable adjustments (see Appendix Tables A.1 and B.1). One reason for the limited sensitivity is that

[14] We also note that Breunig and Mercante (2010) find that a simple OLS-predicted wage has good out-of-sample forecasting performance for non-working individuals.

workforce attachment outcomes for these individuals are more important for determining Social Security benefits and contributions than their wages, as will become clear in the next section.

4.2 *Workforce Attachment*

Section 4.1 documents the procedure we use to build lifetime wage profiles for all workers. In this section we document how we incorporate work stoppages over the career cycle, and workforce exits, into the lifetime earnings projections. In brief summary, the basic idea is to populate a full wage profile over the entire range of potential experience– which we cap at 50 years – for each individual following the approach in Section 4.1, and then go back and "shock" individuals with unemployment spells based on a separate empirical model of workforce attachment.

The unemployment shocks for each individual depend on the sample-average unemployment rate by education and experience level, along with each individual's own observed workforce attachment. They are constructed in the following steps. First, we estimate work likelihoods by experience and education level in the full sample, smoothed by a cubic spline. We do not have enough data for the non-white subgroups to produce precisely estimated employment likelihood parameters at the race-by-education level, and thus we do not allow for differential workforce attachment by race beyond the differential created by education differences. We denote the smoothed, sample average unemployment likelihood at experience level k and education level e estimated by our regression as U_k^e. Second, we calculate the simple average unemployment likelihood for each individual, U_i, as L_i / T_i, where L_i is the number of periods that individual i does not have positive earnings in the SIPP data and T_i is the total number of periods individual i is observed.[15] Third, we define π_{ik} as the

[15] Individual workforce attachment values, U_i, for individuals who are not present in all SIPP waves are coded based on observed data. For example, an individual who is observed in only three waves but has earnings in all three waves has $U_i = 0$.

forecasted (and/or historical) employment-to-population ratio (EPR) corresponding to the year of experience k for individual *i,* divided by the average EPR observed during the SIPP window.[16] π_{ik} controls for the impact of macroeconomic conditions over the worklife on the unemployment likelihood. We then assign unemployment shocks at random throughout the career for individual i with education e and experience k with likelihood:

$$U_{ik}^{e} = (\psi U_{i} + (1-\psi)U_{k}^{e}) * \pi_{ik} \tag{5}$$

The weighting parameter ψ determines the degree to which workforce attachment observed over our short interval of data is parameterized to persist throughout the career. We do not have enough data to perform a direct estimation of ψ and thus we calibrate ψ in the validation cohort so that the forecasted share of individuals who are not eligible for Social Security benefits (i.e., the share who do not earn at least 40 credits) matches the ineligible share reported in the real SSA data. This yields a value of ψ of 0.86. Outside of individuals in the bottom decile of lifetime earnings, our Social Security forecasts are not sensitive to a fairly wide range of alternative parameterizations of ψ (see Appendix Tables A.2 and B.2).[17]

Figure 2 illustrates the workforce-attachment projections, U_{k}^{e}, for each education level over the full (potential) 50-year career. As expected, more educated individuals are more attached to the workforce throughout the career, and work longer. Early in the career, individuals with less education have higher work likelihoods, which is by construction because we do not allow individuals to work

[16] The EPR is specific to the year, *t*, which we do not index for notational brevity. We instead attach the relevant year-specific EPR to each individual by experience matched to the appropriate year. As indicated below, we obtain EPR data from the St. Louis Federal Reserve Bank FRED® data center and the 2015 Annual Trustee Report from the Social Security Administration (for historical and projected values, respectively).

[17] Our forecasts for individuals with the lowest career earnings (in the bottom decile), whose low earnings are driven in large part by low workforce participation rates, are most sensitive to the parameterization of ψ because they have the largest differences between U_{i} and U_{k}^{e}.

while in school. While this is clearly a simplification, our findings are not sensitive to reasonable modifications (see Appendix Tables A.3 and B.3).[18]

Workforce exits are empirically built into the model by the high rates of unemployment late in the career shown by the trends in U_k^e in Figure 2, but the high value of ψ in Equation (5) dulls the impact. The practical implication is that workforce attachment is overextended in our forecasts for highly-attached individuals during the 4-year SIPP window (i.e., individuals with high values of U_i). Workforce attachment profiles late in the career can be made to better match the real data by allowing ψ to vary with work experience – specifically, by shifting more weight to the group trend later in the career, which is a scenario we consider in Appendix Tables A.2 and B.2. However, this has a limited impact on our forecasts because AIME relies on only the highest 35 years of earnings, with the end result being that late-career earnings for highly-attached individuals do not meaningfully affect AIME.

4.3 Social Security Calculations

With the career earnings profiles in hand, we use the Social Security contribution rate and benefit formula to forecast lifetime benefits and contributions for individuals, inclusive of contributions on the behalf of workers by employers. The two Social Security formulas used in our study are for AIME and the PIA. The formula for AIME is:

$$AIME_i = \begin{cases} 0 & if \quad credit < 40 \\ \sum_{t=1}^{35} Index_{it} * Y_{it} / 420 & if \quad credit \geq 40 \end{cases} \tag{6}$$

Where:

[18] The work restriction is built into the forecasting model by the pseudo-experience variable, which assigns the year of workforce entry based on the education level as reported in Section 4.1 (footnote 9). In Appendix Tables A.3 and B.3, we examine the sensitivity of our findings to allowing for work during college for college graduates, with wages equal to a fraction of post-schooling wages, and our results are substantively unaffected. The insensitivity is because Social Security contributions and benefits from part-time jobs in college are only a small part of the lifetime contributions and benefits of college-educated workers.

$$Index_{it} = \begin{cases} \dfrac{AWI_{age=60}}{AWI_{age=age_{it}}} & if \quad age_{it} < 60 \\ 1 & if \quad age_{it} \geq 60 \end{cases} \tag{7}$$

In equation (6) Y_{it} is annual earnings up to the maximum taxable value for individual *i* in year *t*. If more than 35 years of earnings are available, the highest 35 (indexed) values are used. AWI in equation (7) is the average wage index, which accounts for inflation and real wage growth and is used to scale earnings for the purpose of calculating the Social Security benefit.

The content of equations (6) and (7) can be summarized as follows: AIME scales annual earnings by the AWI and is the sum of the highest 35 scaled values, converted into monthly earnings by dividing by 420. Earnings before age 60 are indexed to AWI at age 60; earnings after age 60 are not indexed further. Individuals who accrue less than 40 quarters of earnings (10 years of credit) over the entire career are not Social Security eligible and AIME is set to zero.

The formula for PIA is the key source of progressivity in Social Security. It is a kinked replacement-rate formula for AIME as follows:

$$PIA_i = \begin{cases} 0.9 * AIME_i & if \quad AIME_i \leq BP_1 \\ 0.9 * BP_1 + 0.32 * (AIME_i - BP_1) & if \quad BP_1 < AIME_i \leq BP_2 \\ 0.9 * BP_1 + 0.32 * (BP_2 - BP_1) + 0.15 * (AIME_i - BP_2) & if \quad AIME_i > BP_2 \end{cases} \tag{8}$$

In Equation (8), BP_1 and BP_2 are the two “bend points” in the formula, which is structured to replace low earnings at a higher rate than high earnings. BP_1 and BP_2 are automatically increased each year by the same percentage as the AWI. In 2016, $BP_1 =$ \$856 and $BP_2 =$ \$5,157.

We apply these formulas to the projected earnings profiles and use them to compute lifetime expected Social Security benefits. We assume that individuals start collecting Social Security benefits at the normal retirement age. To estimate lifetime contributions, we use the total contribution rate up to the salary ceiling, which is currently $118,500 with a contribution rate of 10.6 percent as of 2016.[19] Above the salary ceiling no contributions are collected.

Our forecasting model also requires historic and predicted data on the Average Wage Index (AWI), Consumer Price Index (CPI), and the employment to population ratio (EPR). The historical data are obtained from the St. Louis Federal Reserve Bank FRED® data center and future projections are from the 2015 Annual Trustee Report from the Social Security Administration (2015).

4.4 Mortality Rates

Few studies in the literature provide clear documentation of their mortality calculations. Based on the limited information that we have been able to find, one issue in studies that aim to document progressivity trends is that while overall mortality rates are allowed to evolve over time, researchers apply static mortality-rate adjustment factors for socioeconomic groups. For example, Smith, Toder and Iams (2003) use static mortality adjustment factors across groups estimated from a hazard model run on PSID data over the years 1968-94. This structurally prevents differences in mortality trends between groups from affecting estimates of redistributive trends over the timespan of their analysis (1931-60), which will affect estimates of the change in progressivity. Similarly, Gustman, Steinmeier and Tabatabai (2013) apply static mortality-adjustment factors by income taken from Duleep (1989) for cohorts born between 1936-41 and 1948-53.[20] In addition to the findings we report below, studies that reinforce the importance of accounting for mortality trends over time include Auerbach et al., (2017), Bound et al. (2014), Goldman and Orszag (2014), and Waldron (2007).

We use the cohort life tables for men taken from Bell and Miller (2005) as baseline mortality rates in our study. These life tables are uncontroversial and widely used (Walden 2007; İmrohoroğlu

[19] The 10.6 percent total tax is for the retirement-benefit portion of Social Security.
[20] Gustman, Steinmeier and Tabatabai (2013) provide little information about their mortality calculations, noting only that survival rates are calculated using "a life table adjusted for variation in life expectancy with income" (p. 10). We were able to gain additional insight into their procedures and confirm that a static adjustment was used via correspondence with the authors. We thank the authors for their assistance.

and Kitao, 2009). We then apply adjustment factors to these rates so that they capture differential mortality across race and education groups, by cohort. We produce the adjustment factors following the approach of Bound et al. (2014); they are based on mortality data from the Centers for Disease Control (CDC) combined with data from the United States Census. The CDC mortality data are based on deaths in a given year rather than being tied to a birth cohort. Thus, for any living cohort, the newest CDC data provide the most accurate information about differential mortality. In this sense, the case could be made to use the most recent CDC data for both cohorts we study; indeed, this line of thinking may have contributed to previous decisions by researchers to use static differential mortality gaps. However, doing so would mask trends in mortality across groups over time. In order to ensure that mortality trends are captured by our calculations, we use the newest five years of CDC and American Community Survey (ACS) data (2010-14) to estimate mortality adjustment factors by race and education group for the 1975-80 cohort, and data from 13 years prior for the 1962-67 cohort (i.e., CDC data from 1997-2001 and year-2000 decennial Census data).[21] To illustrate the importance of allowing for differential mortality trends, we also estimate trends in Social Security redistributions using static mortality adjustment factors below, as has been done in previous work.

Appendix C shows the final mortality tables that we use in our calculations for transparency and to facilitate replications. These data are also available in a .csv file.

4.5 Methodological Summary and Assumptions

In summary, our forecasting approach takes observed earnings outcomes over a short time interval and uses Bayesian methods to forecast career earnings. We rely on observed earnings outcomes to implicitly account for many of the factors that are explicitly built into available large scale

[21] Ideally we would use ACS data from 1997-2001, but ACS data are not available prior to the year 2000. By using data from the 2000 Census (5 percent sample), the assumption for procedural consistency is that the race and education compositions in these data are similar to the average race and education compositions over the years 1997-2001.

microsimulation models, which offers benefits in terms of transparency and tractability. The key assumptions that underlie our forecasting procedure are documented in Table 3.

5. Validation

In this section we report on the results from our validation exercise. Recall that our objective is to match the distribution of AIME calculated from SSA records with our forecasts for individuals born between 1946-51 in the 1996 SIPP. A distributional match of AIME implies a distributional match of PIA, which is a strictly monotonically increasing function of AIME. As noted above, we rely on the Synthetic SIPP from the U.S. Census Bureau (2015) to gain access to the SSA supplement in the restricted SIPP file – SSA data are available for the validation cohort via the Synthetic SIPP through 2011. The Synthetic SIPP was designed for the specific purpose of making SSA and other restricted-use data in the SIPP broadly accessible, albeit with some limitations that we describe in this section.

Before proceeding further, we first clarify that the validation exercise is more precisely described as a comparison of actual and forecasted distributions of what one could call "restricted" AIME. The reason is that we observe SSA earnings records for the youngest individuals in the validation cohort (born in 1951) only through age-60, and for comparability we cap our AIME calculations at age-60 for everyone. Thus, the AIME comparisons in the validation exercise do not compare true AIME values over all years of work, but rather "restricted" AIME values calculated based on earnings through age-60. Nonetheless, our AIME calculations account for the vast majority of career earnings for individuals and are sufficient to examine the accuracy of our earnings forecasts over most of the worklife.

A first-best scenario for validation would involve linking the SSA data to our forecasts on an individual-by-individual basis. However, the Synthetic SIPP does not have all of the variables from the public-use file that we require for forecasting, and we cannot directly link the public-use and restricted-use files via the Synthetic SIPP. This means that we cannot link our earnings forecasts to SSA records at the individual level. Therefore, we pursue a second-best approach that proceeds in two steps. First, we use the Synthetic SIPP to recover the distribution of AIME from the SSA data for the cohort of interest. Second, we compare this distribution to the forecasted distribution for the same cohort based on earnings records from the public-use file.

If all individuals in the restricted SIPP file had the SSA-data supplement, the distributional comparison would be straightforward. However, 20 percent of individuals do not have linked SSA records and they are a non-random sample.[22] In the public-use data we have information for all individuals but cannot identify which individuals have SSA data and which individuals do not (because the public-use SIPP cannot be directly merged to the Synthetic SIPP). The implication is that a simple comparison between the distributions from the SSA sample in the restricted-use SIPP and the full public-use sample will be confounded by differential selection into the SSA sample.

To resolve this problem we reweight the SSA records within the Synthetic SIPP to construct a plausibly representative sample that matches the representative sample in the public-use SIPP. We obtain the weights from a simple linear regression that predicts having an SSA record as a function of observable characteristics.[23] The predictions from the regression, $\hat{P}_i$, are used to group observationally similar individuals with and without SSA records in the restricted data. Specifically, for each individual without an SSA record, we identify all individuals with an SSA record who have values of $\hat{P}_i$ within 0.01 points (i.e., a 0.01 caliper). We then evenly distribute the weight of the non-SSA individual across

[22] The direction of selection is unclear. Individuals who are more likely to have SSA-linked records are more likely to be white, married, less educated, own a home, and have lower self-reported income but higher net worth. Studies that have examined selection into having linked SSA data in the Health and Retirement Study have found similar evidence of non-random selection (of indeterminate direction) – e.g., see Gustman and Steinmeier (1999, 2001) and Olson (1999).

[23] Gustman and Steinmeier (2001) and Gustman, Steinmeier and Tabatabai (2013) employ a similar method to handle individuals without linked SSA records. The characteristics we use for matching are race, gender, education level, birth year, occupational category, marital status, home ownership status, total asset value, and total reported income across all survey waves. For more information about matching-based imputation methods see Bhuller, Mogstad and Salvanes (2011) and Watson and Starick (2011).

all matched SSA individuals. For example, if a non-SSA individual has five SSA individuals within his 0.01 caliper, we add 0.20 weight to each matched record that contains SSA data. We use the weighted sample of individuals with SSA records in the Synthetic SIPP to produce a plausibly representative distribution of AIME based on real Social Security data for the individuals in the validation cohort. Note that the 0.01 caliper is sufficiently large to match all non-SSA individuals with at least one SSA individual, which preserves the full weight of the data, and our findings are not sensitive to reasonable caliper adjustments.

Next, from the 1996 public-use SIPP file we pull data for men from the same cohort – born between 1946-51 – and produce forecasted lifetime earnings profiles based on the procedure described above, stopping at age-60 for comparability purposes. We apply the AIME formula to these profiles and see how closely the distribution of our forecasted values matches the distribution estimated based on real SSA data.

Table 4 shows results from the validation exercise. We report forecasted and SSA-data-based estimates of the mean of AIME, the share of individuals without coverage (i.e., who work fewer than 40 quarters), and AIME values at each decile of the distribution.[24] For each estimate we report the standard deviation obtained by bootstrapping with replacement, using 250 repetitions. Each bootstrap iteration using the SSA data replicates the full procedure including the matching process; similarly, each bootstrap iteration for the forecasts replicates the entire forecasting procedure.

The results in Table 4 demonstrate that the distributions based on SSA data and our forecasts are a close match. There are no statistically significant differences at any decile, as indicated by the p-values in the final column. While like with any true projection there is no assurance that forecasts for other cohorts will be accurate, the ability of our method to recover the distribution of AIME for the

[24] Recall that the share of uncovered individuals is all but assured to match between the forecasted and SSA data because we use this share to calibrate the value of ψ in equation (5).

validation cohort lends credence to the approach, and there is no reason *a priori* to expect systematic bias in extensions.

6. Estimation of Social Security Redistributions for Younger Cohorts of Workers

6.1 Primary Results

In this section we use our forecasting procedure to study the younger cohorts from the 1996 and 2008 SIPP. From the 1996 SIPP we examine individuals aged 29-34 in 1996, born between 1962-67; from the 2008 SIPP we examine individuals aged 29-34 in 2009, born between 1975-80. Recall that Table 2 provides descriptive statistics for the full SIPP waves and the forecasting samples.

We compare race and education groups by the expected Social Security IRR. The expected IRR is defined by the following formula:

$$0 = \sum_{a=18}^{100} \frac{CF_a}{(1+IRR)^a} \qquad (9)$$

In equation (9), CF_a is the total expected cash flow in real dollars at age *a*. At younger ages the cash flow is negative because individuals are in the "contributing phase" of the Social Security system. When benefits are collected later in life, the cash flow is positive. The IRR is the discount rate at which the present value of expected contributions equals the present value of expected benefits in real dollars. We calculate the IRR for each race and education group after aggregating total contributions and benefits, as in previous studies. This facilitates the inclusion of individuals who do not attain benefit eligibility and/or do not survive to the collection age (i.e., whose individual IRR values are undefined, but who may have made contributions over the course of the worklife and thus influence the group-level IRR).

Table 5 shows the Social Security IRR based on our forecasts, overall and for each race and education subgroup, by cohort. We omit other-race individuals from the IRR calculations because the calculations require specifying mortality rates and it is not obvious how to do this with the other-race group.[25] We use mortality tables starting at age-35 for all individuals, which is one year after the age range we use to select individuals into the forecasting sample (29-34). This assumes that all individuals survive until age-35. In Appendix C, we show that the substance of our findings is not sensitive to using mortality tables retroactive to age-21. This is because survival rates are very high for all individuals between ages 21-35.[26] However, if we push the use of mortality tables back to the collection age of 65 – i.e., if we assume everyone survives until age-65 – we obtain substantially different estimates of the IRR and of redistributions, which we also show in Appendix C.

The first row of estimates in Table 5 shows that the overall Social Security IRR increased from 1.10 to 1.40 percent between the cohorts we study. This is primarily because life expectancies increased and there are not offsetting changes to the benefit formula or contribution rate. In terms of redistributions the picture is mixed. Starting with the estimates by race, Hispanic individuals have much higher returns than the other groups for both cohorts. This is because (a) Hispanics have weaker labor market outcomes compared to whites (similarly to African Americans) and (b) Hispanics have much lower mortality rates than the other races (see Appendix C; also Palloni and Arias, 2004). These two factors are apparent below when we decompose the gaps. Comparing African American and white workers, redistributions are regressive for the 1962-67 cohort and essentially even for the 1975-80 cohort.

Next we turn to the estimates by education level. Individuals from the 1962-67 cohort without a high school diploma have the largest IRR, but among the other education groups there is not a clear IRR trend. For the 1975-80 cohort there is a modest improvement in progressivity.

[25] The other-race group is small (see Table 2) and diverse, which makes it difficult to construct defensible mortality rates.
[26] Note that because the individuals in our data must have lived to at least age-29 to be included in our sample, using mortality tables that go back to an earlier age involves imposing a new realization of survival rates based on population averages to these individuals, as in Liebman (2002).

Before turning to the decompositions, in Table 6 we briefly examine the potential for changes in how we use mortality rates to influence our findings. Specifically, we replicate the results for the 1975-80 cohort following all of our procedures as described above, except that we use the same mortality adjustment factors for racial/ethnic and education groups that we use for the 1962-67 cohort. That is, while we allow the overall mortality rate to evolve over time between cohorts, we hold relative mortality rates across groups fixed using a static adjustment. This mimics the procedure of other studies.[27]

For the estimates by race/ethnicity, the different mortality adjustments have little bearing on the findings substantively. However, for the estimates by education level, Table 6 shows that holding the mortality adjustment factors fixed leads to a substantial overstatement of the Social Security progressivity trend. For example, the IRR gap between the least- and most-educated groups in the 1975-80 cohort using our preferred, time-adjusted mortality tables is 0.57 percentage points (1.87 minus 1.30); but without allowing the mortality adjustment factors to change over time, this estimate increases by almost 50 percent to 0.84 percentage points (1.98 minus 1.14). This exercise makes clear that failing to account for trends in differential mortality across race and education groups can lead to inaccurate estimates of progressivity trends.

6.2 Decompositions

In Table 7 we decompose the race- and education-based IRR gaps for each cohort to determine the shares of the gaps attributable to differential mortality and labor-market dynamics. Our focus on men's own benefits, ignoring spousal and survivor benefits, ensures that the differences we observe are attributable to just these two factors. The top panel of Table 7 shows decomposition results by race, comparing non-Hispanic blacks and Hispanics to non-Hispanic whites. The bottom

---

[27] The estimated IRR overall in Table 6 is unchanged using either set of mortality rates (row 1). A near-match is expected because the overall mortality rate evolves the same over time between cohorts (it is just the adjustment factors that differ), but the exact match (to the 100th decimal place) is coincidental.

panel shows decompositions by education level, comparing less-educated individuals to those with at least a 4-year college degree. Because the gaps are relative to the more-advantaged group in each comparison, positive numbers indicate a progressive influence and negative numbers indicate a regressive influence.

To obtain the shares of the IRR gaps in Table 5 explained by the two factors, we sequentially substitute mortality rates and earnings profiles between groups. These factors are not separable and thus their ordering in the decomposition is consequential. We are not aware of a straightforward order-invariant decomposition method in our application (unlike in Gelbach, 2016); therefore, we perform the decomposition using both orderings and report the average factor shares in Table 7.

We illustrate our approach using the black-white IRR gap for individuals born between 1975-80, which is 0.07 percentage points as shown in the upper-right panel of Table 7 (this total IRR gap is taken directly from Table 5). To obtain the portion of the gap explained by mortality rates alone, we substitute white mortality rates for African Americans and otherwise replicate our calculations. Next, for the earnings and workforce attachment profiles, we construct a counterfactual AIME distribution for African Americans that matches the distribution for whites. The distributional shift is a two-step process following Bound, Lovenheim and Turner (2010) and Arcidiacono and Koedel (2014), in which we map African Americans to counterfactual AIME values to produce a distribution that matches the distribution for whites. Then we perform the same substitutions, but in reverse order, and Table 7 reports average values across both orderings.

Consistent with previous research, Table 7 indicates that the difference in mortality rates between African Americans and whites puts regressive pressure on Social Security redistributions. Specifically, the higher mortality rates of African Americans from the 1975-80 cohort lower their IRR by 0.27 percentage points relative to what their IRR would be if they had white mortality rates. Earnings dynamics work in the opposite direction given the progressive PIA formula; African

Americans' lower lifetime earnings result in an IRR that is 0.34 percentage points higher. On net, these two influences roughly cancel out, resulting in what are essentially even Social Security returns for African Americans and whites born between 1975-80. As noted above, Hispanics have a higher group-level IRR than whites because of both earnings dynamics and differential mortality – the estimates in Table 7 document the relative importance of these two factors.

In terms of trends over time by race, mortality rates for Hispanics and non-Hispanic blacks improved relative to whites over the time period we study, which is the primary driver of their relative IRR improvements. In fact, while both groups have lower earnings than whites in both cohorts (as indicated by the positive entries for the earnings gap in Table 7), for Hispanics there is a modest decline in the role of earnings dynamics in fueling the IRR difference. This suggests a small gain for this group relative to whites in terms of AIME.

Next we turn to redistributions across education groups. Like with the analysis by race, mortality gaps play an important role in curbing progressive redistributions. However, unlike in the comparisons by race, trends in mortality rates across education groups are diverging between the cohorts we study in a way that applies increasing regressive pressure. This can be seen by the consistently more-negative entries in the mortality column for the 1975-80 cohort versus 1962-67 cohort in Table 7; for all three less-educated groups, the regressive influence of differential mortality became stronger. The stronger progressive influence of the Social Security replacement formula, within the context of rising earnings inequality during this time period (Autor, Katz and Kearny, 2008), is also clearly reflected in the estimates in Table 7. For the most part the progressive effect of the replacement formula is larger than the regressive effect of changing mortality, but the net gain in progressivity is modest and inconsistent.

7. Conclusion

We estimate redistributions via the United States Social Security retirement system based on men's own contributions and benefits for cohorts born during the second half of the 20th century. Our most recent estimates are for individuals born between 1975-80. We find that own-earnings returns in the Social Security system for the cohorts we study are only modestly progressive. Moreover, the trend of increasing progressivity is weak, especially for redistributions across education groups. This is somewhat surprising in light of documentation in previous research of progressive Social Security redistributions and a trend toward greater progressivity earlier in the 20th century (Congressional Budget Office, 2015; Gustman and Steinmeier, 2001; Gustman, Steinmeier and Tabatabai, 2013; Smith, Toder and Iams, 2003), combined with rising earnings inequality in the United States that affects the cohorts we study (Autor, Katz, and Kearney, 2008).[28]

We use decompositions to show that mortality gaps across race and education groups significantly dull the progressive influence of the Social Security earnings-replacement formula. In addition to differences in mortality in levels, differences in mortality trends are also important. Mortality rates improved for African Americans and Hispanics relative to whites over the time period we study, putting some progressive pressure on redistributions across racial/ethnic groups, but diverging mortality rates between individuals with different levels of education have led to less progressive redistributions over time.

We focus our analysis on men's own-earnings returns. We do not study women, nor do we account for spousal and survivor benefits. The interaction of households with the Social Security system has been the focus of several recent studies of Social Security redistributions (e.g., Gustman and Steinmeier, 2001; Gustman, Steinmeier and Tabatabai, 2013; Smith, Toder and Iams, 2003). A

[28] While these general themes rise out of the Social Security literature, we again note that there is disagreement in some microsimulation studies about the progressivity trend (e.g., see Biggs, Sarney and Tamborini, 2009; Caldwell et al., 1999).

general takeaway is that the increased integration of women into the workforce over the 20th century has increased the progressivity of household-level redistributions (by, for example, reducing transfers within households from high-earning husbands to low-earning wives). That said, extensions of our work to capture household redistributions would be expected to show lower progressivity because household-level redistributions remain less progressive than individual-level redistributions (Gustman, Steinmeier and Tabatabai, 2013). Not as much attention has been paid of late to how progressivity is evolving with respect to own-benefit, formula-driven redistributions and it is this question we aim to inform.

Our forecasting method is new to the Social Security literature, which has historically relied on complex and proprietary microsimulation models for forecasting. While microsimulation models offer a number of advantages, they have limitations as well – perhaps most notably, their exclusive use in prospective Social Security research creates barriers to entry. Our method is straightforward to replicate and can be extended based on the information provided in this article (also see Gu and Koenker, 2017; Koenker and Mizera, 2017). Our hope is that more transparent and accessible methods can improve the capacity for scientific inquiry in this area of research.

<u>References</u>

Arcidiacono, P. and Koedel, C. (2014). Race and College Success: Evidence from Missouri. *American Economic Journal: Applied Economics* 6(3), 20-57.

Auerbach, A.J., Charles, K.K., Coile, C.C., Gale, W., Goldman, D., Lee, R., Lucus, C.M., Orszag, P.R., Sheiner, L.M., Tysinger, B., Weil, D.N., Wolfers, J., & Wong, R. (2017). How the Growing Gap in Life Expectancy May Affect Retirement Benefits and Reforms. *The Geneva Papers on Risk and Insurance—Issues and Practice* 42(3), 475-499.

Autor, D.H., Katz, L.F., & Kearney, M.S. (2008). Trends in U.S. Wage Inequality: Revising the Revisionists. *Review of Economics and Statistics* 90(2), 300-323.

Bell, F. C., & Miller, M. L. (2005). Life Tables for the United States Social Security Area, 1900-2100. Actuarial Study no. 120.

Benedetto, Gary, Martha H. Stinson and John M. Abowd. (2013). The Creation and Use of the SIPP Synthetic Beta. Washington, DC: United States Census Bureau.

Biggs, A. G., Sarney, M., & Tamborini, C. R. (2009). A Progressivity Index for Social Security. Issue Paper 2009-01.

Boll, S., Raffelhüschen, B., & Walliser, J. (1994). Social Security and Intergenerational Redistribution: A Generational Accounting Perspective. *Public Choice* 81(1-2), 79-100.

Bound, J., Geronimus, A., Rodriguez, J., & Waidman, T. (2014). The Implications of Differential Trends in Mortality for Social Security Policy. University of Michigan Retirement Research Center (MRRC) Working Paper.

Bound, J., Lovenheim, M. F., & Turner, S. (2010). Why have college completion rates declined? An analysis of changing student preparation and collegiate resources. *American Economic Journal: Applied Economics*, 2(3), 129-157.

Breunig, R., & Mercante, J. (2010). The Accuracy of Predicted Wages of the Non‐Employed and Implications for Policy Simulations from Structural Labour Supply Models. *Economic Record*, 86(272), 49-70.

Brown, J., Liebman, J. B., & Pollet, J. (2002). Appendix. Estimating Life Tables That Reflect Socioeconomic Differences In Mortality. In M. Feldstein and J.B. Liebman (Eds.), *Distributional Aspects of Social Security and Social Security Reform* (pp. 447-458). Chicago: University of Chicago Press.

Butrica, B. A., Iams, H. M., & Smith, K. E. (2003). The Changing Impact of Social Security on Retirement Income in the United States. *Social Security Bulletin*, *65*(3).

Butrica, B. A., & Smith, K. E. (2012). The Impact of Changes in Couples' Earnings on Married Women's Social Security Benefits. *Social Security Bulletin*, *72*(1), 1-9.

Caldwell, S., Favreault, M., Gantman, A., Gokhale, J., Johnson, T., & Kotlikoff, L. J. (1999). Social Security's Treatment of Postwar Americans. In Poterba J.M. (Eds.) *Tax Policy and the Economy*, Volume 13 (pp. 109-148). Boston, MA: MIT Press.

Cassells, R., Harding, A., & Kelly, S. (2006). Problems and Prospects for Dynamic Microsimulation: A Review and Lessons for APPSIM. NATSEM, University of Canberra.

Congressional Budget Office. (2015). CBO's 2015 Long-Term Projections for Social Security: Additional Information. Washington, DC: Congressional Budget Office.

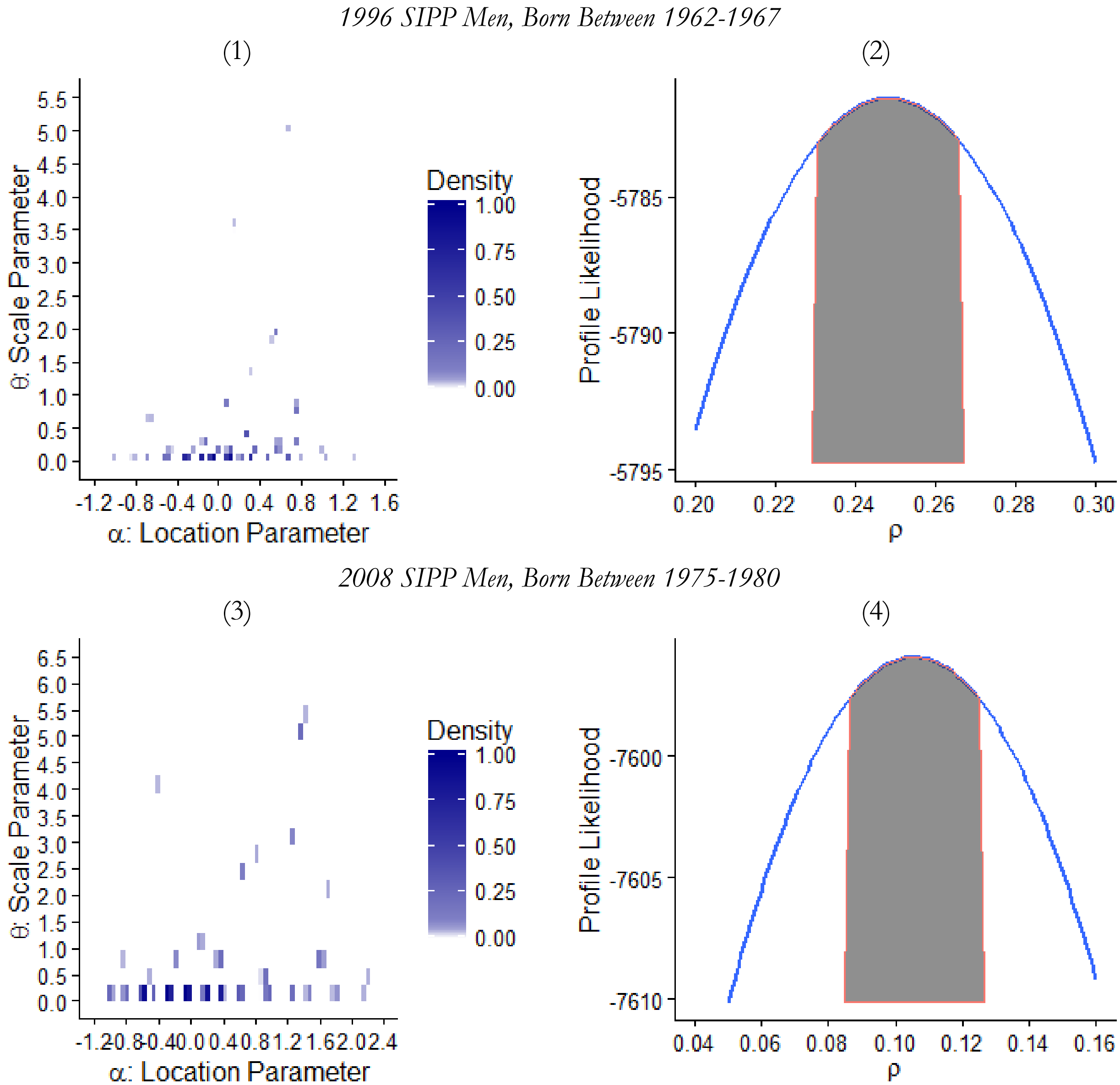
Coronado, J.L., Fullerton, D., & Glass, T. (2011). The Progressivity of Social Security. *The B.E. Journal of Economic Analysis & Policy* 11(1), 1-45.

Cumpston, J. R. (2011). New Techniques for Household Microsimulation, and Their Application to Australia. Doctoral thesis, Australian National University: Canberra, Australia (retrieved from https://digitalcollections.anu.edu.au/handle/1885/9046?mode=full).

Duleep, H. O. (1989). Measuring socioeconomic mortality differentials over time. *Demography*, 26(2), 345-351.
Favreault, M. & Smith, K. (2004). A Primer on the Dynamic Simulation of Income Model (DYNASIM3). Washington, DC: The Urban Institute.
Fullerton, D., & Mast, B. (2005). *Income Redistribution from Social Security*. Washington, DC: American Enterprise Institute Press.
Garrett, D.M. (1995). The Effects of Differential Mortality Rates on the Progressivity of Social Security. *Economic Inquiry* 33(3), 457-475.
Gelbach, J. B. (2016). When do covariates matter? And which ones, and how much?. *Journal of labor Economics* 34(2), 509-543.
Goda, G. S., Shoven, J. B., & Slavov, S. N. (2011). Differential Mortality by Income and Social Security Progressivity. In David A. Wise (Eds.), *Explorations in the Economics of Aging* (pp. 189-204). Chicago, IL: University of Chicago Press.
Goldman, D. P., Lakdawalla, D., Michaud, P. C., Eibner, C., Zheng, Y., Gailey, A., Vaynman, I., Sullivan, Y., Tysinger, B. & Leaf, D. E. (2015). The Future Elderly Model: Technical Documentation. University of South California.
Goldman, D. P., & Orszag, P. R. (2014). The Growing Gap in Life Expectancy: Using the Future Elderly Model to Estimate Implications for Social Security and Medicare. *The American economic review*, 104(5), 230-233.
Gu, J., & Koenker, R. (2017). Unobserved Heterogeneity in Income Dynamics: An Empirical Bayes Perspective. *Journal of Business & Economic Statistics* 35(1), 1-16.
Gustman, A. L., & Steinmeier, T. L. (1999). What People Don't Know About Their Pensions and Social Security: An Analysis Using Linked Data from the Health and Retirement Study. NBER Working Paper No. 7368.
Gustman, A.L., & Steinmeier, T.L. (2001). How Effective is Redistribution Under the Social Security Benefit Formula? *Journal of Public Economics* 82, 1-28.
Gustman, A. L., Steinmeier, T. L., & Tabatabai, N. (2013). Redistribution Under the Social Security Benefit Formula at the Individual and Household Levels, 1992 and 2004. *Journal of Pension Economics and Finance*, 12(01), 1-27.
İmrohoroğlu, S., & Kitao, S. (2009). Labor supply elasticity and social security reform. *Journal of Public Economics*, 93(7), 867-878.
Jeong, H., Kim, Y., & Manovskii, I. (2015). The Price of Experience. *The American Economic Review 105*(2), 784-815.
Jiang, W., & Zhang, C. H. (2009). General Maximum Likelihood Empirical Bayes Estimation of Normal Means. *The Annals of Statistics* 37(4), 1647-1684.
Kashin, K., King, G., & Soneji, S. (2015). Systematic Bias and Nontransparency in US Social Security Administration Forecasts. *The Journal of Economic Perspectives* 29(2), 239-257.
Kiefer, J., & Wolfowitz, J. (1956). Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters. *The Annals of Mathematical Statistics*, 887-906.
Koenker R. and Gu, J. (forthcoming). REBayes: An R Package for Empirical Bayes Mixture Methods. *Journal of Statistical Software*, forthcoming.
Koenker, R., & Mizera, I. (2014). Convex Optimization, Shape Constraints, Compound Decisions, and Empirical Bayes Rules. *Journal of the American Statistical Association* 109(506), 674-685.
Li, J., & O'Donoghue, C. (2013). A survey of dynamic microsimulation models: uses, model structure and methodology. *International Journal of Microsimulation* 6(2), 3-55.

Liebman, J.B. (2002). Redistribution in the current U.S. Social Security system. In M. Feldstein and J.B. Liebman (Eds.), *The Distributional Aspects of Social Security and Social Security Reform* (pp. 11-41). Chicago: University of Chicago Press.
Meyerson, N., Pineles-Mark, C., Schwabish, J. A., Simpson, M., & Topoleski, J. (2009). CBO's Long-Term Model: An Overview. Washington, DC: Congressional Budget Office.
Murphy, K. M., & Welch, F. (1990). Empirical Age-Earnings Profiles. *Journal of Labor Economics* 8(2), 202-229.
National Academies of Sciences, Engineering, and Medicine. (2015). *The Growing Gap in Life Expectancy by Income: Implications for Federal Programs and Policy Responses*. Washington, DC: The National Academies Press.
Olson, Janice A. (1999). Linkages with Data from Social Security Administrative Records in the Health and Retirement Study. *Social Security Bulletin* 62, 73.
Palloni, A., & Arias, E. (2004). Paradox Lost: Explaining the Hispanic Adult Mortality Advantage. *Demography 41*(3), 385-415.
Smith, K. & Favreault, M. (2013). A Primer on Modeling Income in the Near Term, Version 7 (MINT7). Washington, DC: The Urban Institute.
Smith, K., Toder, E., & Iams, H. (2003). Lifetime Distributional Effects of Social Security Retirement benefits. *Social Security Bulletin* 65(1), 33-61.
Smock, P.J., & Manning, W.D. (1997). Cohabitating partners' economic circumstances and marriage. *Demography* 34(3), 331-341.
U.S. Census Bureau. (2015). *SIPP Synthetic Beta: Version 6.0.1* [Computer file]. Washington DC: U.S. Census Bureau; Ithaca NY: Cornell University, Synthetic Data Server [distributor].
U.S. Social Security Administration. (2015). The 2015 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds. Washington, DC: U.S. Social Security Administration.
Waldron, H. (2007). Trends in Mortality Differentials and Life Expectancy for Male Social Security-Covered Workers, by Socioeconomic Status. *Social Security Bulletin* 67(3), 1-28.
Waldron, H. (2013). Mortality Differentials by Lifetime Earnings Decile: Implications for Evaluations of Proposed Social Security Law Changes. *Social Security Bulletin* 73(1), 1-37.
Watson, N., & Starick, R. (2011). Evaluation of Alternative Income Imputation Methods for a Longitudinal Survey. *Journal of Official Statistics* 27(4), 693.

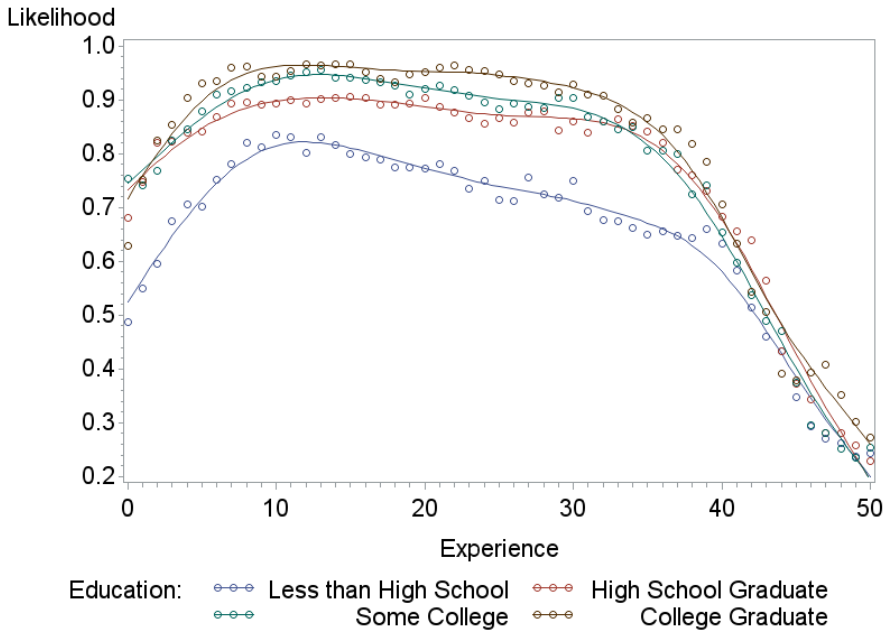
Figure 1. Estimated Prior Distributions and Distributions of ρ



Notes: Panels (1) and (3) show heat maps of the prior distributions for the 1996 and 2008 samples, respectively. The highest-density cell is rescaled to a density value of 1.0. For panels (1) and (3) numbers on the x-axis are in \$10,000 units and numbers on the y-axis are in $(\$10{,}000)^2$ units. Panels (2) and (4) are the likelihood plots of ρ. The 95% confidence interval is the shaded area.

Figure 2. Estimated Employment Likelihood over the Career, by Education Level, for the 1996 and 2008 SIPP Samples.

1996 SIPP Men



2008 SIPP Men

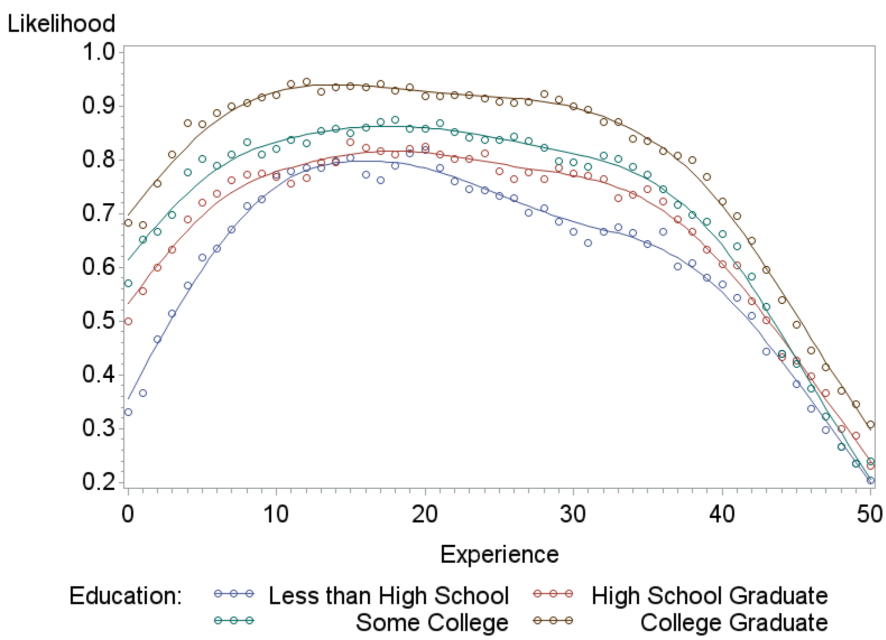


Table 1. Descriptive Statistics for the 1996 SIPP and Validation Subsample.

	All 1996 SIPP Men	Validation Cohort (Born 1946-1951)
Non-Hispanic Black	0.11	0.10
Hispanic	0.11	0.09
Non-Hispanic White	0.74	0.77
Other Race	0.04	0.04
Less than High School	0.15	0.12
High School	0.34	0.27
Some College	0.27	0.29
4-Year College Graduate	0.24	0.32
Number of Individuals	33,907	4,228
Average Nonzero Earnings	9820	13131
Positive Earnings Ratio	0.81	0.87
Number of Records	280,839	37,205

Notes: Column 1 includes all men in 1996 SIPP with 0 to 50 years of experience in the year 1996. Each person can have up to 12 records. The positive earnings ratio shows the ratio of records with positive earnings to all records. Each record is based on a 4-month interval.

Table 2. Descriptive Statistics for the 1996 and 2008 SIPP Files and Forecasting Subsamples.

	1996 SIPP		2008 SIPP	
	All Men	Forecasting Cohort (Born 1962-1967)	All Men	Forecasting Cohort (Born 1975-1980)
Non-Hispanic Black	0.11	0.13	0.11	0.12
Hispanic	0.11	0.15	0.17	0.24
Non-Hispanic White	0.74	0.67	0.65	0.56
Other Race	0.04	0.05	0.07	0.08
Less than High School	0.15	0.14	0.11	0.12
High School	0.34	0.35	0.29	0.28
Some College	0.27	0.29	0.34	0.33
4-Year College Graduate	0.24	0.23	0.27	0.27
Number of Individuals	33,907	5,040	39,064	4,750
Average Nonzero Earnings	9820	10716	17020	15660
Positive Earnings Ratio	0.81	0.92	0.74	0.85
Number of Records	280,839	41,171	319,035	36,736

Notes: Column 1 includes all men in the 1996 SIPP with 0 to 50 years of experience in 1996 (this column replicates the information in column 1 of Table 1); column 3 includes all men in the 2008 SIPP with 0 to 50 years of experience in 2009. Each person can have up to 12 records. The positive earnings ratio shows the ratio of records with positive earnings to all records. Each record is based on a 4-month interval.

Table 3. Documentation of the Key Assumptions Required for the Forecasting Procedure.

1. Functional form assumptions as indicated in equations (1) – (4). These include that the real returns to education and experience over the career are captured by the quartic experience terms by race and education level in equation (1), and that the residuals from the regression can be decomposed into a permanent, individual-specific parameter and a correlated temporary shock that follows an AR(1) process. We further assume that the profile of workforce attachment over the career can be estimated using a cubic spline, although this assumption is less important in practice because of the high weight applied to individuals' observed workforce attachment in equation (5).

2. The real returns to experience by race and education group, as estimated across the spectrum of the contemporaneous workforce, can be applied to a subsample of workers. In our primary forecasts for younger cohorts, the important assumption is that the returns estimated using data from older workers apply to younger workers. While this assumption is almost certainly inaccurate to some degree as the labor market is constantly evolving, older contemporary workers provide the closest data by which to estimate differential earnings returns for younger workers.

3. Similar to large scale microsimulation models, we rely on the SSA for long-term projections of the AWI, the CPI, and the EPR. These values are based on the SSA's projections using intermediate assumptions, which reflect the Trustees' best estimate for future experience.

4. The forecasted mortality rates are accurate, and perhaps more importantly given our research question, properly capture differential mortality trends across groups over time. Again, the mortality tables we use in our calculations are reported in Appendix C and available as a .csv file.

Table 4. Distributional AIME Comparison (Through Age-60) for the Validation Cohort.

		Forecasted Distribution	SSA-Data Distribution	P-Value from Equality Test
Mean AIME		3709.92	3726.89	0.78
		(45.32)	(39.16)	
Uncovered Share		0.04	0.04	1.00
		(0.00)	(0.00)	
AIME Deciles	1	705.98	722.3	0.81
		(47.26)	(46.37)	
	2	1555.34	1511.64	0.64
		(64.96)	(67.31)	
	3	2171.82	2284.16	0.19
		(65.46)	(56.12)	
	4	2880.61	2942.21	0.43
		(64.10)	(45.69)	
	5	3590.21	3618.55	0.77
		(68.87)	(66.47)	
	6	4273.65	4291.35	0.83
		(66.87)	(51.27)	
	7	5155.52	5099.57	0.57
		(83.37)	(50.40)	
	8	6072.04	5942.88	0.14
		(65.72)	(57.53)	
	9	6834.77	6914.32	0.30
		(47.74)	(59.54)	

Notes: The uncovered share is the share of individuals who are not eligible for Social Security benefits (less than 40 quarters with sufficient earnings). As noted in the text, we calculate AIME for the validation cohort through age-60. Standard deviations obtained via bootstrapping are reported in parentheses.

Table 5. Forecasted Internal Rates of Return, Overall and by Subgroup, for Each Forecasting Cohort.

	Cohorts	
	1962-1967	1975-1980
Overall	1.10%	1.40%
Subgroups by Race		
Non-Hispanic Black	0.83	1.29
Hispanic	1.72	1.96
Non-Hispanic White	1.03	1.22
Subgroups by Education Level		
Less than High School	1.45	1.87
High School Graduate	1.01	1.43
Some College	1.20	1.38
4-Year College Graduate	1.01	1.30

Notes: Each cell reports an estimated internal rate of return (IRR). The IRR formula is as shown in the text.

Table 6. Forecasted Internal Rates of Return for the 1975-1980 Cohort Using Race and Education Mortality Adjustment Factors from the 1962-67 Cohort.

	1975-1980 Cohort	
	Baseline (from Table 5)	Fixed Mortality Adjustment Factors
Overall	1.40%	1.40%
Subgroups by Race		
Non-Hispanic Black	1.29	1.28
Hispanic	1.96	1.89
Non-Hispanic White	1.22	1.26
Subgroups by Education Level		
Less than High School	1.87	1.98
High School Graduate	1.43	1.49
Some College	1.38	1.51
4-Year College Graduate	1.30	1.14

Notes: Each cell reports an estimated internal rate of return (IRR). The IRR formula is as shown in the text. Column 1 is taken from Table 5. Column 2 shows IRR estimates for the 1975-1980 cohort using the mortality adjustment factors by race and education group used for the 1962-1967 cohort.

Table 7. Decompositions of IRR Gaps Across Subgroups.

	1962-1967 Cohort			1975-1980 Cohort		
	Total Gap	Mortality Gap	Earnings Gap	Total Gap	Mortality Gap	Earnings Gap
Non-Hispanic Black	-0.20%	-0.44%	0.24%	0.07%	-0.27%	0.34%
Hispanic	0.69%	0.38%	0.31%	0.74%	0.50%	0.24%
	Total Gap	Mortality Gap	Earnings Gap	Total Gap	Mortality Gap	Earnings Gap
Less than High School	0.44%	-0.79%	1.23%	0.57%	-1.01%	1.58%
High School Graduate	0.00%	-0.56%	0.56%	0.13%	-0.72%	0.85%
Some College	0.19%	-0.25%	0.44%	0.08%	-0.51%	0.59%

Notes: The total gaps are taken directly from Table 5. Gaps by race are relative to whites and gaps by education level are relative to 4-year college graduates. Thus, positive numbers indicate progressive sources of redistributions and negative numbers indicate regressive sources.

Appendices
(for posting online)

Appendix A
Sensitivity Analyses: Validation Exercise

This appendix reports briefly on sensitivity analyses with respect to the validation exercise. First, in Table A.1 we examine the sensitivity of our estimates to allowing for individuals with incomplete earnings records to have lower earnings than would be implied by using the prior, which is how we handle these individuals in the main text (reported as "baseline" in Table A.1 for comparison). We consider scenarios where these individuals' forecasted earnings are reduced to 90, 80 and 70 percent of what is estimated by the prior. Our estimated AIME distribution exhibits very little sensitivity to this dimension of our procedure, although there is some sensitivity in the bottom tail. Overall, using the prior to forecast earnings for these individuals results in the best match to the AIME distribution as estimated using SSA data.

Second, in Table A.2 we examine sensitivity to adjusting the persistence of workforce attachment (ψ in Equation 5 in the text). Recall that we parameterize ψ to match the share of covered workers in the SSA data (who have qualified earnings for at least 40 quarters). Our forecasts of the covered share are quite sensitive to the value of ψ, but the forecasted AIME distribution is not, except at the bottom decile. Again, our baseline forecasting approach results in the best match.

Third, in Table A.3 we allow for work during college for the highest-educated group (individuals with a four-year college degree). Our findings are not meaningfully sensitive to reasonable modifications along this dimension, primarily because these individuals already typically accrue 35+ years of earnings without working while in college. In Table A.3 we report results from scenarios focusing on college completers only. However, in unreported results omitted for brevity, we also verify that our findings are not sensitive to allowing for work during school for individuals who attend some college but do not complete a 4-year degree, and for high school graduates during the last year of high school.

Table A.1. Sensitivity of AIME Distribution Forecasts to Lowering Earnings Forecasts for Individuals With Fewer than Three Observed Earnings Records.

Variable		Forecasted Distribution				SSA-Data Distribution
		Baseline	0.9	0.8	0.7	
Mean		3709.92	3704.19	3698.25	3692.11	3726.89
		(45.32)	(45.29)	(45.26)	(45.24)	(39.16)
Uncovered Share		0.04	0.04	0.04	0.04	0.04
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Decile	1	705.98	671.86	638.08	591.53	722.30
		(47.26)	(45.51)	(43.95)	(43.47)**	(46.37)
	2	1555.34	1540.7	1535.74	1523.12	1511.64
		(64.96)	(65.79)	(66.69)	(67.09)	(67.31)
	3	2171.82	2170.69	2166.17	2154.19	2284.16
		(65.46)	(65.68)	(65.22)	(64.74)	(56.12)
	4	2880.61	2875.63	2871.77	2863.53	2942.21
		(64.10)	(64.13)	(64.77)	(64.52)	(45.69)
	5	3590.21	3584.45	3575.5	3574.42	3618.55
		(68.87)	(68.91)	(68.84)	(68.99)	(66.47)
	6	4273.65	4270.85	4270.85	4270.78	4291.35
		(66.87)	(67.12)	(66.97)	(66.92)	(51.27)
	7	5155.52	5155.52	5153.59	5153.2	5099.57
		(83.37)	(83.48)	(83.57)	(83.54)	(50.40)
	8	6072.04	6071.86	6071.86	6071.86	5942.88
		(65.72)	(65.70)	(65.72)	(65.73)	(57.53)
	9	6834.77	6834.77	6834.77	6834.77	6914.32
		(47.74)	(47.72)	(47.73)	(47.73)	(59.54)

Notes: In column 2, 3 and 4, forecasted earnings values (based on the prior as described in the text) for individuals with three or fewer observed earnings records are multiplied by 0.9, 0.8, and 0.7, respectively. Statistical significance for each forecasted estimate is reported with respect to the SSA-Data Distribution, as in Table 4.
***/**/* indicates statistical significance at the 0.01/0.05/0.10 levels relative to the SSA-data estimate.

Table A.2. Sensitivity of AIME Distribution Forecasts to Adjusting the Individual Persistence Parameter Governing Workforce Participation, ψ .

Variable		Forecasted Distribution				SSA-Data Distribution
		Baseline ($\psi = 0.86$)	$\psi = 0.9$	$\psi = 0.8$	$\psi = 0.7$	
Mean		3709.92	3703	3727.09	3734.6	3726.89
		(45.32)	(45.41)	(45.19)	(44.75)	(39.16)
Uncovered Share		0.04	0.06	0.01	0.01	0.04
		(0.00)	(0.00)***	(0.00)***	(0.00)***	(0.00)
Decile	1	705.98	592.13	845.94	933.77	722.30
		(47.26)	(58.55)	(45.56)*	(45.55)***	(46.37)
	2	1555.34	1541.41	1569.66	1612.36	1511.64
		(64.96)	(67.22)	(59.66)	(53.99)	(67.31)
	3	2171.82	2176.46	2164.12	2163.69	2284.16
		(65.46)	(64.90)	(66.29)	(63.72)	(56.12)
	4	2880.61	2898.99	2875.1	2850.9	2942.21
		(64.10)	(63.53)	(63.61)	(64.00)	(45.69)
	5	3590.21	3623.98	3579.31	3577.61	3618.55
		(68.87)	(68.84)	(68.99)	(67.62)	(66.47)
	6	4273.65	4293.84	4261.59	4248.42	4291.35
		(66.87)	(68.68)	(67.59)	(66.08)	(51.27)
	7	5155.52	5174.5	5141.6	5111.47	5099.57
		(83.37)	(84.41)	(83.83)	(84.03)	(50.40)
	8	6072.04	6074.95	6034.65	5986.92	5942.88
		(65.72)	(65.51)	(65.78)	(66.88)	(57.53)
	9	6834.77	6841.29	6815.86	6785.34	6914.32
		(47.74)	(46.90)	(47.08)	(47.29)*	(59.54)

Notes: Statistical significance for each forecasted estimate is reported with respect to the SSA-Data Distribution as in Table 4.
***/**/* indicates statistical significance at the 0.01/0.05/0.10 levels relative to the SSA-data estimate.

Table A.3. Sensitivity of AIME Distribution Forecasts to Allowing for Work During College for College-Educated Workers.

Variable		Forecasted Distribution			SSA-Data
		Baseline (zero earnings during college)	During-college earnings set to half of the average wage during first 3 post-college years	During-college earnings set to the average wage during first 3 post-college years	
Mean		3709.92	3710.39	3713.44	3726.89
		(45.32)	(45.33)	(45.34)	(39.16)
Uncovered Share		0.04	0.04	0.04	0.04
		(0.00)	(0.00)	(0.00)	(0.00)
Decile	1	705.98	708.57	708.57	722.30
		(47.26)	(47.25)	(47.25)	(46.37)
	2	1555.34	1555.71	1558.04	1511.64
		(64.96)	(65.03)	(65.04)	(67.31)
	3	2171.82	2171.82	2174.7	2284.16
		(65.46)	(65.52)	(65.52)	(56.12)
	4	2880.61	2880.61	2880.94	2942.21
		(64.10)	(64.16)	(64.25)	(45.69)
	5	3590.21	3590.21	3591.88	3618.55
		(68.87)	(68.88)	(69.05)	(66.47)
	6	4273.65	4273.65	4274.03	4291.35
		(66.87)	(66.88)	(67.03)	(51.27)
	7	5155.52	5157	5180.32	5099.57
		(83.37)	(83.39)	(83.78)	(50.40)
	8	6072.04	6072.04	6082.88	5942.88
		(65.72)	(65.75)	(65.60)	(57.53)
	9	6834.77	6834.77	6836.24	6914.32
		(47.74)	(47.73)	(47.45)	(59.54)

Notes: The first column replicates our baseline findings. The second column assigns college completers earnings equal to half of average earnings during the first three post-college years. The third column assigns college completers earnings equal to average earnings during the first three post-college years. Statistical significance for each forecasted estimate is reported with respect to the SSA-Data Distribution as in Table 4.
***/**/* indicates statistical significance at the 0.01/0.05/0.10 levels relative to the SSA-data estimate.

Appendix B
Sensitivity Analyses: Social Security IRR Forecasts for 1962-67 and 1975-80 Cohorts

This appendix reports sensitivity analyses analogous to what we show in Appendix A, but focusing on the IRR forecasts for cohorts born between 1962-67 and 1975-80. Tables B.1, B.2 and B.3 correspond to Tables A.1, A.2 and A.3, respectively.

Table B.1 shows that the IRR forecasts for young workers are generally insensitive to lowering the earnings forecasts for individuals with incomplete earnings records.

In Table B.2, there is modest sensitivity of the IRR forecasts overall and by group when we modify the individual persistence parameter governing workforce attachment (ψ from Equation 5 in the text). A general trend is that as ψ decreases, the Social Security IRR increases overall. There is no indication of differential sensitivity by race/ethnicity, but the IRR forecasts for less-educated individuals (less than a high school diploma and high school graduates) are modestly more sensitive than the IRR forecasts for more-educated individuals (individuals with some college and at least a 4-year college degree).

The forecasted IRR values rise overall as we lower ψ because a lower ψ generates less heterogeneity in the workforce attachment profiles across individuals, which increases total returns because of the non-linear PIA formula. The fact that less-educated groups are more affected implies that the convergence of workforce attachment profiles associated with a reduction in ψ has a larger effect on those groups, which is intuitive because the least-educated workers are more likely to have the lowest U_i values, which are farthest from any group trend. Also note that in Table B.2 we consider a scenario where we allow for the value of ψ to decline as workers approach the end of the career. Specifically, we set $\psi = 0.86$ for the first 20 years of the career, but beyond that we set $\psi = 0.86*[(50-e)/30]$, where e is experience years. This function for ψ is such that the forecasted workforce-attachment profiles better reflect empirical regularities with respect to workforce exit late in the career.

Overall, our preferred IRR estimates come from the baseline specification where we set $\psi = 0.86$, which performs best in our validation exercise. Changes to ψ have no qualitative implications for our estimated redistributions across racial/ethnic groups. With respect to the education groups, lower values of ψ imply somewhat more progressive redistributions than what we report in the main text, although the changes are small.

Finally, in Table B.3, our IRR estimates are also only modestly affected by allowing for work during college among college graduates. The effect of this adjustment is to lower the IRR for these individuals. The reason is that they (and their employers) pay into Social Security on these earnings, but the earnings have little bearing on their final AIME values because AIME depends only on the highest 35 years of earnings and these individuals typically have more than 35 years of earnings, even without work during college. Like with Table A.3, we also note that in unreported results omitted for brevity, we verify our IRR forecasts are not meaningfully sensitive to allowing for work during school for individuals who attend some college but do not complete a 4-year degree, and for high school graduates during the last year of high school.

Table B.1. Sensitivity of IRR Forecasts to Lowering Earnings Forecasts for Individuals With Fewer than Three Observed Earnings Records.

	Baseline		0.9		0.8		0.7	
	1962-67	1975-80	1962-67	1975-80	1962-67	1975-80	1962-67	1975-80
Overall	1.10%	1.40%	1.10%	1.40%	1.10%	1.40%	1.10%	1.40%
Non-Hispanic Black	0.83%	1.29%	0.83%	1.29%	0.83%	1.29%	0.83%	1.29%
Hispanic	1.72%	1.96%	1.72%	1.96%	1.72%	1.96%	1.72%	1.96%
Non-Hispanic White	1.03%	1.22%	1.03%	1.22%	1.03%	1.23%	1.03%	1.22%
Less than High School	1.45%	1.87%	1.45%	1.87%	1.45%	1.87%	1.46%	1.88%
High School	1.01%	1.43%	1.01%	1.44%	1.01%	1.44%	1.01%	1.44%
Some College	1.20%	1.38%	1.20%	1.38%	1.20%	1.38%	1.20%	1.38%
College Graduate	1.01%	1.30%	1.01%	1.30%	1.01%	1.30%	1.01%	1.30%

Notes: This table is the analog to Appendix Table A.1 for the IRR forecasts. In the second, third and fourth vertical panels, forecasted earnings values (based on the prior as described in the text) for individuals with three or fewer observed earnings records are multiplied by 0.9, 0.8, and 0.7, respectively.

Table B.2. Sensitivity of IRR Forecasts to Adjusting the Individual Persistence Parameter Governing Workforce Participation, ψ .

	Baseline (ψ =0.86)		$\psi = 0.9$		$\psi = 0.8$		$\psi = 0.7$		Variable ψ	
	1962-67	1975-80	1962-67	1975-80	1962-67	1975-80	1962-67	1975-80	1962-67	1975-80
Overall	1.10%	1.40%	1.09%	1.39%	1.12%	1.42%	1.13%	1.43%	1.15%	1.46%
Non-Hispanic Black	0.83%	1.29%	0.81%	1.27%	0.85%	1.32%	0.88%	1.34%	0.92%	1.36%
Hispanic	1.72%	1.96%	1.71%	1.95%	1.74%	1.99%	1.76%	2.00%	1.77%	2.04%
Non-Hispanic White	1.03%	1.22%	1.02%	1.21%	1.05%	1.25%	1.06%	1.25%	1.07%	1.27%
Less than High School	1.45%	1.87%	1.44%	1.87%	1.47%	1.89%	1.51%	1.95%	1.57%	2.02%
High School	1.01%	1.43%	0.98%	1.42%	1.03%	1.48%	1.05%	1.48%	1.08%	1.51%
Some College	1.20%	1.38%	1.19%	1.37%	1.21%	1.42%	1.21%	1.43%	1.23%	1.44%
College Graduate	1.01%	1.30%	1.00%	1.28%	1.02%	1.31%	1.02%	1.31%	1.03%	1.32%

Notes: This table is the analog to Appendix Table A.2 for the IRR forecasts. Note that in the final vertical panel we include an additional set of estimates that reduce the value of ψ as workers approach retirement. Specifically, we set $\psi = 0.86$ for the first 20 years of the career, but beyond that we set $\psi = 0.86*[(50-e)/30]$, where e is experience years. This function for ψ makes it so the forecasted workforce-attachment profiles better reflect empirical regularities with respect to workforce exit.

Table B.3. Sensitivity to Allowing for Work During College for College-Educated Workers.

	Baseline (zero earnings during college)		During-college earnings set to half of the average wage during first 3 post-college years		During-college earnings set to the average wage during first 3 post-college years	
	1962-1967	1975-1980	1962-1967	1975-1980	1962-1967	1975-1980
Overall	1.10%	1.40%	1.08%	1.37%	1.06%	1.35%
Non-Hispanic Black	0.83%	1.29%	0.82%	1.26%	0.81%	1.23%
Hispanic	1.72%	1.96%	1.71%	1.95%	1.70%	1.93%
Non-Hispanic White	1.03%	1.22%	1.00%	1.20%	0.98%	1.17%
Less than High School	1.45%	1.87%	1.45%	1.87%	1.45%	1.87%
High School	1.01%	1.43%	1.01%	1.43%	1.01%	1.43%
Some College	1.20%	1.38%	1.20%	1.38%	1.20%	1.38%
College Graduate	1.01%	1.30%	0.94%	1.23%	0.88%	1.17%

Notes: This table is the analog to Appendix Table A.3 for the IRR forecasts.

Appendix C
Mortality Rates and Sensitivity

In this appendix we report on the sensitivity of our findings to using mortality tables that start age age-21 and age-65, rather than the approach used in the text of allowing for mortality starting at age-35. We replicate the analysis shown in Table 5 in the main text using these two alternatives in appendix Tables C.1 and C.2. As indicated briefly in the text, the findings are qualitatively unaffected if we use mortality tables applied retroactively to age-21 for individuals in our data. However, the findings change dramatically if we do not apply mortality tables until age-65; effectively assuming that all individuals survive until full retirement.

We also show the mortality tables used in our calculations for the 1962-67 and 1975-80 cohorts in Tables C.3 and C.4, respectively. As described in the text, we start with the cohort life tables for men taken from Bell and Miller (2005) to construct baseline mortality rates and then make relative adjustments to these rates by race and education level using CDC and Census data following the approach of Bound et al. (2014). We are not aware of any other recent economic studies that provide detailed mortality tables. The mortality rate data are also available in .csv files from the authors.

Table C.1. Forecasted Internal Rates of Return, Overall and by Subgroup, for Each Forecasting Cohort. Mortality Tables Retroactive to Age-21.

	Cohorts	
	1962-1967	1975-1980
Overall	1.09%	1.39%
<u>Subgroups by Race</u>		
Non-Hispanic Black	0.81	1.28
Hispanic	1.71	1.95
Non-Hispanic White	1.02	1.22
<u>Subgroups by Education Level</u>		
Less than High School	1.43	1.86
High School Graduate	0.99	1.43
Some College	1.19	1.37
4-Year College Graduate	1.00	1.30

Notes: Each cell reports an estimated internal rate of return (IRR). The IRR formula is as shown in the text. Mortality rates are applied retroactively to age-21.

Table C.2. Forecasted Internal Rates of Return, Overall and by Subgroup, for Each Forecasting Cohort. Mortality Table Implementation Delayed Until Age-65.

	Cohorts	
	1962-1967	1975-1980
Overall	1.49%	1.69%
<u>Subgroups by Race</u>		
Non-Hispanic Black	1.47	1.74
Hispanic	2.05	2.20
Non-Hispanic White	1.40	1.51
<u>Subgroups by Education Level</u>		
Less than High School	2.00	2.26
High School Graduate	1.51	1.86
Some College	1.56	1.70
4-Year College Graduate	1.26	1.47

Notes: Each cell reports an estimated internal rate of return (IRR). The IRR formula is as shown in the text. Mortality rates are not applied until age-65; i.e., these results assume all individuals survive until age-65.

Table C.3. Mortality Rates Used in our Forecasts for the 1962-67 Cohort.

	Non-Hispanic Black				Hispanic				Non-Hispanic White			
Age	Edu1	Edu2	Edu3	Edu4	Edu1	Edu2	Edu3	Edu4	Edu1	Edu2	Edu3	Edu4
21	0.00604	0.00358	0.00227	0.00148	0.002	0.00141	0.00099	0.00069	0.00405	0.00203	0.00119	0.0007
22	0.00584	0.00347	0.0022	0.00143	0.00193	0.00137	0.00095	0.00066	0.00392	0.00196	0.00115	0.00068
23	0.00551	0.00327	0.00207	0.00135	0.00182	0.00129	0.0009	0.00063	0.0037	0.00185	0.00109	0.00064
24	0.00554	0.00329	0.00209	0.00135	0.00183	0.0013	0.00091	0.00063	0.00372	0.00186	0.0011	0.00064
25	0.00548	0.00325	0.00206	0.00134	0.00181	0.00128	0.0009	0.00062	0.00368	0.00184	0.00108	0.00063
26	0.00574	0.00341	0.00216	0.0014	0.0019	0.00134	0.00094	0.00065	0.00386	0.00193	0.00113	0.00066
27	0.00574	0.00341	0.00216	0.0014	0.0019	0.00134	0.00094	0.00065	0.00386	0.00193	0.00113	0.00066
28	0.00604	0.00358	0.00227	0.00148	0.002	0.00141	0.00099	0.00069	0.00405	0.00203	0.00119	0.0007
29	0.0065	0.00386	0.00245	0.00159	0.00215	0.00152	0.00106	0.00074	0.00436	0.00218	0.00128	0.00075
30	0.00673	0.004	0.00253	0.00164	0.00222	0.00158	0.0011	0.00077	0.00452	0.00226	0.00133	0.00078
31	0.00696	0.00413	0.00262	0.0017	0.0023	0.00163	0.00114	0.00079	0.00468	0.00234	0.00138	0.00081
32	0.00729	0.00433	0.00274	0.00178	0.00241	0.00171	0.00119	0.00083	0.0049	0.00245	0.00144	0.00084
33	0.00795	0.00472	0.00299	0.00194	0.00263	0.00186	0.0013	0.0009	0.00534	0.00267	0.00157	0.00092
34	0.00835	0.00496	0.00314	0.00204	0.00276	0.00195	0.00136	0.00095	0.00561	0.0028	0.00165	0.00097
35	0.00786	0.00487	0.00312	0.00239	0.00232	0.00238	0.00147	0.00112	0.00588	0.00285	0.0019	0.0011
36	0.0072	0.00446	0.00286	0.00219	0.00213	0.00218	0.00134	0.00103	0.00538	0.00261	0.00174	0.00101
37	0.00659	0.00409	0.00262	0.00201	0.00195	0.00199	0.00123	0.00094	0.00493	0.00239	0.0016	0.00092
38	0.0068	0.00422	0.0027	0.00207	0.00201	0.00206	0.00127	0.00097	0.00509	0.00247	0.00165	0.00095
39	0.00723	0.00448	0.00287	0.0022	0.00214	0.00219	0.00135	0.00103	0.0054	0.00262	0.00175	0.00101
40	0.0078	0.00484	0.0031	0.00237	0.00231	0.00236	0.00146	0.00111	0.00583	0.00283	0.00189	0.00109
41	0.00844	0.00523	0.00335	0.00257	0.00249	0.00255	0.00157	0.0012	0.00631	0.00306	0.00204	0.00118
42	0.00895	0.00555	0.00356	0.00272	0.00264	0.00271	0.00167	0.00128	0.00669	0.00325	0.00217	0.00125
43	0.00971	0.00602	0.00386	0.00295	0.00287	0.00294	0.00181	0.00138	0.00726	0.00352	0.00235	0.00136
44	0.01049	0.00651	0.00417	0.00319	0.0031	0.00317	0.00196	0.0015	0.00784	0.00381	0.00254	0.00147
45	0.01084	0.0076	0.00485	0.00377	0.00306	0.00362	0.00236	0.00195	0.00787	0.00424	0.00297	0.00195
46	0.01168	0.00819	0.00522	0.00407	0.0033	0.0039	0.00254	0.0021	0.00848	0.00457	0.0032	0.0021
47	0.01238	0.00867	0.00553	0.00431	0.00349	0.00413	0.00269	0.00222	0.00899	0.00484	0.00339	0.00223
48	0.01292	0.00906	0.00578	0.0045	0.00365	0.00431	0.00281	0.00232	0.00939	0.00506	0.00354	0.00233

49	0.01336	0.00936	0.00597	0.00465	0.00377	0.00446	0.0029	0.0024	0.0097	0.00522	0.00366	0.00241
50	0.01382	0.00969	0.00618	0.00481	0.0039	0.00461	0.003	0.00248	0.01004	0.00541	0.00378	0.00249
51	0.0144	0.01009	0.00644	0.00501	0.00407	0.0048	0.00313	0.00259	0.01046	0.00563	0.00394	0.00259
52	0.01521	0.01066	0.0068	0.00529	0.00429	0.00507	0.00331	0.00273	0.01104	0.00595	0.00416	0.00274
53	0.01622	0.01137	0.00725	0.00565	0.00458	0.00541	0.00353	0.00292	0.01178	0.00634	0.00444	0.00292
54	0.01746	0.01224	0.00781	0.00608	0.00493	0.00583	0.0038	0.00314	0.01268	0.00683	0.00478	0.00315
55	0.01342	0.01162	0.00737	0.00638	0.00517	0.00607	0.00397	0.00383	0.01087	0.00687	0.00525	0.00366
56	0.01455	0.0126	0.00799	0.00692	0.00561	0.00659	0.00431	0.00416	0.01179	0.00745	0.00569	0.00397
57	0.01577	0.01365	0.00865	0.0075	0.00608	0.00714	0.00467	0.0045	0.01277	0.00807	0.00617	0.0043
58	0.01708	0.01479	0.00938	0.00812	0.00658	0.00773	0.00506	0.00488	0.01383	0.00874	0.00668	0.00466
59	0.0185	0.01601	0.01015	0.00879	0.00713	0.00837	0.00548	0.00528	0.01498	0.00947	0.00723	0.00504
60	0.02008	0.01738	0.01102	0.00955	0.00774	0.00909	0.00595	0.00574	0.01627	0.01028	0.00785	0.00548
61	0.02187	0.01893	0.012	0.0104	0.00843	0.0099	0.00648	0.00625	0.01771	0.0112	0.00855	0.00596
62	0.02387	0.02066	0.0131	0.01135	0.0092	0.0108	0.00707	0.00682	0.01933	0.01222	0.00933	0.00651
63	0.02613	0.02261	0.01434	0.01242	0.01007	0.01182	0.00774	0.00746	0.02116	0.01337	0.01022	0.00712
64	0.02864	0.02478	0.01572	0.01361	0.01103	0.01296	0.00848	0.00818	0.02319	0.01466	0.0112	0.00781
65	0.02148	0.02321	0.01675	0.01491	0.01157	0.01404	0.00985	0.01026	0.02016	0.01561	0.01332	0.01035
66	0.02355	0.02544	0.01836	0.01635	0.01269	0.01539	0.0108	0.01124	0.0221	0.01711	0.0146	0.01134
67	0.02571	0.02778	0.02004	0.01785	0.01385	0.0168	0.01179	0.01227	0.02413	0.01868	0.01594	0.01238
68	0.02796	0.03021	0.0218	0.01941	0.01506	0.01827	0.01282	0.01335	0.02624	0.02031	0.01734	0.01347
69	0.03032	0.03276	0.02364	0.02105	0.01634	0.01981	0.01391	0.01447	0.02846	0.02203	0.0188	0.0146
70	0.03302	0.03567	0.02574	0.02292	0.01779	0.02157	0.01514	0.01576	0.03099	0.02399	0.02047	0.0159
71	0.03598	0.03888	0.02805	0.02498	0.01939	0.02351	0.0165	0.01718	0.03377	0.02614	0.02231	0.01733
72	0.039	0.04214	0.03041	0.02707	0.02101	0.02548	0.01789	0.01862	0.03661	0.02834	0.02418	0.01879
73	0.04201	0.04539	0.03275	0.02916	0.02263	0.02745	0.01927	0.02006	0.03943	0.03052	0.02605	0.02023
74	0.0452	0.04883	0.03524	0.03137	0.02435	0.02953	0.02073	0.02158	0.04243	0.03284	0.02802	0.02177
75	0.03829	0.04727	0.03511	0.03058	0.02329	0.03037	0.02387	0.02579	0.03987	0.0361	0.03265	0.02732
76	0.04175	0.05154	0.03829	0.03334	0.0254	0.03312	0.02603	0.02812	0.04347	0.03937	0.0356	0.02979
77	0.04548	0.05615	0.04171	0.03632	0.02767	0.03608	0.02836	0.03064	0.04736	0.04289	0.03879	0.03246
78	0.04948	0.06109	0.04538	0.03952	0.0301	0.03925	0.03085	0.03333	0.05152	0.04666	0.0422	0.03531
79	0.05392	0.06657	0.04945	0.04306	0.0328	0.04277	0.03362	0.03632	0.05615	0.05085	0.04599	0.03848

80	0.05886	0.07267	0.05399	0.04701	0.03581	0.04669	0.0367	0.03965	0.06129	0.05551	0.0502	0.042
81	0.06471	0.07989	0.05935	0.05168	0.03936	0.05133	0.04034	0.04359	0.06738	0.06102	0.05519	0.04617
82	0.0719	0.08877	0.06595	0.05743	0.04374	0.05704	0.04483	0.04843	0.07487	0.06781	0.06132	0.05131
83	0.08072	0.09966	0.07404	0.06447	0.04911	0.06403	0.05033	0.05437	0.08406	0.07612	0.06885	0.0576
84	0.09098	0.11233	0.08345	0.07267	0.05535	0.07217	0.05673	0.06129	0.09474	0.0858	0.0776	0.06493
85	0.08494	0.0881	0.07278	0.07883	0.07051	0.07844	0.07341	0.08369	0.10473	0.08835	0.08996	0.08868
86	0.09493	0.09846	0.08134	0.0881	0.0788	0.08767	0.08204	0.09353	0.11705	0.09874	0.10054	0.0991
87	0.10534	0.10926	0.09026	0.09776	0.08744	0.09728	0.09103	0.10379	0.12989	0.10957	0.11156	0.10997
88	0.11611	0.12043	0.0995	0.10776	0.09638	0.10723	0.10034	0.1144	0.14317	0.12077	0.12297	0.12122
89	0.12731	0.13204	0.10909	0.11815	0.10567	0.11756	0.11002	0.12543	0.15697	0.13241	0.13482	0.1329
90	0.13903	0.1442	0.11913	0.12902	0.1154	0.12839	0.12015	0.13698	0.17143	0.14461	0.14724	0.14514
91	0.15139	0.15702	0.12972	0.1405	0.12566	0.1398	0.13083	0.14916	0.18667	0.15746	0.16033	0.15805
92	0.16451	0.17063	0.14097	0.15267	0.13655	0.15192	0.14217	0.16208	0.20284	0.17111	0.17422	0.17174
93	0.17849	0.18513	0.15294	0.16564	0.14815	0.16482	0.15425	0.17585	0.22008	0.18565	0.18902	0.18633
94	0.19343	0.20063	0.16575	0.17951	0.16056	0.17862	0.16716	0.19058	0.2385	0.20119	0.20485	0.20193
95	0.2081	0.21584	0.17832	0.19312	0.17273	0.19217	0.17984	0.20503	0.25659	0.21645	0.22038	0.21725
96	0.22226	0.23054	0.19046	0.20627	0.18449	0.20525	0.19208	0.21899	0.27406	0.23118	0.23539	0.23204
97	0.23564	0.24441	0.20191	0.21868	0.19559	0.2176	0.20364	0.23216	0.29055	0.24509	0.24955	0.246
98	0.24798	0.25721	0.21249	0.23014	0.20584	0.229	0.2143	0.24433	0.30577	0.25793	0.26262	0.25888
99	0.25895	0.26858	0.22189	0.24031	0.21494	0.23913	0.22378	0.25513	0.31929	0.26934	0.27424	0.27033
100	0.27046	0.28053	0.23175	0.251	0.2245	0.24976	0.23373	0.26647	0.33348	0.28131	0.28643	0.28235

Notes: The numbers in this table are mortality rates for racial/ethnic-by-education groups. Edu1, Edu2, Edu3 and Edu4 correspond to groups Less than High School, High School Graduate, Some College, and 4-Year College Graduate, respectively. The mortality rates are constructed as described in the text. Our primary results begin mortality calculations at age-35.

Table C.4. Mortality Rates Used in our Forecasts for the 1975-80 Cohort.

	Non-Hispanic Black				Hispanic				Non-Hispanic White			
Age	Edu1	Edu2	Edu3	Edu4	Edu1	Edu2	Edu3	Edu4	Edu1	Edu2	Edu3	Edu4
21	0.00574	0.00291	0.00165	0.00072	0.00194	0.0011	0.00081	0.00032	0.00509	0.00218	0.00128	0.00045
22	0.00563	0.00286	0.00162	0.0007	0.00191	0.00108	0.00079	0.00031	0.00499	0.00214	0.00126	0.00044
23	0.00588	0.00299	0.00169	0.00074	0.00199	0.00113	0.00082	0.00032	0.00521	0.00223	0.00131	0.00046
24	0.00581	0.00295	0.00167	0.00073	0.00197	0.00111	0.00082	0.00032	0.00515	0.0022	0.0013	0.00046
25	0.00563	0.00286	0.00162	0.0007	0.00191	0.00108	0.00079	0.00031	0.00499	0.00214	0.00126	0.00044
26	0.00503	0.00256	0.00145	0.00063	0.00171	0.00097	0.00071	0.00028	0.00446	0.00191	0.00113	0.00039
27	0.00472	0.0024	0.00136	0.00059	0.0016	0.0009	0.00066	0.00026	0.00418	0.00179	0.00105	0.00037
28	0.00451	0.00229	0.0013	0.00056	0.00153	0.00086	0.00063	0.00025	0.00399	0.00171	0.00101	0.00035
29	0.00461	0.00234	0.00133	0.00058	0.00156	0.00088	0.00065	0.00025	0.00409	0.00175	0.00103	0.00036
30	0.00475	0.00241	0.00137	0.00059	0.00161	0.00091	0.00067	0.00026	0.00421	0.0018	0.00106	0.00037
31	0.005	0.00254	0.00144	0.00063	0.00169	0.00096	0.0007	0.00027	0.00443	0.0019	0.00112	0.00039
32	0.00493	0.0025	0.00142	0.00062	0.00167	0.00095	0.00069	0.00027	0.00437	0.00187	0.0011	0.00039
33	0.00517	0.00263	0.00149	0.00065	0.00175	0.00099	0.00073	0.00028	0.00459	0.00196	0.00116	0.00041
34	0.00549	0.00279	0.00158	0.00069	0.00186	0.00105	0.00077	0.0003	0.00487	0.00208	0.00123	0.00043
35	0.00395	0.00273	0.00175	0.00092	0.00123	0.00107	0.00082	0.00041	0.00413	0.00221	0.00139	0.00061
36	0.00426	0.00294	0.00188	0.00099	0.00133	0.00116	0.00089	0.00044	0.00445	0.00238	0.0015	0.00066
37	0.00459	0.00317	0.00203	0.00107	0.00143	0.00125	0.00096	0.00048	0.0048	0.00257	0.00162	0.00071
38	0.00495	0.00342	0.00219	0.00115	0.00155	0.00135	0.00103	0.00052	0.00518	0.00277	0.00175	0.00076
39	0.00537	0.00371	0.00237	0.00125	0.00168	0.00146	0.00112	0.00056	0.00561	0.003	0.00189	0.00083
40	0.00578	0.00399	0.00256	0.00134	0.0018	0.00157	0.0012	0.0006	0.00605	0.00323	0.00204	0.00089
41	0.00619	0.00428	0.00274	0.00144	0.00193	0.00169	0.00129	0.00065	0.00648	0.00346	0.00218	0.00095
42	0.00666	0.0046	0.00294	0.00155	0.00208	0.00181	0.00138	0.00069	0.00696	0.00372	0.00235	0.00103
43	0.00712	0.00492	0.00315	0.00166	0.00222	0.00194	0.00148	0.00074	0.00745	0.00398	0.00251	0.0011
44	0.00759	0.00524	0.00335	0.00176	0.00237	0.00207	0.00158	0.00079	0.00794	0.00424	0.00267	0.00117
45	0.00705	0.00531	0.00326	0.0019	0.00246	0.00241	0.00172	0.00096	0.00776	0.00406	0.00266	0.00128
46	0.00752	0.00567	0.00348	0.00203	0.00263	0.00257	0.00184	0.00103	0.00827	0.00434	0.00284	0.00137
47	0.0079	0.00596	0.00365	0.00213	0.00276	0.0027	0.00193	0.00108	0.00869	0.00456	0.00299	0.00144
48	0.00822	0.00619	0.0038	0.00222	0.00287	0.00281	0.00201	0.00112	0.00904	0.00474	0.0031	0.00149

49	0.00846	0.00638	0.00391	0.00229	0.00295	0.00289	0.00207	0.00116	0.00931	0.00488	0.0032	0.00154
50	0.00876	0.0066	0.00404	0.00236	0.00306	0.00299	0.00214	0.00119	0.00963	0.00505	0.00331	0.00159
51	0.00916	0.0069	0.00423	0.00247	0.0032	0.00313	0.00224	0.00125	0.01007	0.00528	0.00346	0.00166
52	0.00967	0.00729	0.00447	0.00261	0.00338	0.0033	0.00236	0.00132	0.01064	0.00557	0.00365	0.00176
53	0.01032	0.00778	0.00477	0.00279	0.0036	0.00353	0.00252	0.00141	0.01135	0.00595	0.0039	0.00187
54	0.01115	0.0084	0.00515	0.00301	0.00389	0.00381	0.00272	0.00152	0.01226	0.00643	0.00421	0.00202
55	0.01113	0.00983	0.00632	0.00435	0.00419	0.00453	0.00353	0.00215	0.01164	0.00667	0.00496	0.00262
56	0.0121	0.01069	0.00687	0.00472	0.00455	0.00493	0.00384	0.00234	0.01265	0.00726	0.00539	0.00285
57	0.01313	0.0116	0.00746	0.00513	0.00494	0.00534	0.00416	0.00254	0.01373	0.00787	0.00585	0.0031
58	0.01424	0.01258	0.00809	0.00556	0.00536	0.0058	0.00452	0.00275	0.0149	0.00854	0.00635	0.00336
59	0.01544	0.01363	0.00877	0.00603	0.00581	0.00628	0.00489	0.00298	0.01615	0.00926	0.00688	0.00364
60	0.01676	0.0148	0.00952	0.00654	0.0063	0.00682	0.00531	0.00324	0.01752	0.01005	0.00747	0.00395
61	0.01826	0.01613	0.01037	0.00713	0.00687	0.00743	0.00579	0.00353	0.0191	0.01095	0.00814	0.00431
62	0.02001	0.01768	0.01137	0.00782	0.00753	0.00815	0.00635	0.00387	0.02093	0.012	0.00892	0.00472
63	0.02203	0.01946	0.01252	0.0086	0.00829	0.00897	0.00699	0.00426	0.02304	0.01321	0.00982	0.0052
64	0.0243	0.02146	0.0138	0.00949	0.00914	0.00989	0.00771	0.0047	0.02541	0.01457	0.01084	0.00573
65	0.02187	0.02124	0.01449	0.01103	0.01042	0.01104	0.00846	0.00643	0.02309	0.01507	0.01256	0.00742
66	0.02408	0.0234	0.01595	0.01214	0.01147	0.01216	0.00932	0.00708	0.02543	0.0166	0.01384	0.00818
67	0.02638	0.02563	0.01748	0.0133	0.01257	0.01332	0.01021	0.00776	0.02785	0.01818	0.01516	0.00896
68	0.02873	0.02791	0.01903	0.01449	0.01369	0.0145	0.01112	0.00845	0.03033	0.0198	0.01651	0.00975
69	0.03118	0.03029	0.02066	0.01572	0.01485	0.01574	0.01207	0.00917	0.03292	0.02149	0.01791	0.01059
70	0.03395	0.03298	0.02249	0.01712	0.01617	0.01714	0.01314	0.00998	0.03584	0.0234	0.0195	0.01153
71	0.03704	0.03598	0.02454	0.01867	0.01764	0.0187	0.01433	0.01089	0.0391	0.02553	0.02128	0.01258
72	0.04014	0.039	0.02659	0.02024	0.01912	0.02027	0.01554	0.01181	0.04238	0.02767	0.02306	0.01363
73	0.04326	0.04203	0.02866	0.02181	0.02061	0.02184	0.01674	0.01272	0.04568	0.02982	0.02486	0.01469
74	0.04653	0.04521	0.03083	0.02346	0.02217	0.02349	0.01801	0.01369	0.04913	0.03207	0.02674	0.0158
75	0.03568	0.041	0.03007	0.02589	0.02304	0.02535	0.02272	0.01867	0.04046	0.03163	0.03104	0.02239
76	0.03896	0.04477	0.03284	0.02828	0.02516	0.02769	0.02481	0.02039	0.04419	0.03454	0.03389	0.02445
77	0.04243	0.04877	0.03577	0.0308	0.0274	0.03016	0.02703	0.02221	0.04813	0.03762	0.03692	0.02663
78	0.04606	0.05294	0.03883	0.03343	0.02974	0.03273	0.02934	0.0241	0.05224	0.04084	0.04007	0.0289
79	0.05006	0.05753	0.0422	0.03633	0.03232	0.03557	0.03188	0.02619	0.05678	0.04438	0.04355	0.03141

80	0.0545	0.06264	0.04595	0.03956	0.03519	0.03873	0.03471	0.02852	0.06182	0.04832	0.04742	0.0342
81	0.0599	0.06884	0.05049	0.04347	0.03868	0.04257	0.03815	0.03134	0.06794	0.05311	0.05211	0.03758
82	0.06686	0.07684	0.05636	0.04852	0.04317	0.04751	0.04258	0.03499	0.07583	0.05928	0.05816	0.04195
83	0.0757	0.087	0.06382	0.05494	0.04888	0.0538	0.04822	0.03961	0.08586	0.06712	0.06586	0.0475
84	0.08613	0.09898	0.0726	0.06251	0.05561	0.0612	0.05485	0.04507	0.09768	0.07636	0.07492	0.05404
85	0.07547	0.09378	0.07579	0.06626	0.05787	0.06342	0.06701	0.05486	0.09316	0.08454	0.09198	0.07271
86	0.08483	0.10541	0.08519	0.07447	0.06504	0.07128	0.07532	0.06167	0.10471	0.09502	0.10339	0.08173
87	0.09445	0.11737	0.09486	0.08292	0.07242	0.07937	0.08387	0.06866	0.11659	0.1058	0.11512	0.091
88	0.10429	0.1296	0.10473	0.09156	0.07997	0.08763	0.0926	0.07581	0.12873	0.11682	0.1271	0.10048
89	0.11444	0.14221	0.11493	0.10047	0.08775	0.09616	0.10161	0.08319	0.14126	0.12819	0.13947	0.11026
90	0.12501	0.15534	0.12554	0.10975	0.09585	0.10504	0.111	0.09087	0.1543	0.14003	0.15235	0.12044
91	0.13615	0.16919	0.13673	0.11953	0.1044	0.11441	0.12089	0.09898	0.16806	0.15251	0.16594	0.13118
92	0.14803	0.18395	0.14866	0.12996	0.11351	0.12439	0.13144	0.10761	0.18273	0.16582	0.18042	0.14262
93	0.16077	0.19978	0.16145	0.14114	0.12327	0.13509	0.14275	0.11687	0.19844	0.18008	0.19594	0.15489
94	0.17448	0.21682	0.17522	0.15318	0.13379	0.14661	0.15492	0.12684	0.21537	0.19544	0.21265	0.1681
95	0.18795	0.23355	0.18875	0.165	0.14411	0.15793	0.16688	0.13663	0.23199	0.21053	0.22906	0.18108
96	0.20094	0.2497	0.2018	0.17642	0.15408	0.16885	0.17842	0.14608	0.24804	0.22509	0.2449	0.1936
97	0.21321	0.26494	0.21412	0.18718	0.16348	0.17915	0.18931	0.15499	0.26318	0.23883	0.25985	0.20542
98	0.2245	0.27898	0.22546	0.1971	0.17214	0.18864	0.19934	0.1632	0.27712	0.25148	0.27362	0.2163
99	0.23455	0.29146	0.23555	0.20592	0.17985	0.19708	0.20826	0.1705	0.28952	0.26273	0.28586	0.22598
100	0.24506	0.30452	0.2461	0.21514	0.1879	0.20591	0.21759	0.17814	0.30249	0.2745	0.29866	0.2361

Notes: The numbers in this table are mortality rates for racial/ethnic-by-education groups. Edu1, Edu2, Edu3 and Edu4 correspond to groups Less than High School, High School Graduate, Some College, and 4-Year College Graduate, respectively. The mortality rates are constructed as described in the text. Our primary results begin mortality calculations at age-35.