Working Paper Series
Congressional Budget Office
Washington, D.C.

Modeling Individual Earnings in CBO's Long-Term Microsimulation Model

Jonathan A. Schwabish
Congressional Budget Office
jonathan.schwabish@cbo.gov

Julie H. Topoleski
Congressional Budget Office
julie.topoleski@cbo.gov

June 2013

Working Paper 2013-04

The authors thank Linda Bilheimer, Melissa Favreault, Ed Harris, Joyce Manchester, Noah Meyerson, Shinichi Nishiyama, Ben Page, Charles Pineles-Mark, John Sabelhaus, Stephen Shore, Michael Simpson, and Karen Smith for comments and suggestions.

Abstract

This paper describes the methods developed to project individual earnings in the Congressional Budget Office Long-Term (CBOLT) microsimulation model. CBOLT is used to assess the fiscal situations of the Social Security system and the federal government as a whole. Unlike many other models that project Social Security's finances, CBOLT projects behavior at the individual level. For each individual in the model, CBOLT projects levels of educational attainment, transitions in and out of marriage, labor force participation and employment transitions, immigration and emigration, and claiming patterns for Social Security benefits. An important feature of CBOLT is that it models each worker's annual earnings over that worker's lifetime. Those lifetime earnings patterns are the key determinants of individual payroll taxes paid and Social Security benefits received, and thus of aggregate Social Security finances.

In CBO's modeling, the historical pattern of rising earnings inequality continues for the next two decades, but earnings inequality generally ceases to rise by the mid-2030s. The method for projecting individuals' earnings that was developed for CBOLT and described in this paper closely follows the method first documented by Carroll (1992), but it is also informed by the work of other researchers. In general, individual earnings are perturbed by a pair of estimated earnings "shocks." The first shock is permanent and measures the long-run gap between a worker's earnings and the average earnings of that worker's group (where the group may be defined by age, sex, and education). Permanent shocks could be caused by, for example, receiving a promotion or attaining a higher level of education. The second shock is transitory and measures any additional but temporary variation in a person's earnings. Transitory shocks could arise from, for example, receiving a bonus or missing work because of illness.

Introduction

This paper describes how individual earnings are projected in the Congressional Budget Office Long-Term (CBOLT) microsimulation model, which is used to assess the fiscal situation of the Social Security system and the federal government as a whole. Unlike many other models that project Social Security finances, CBOLT projects behavior at the individual level, focusing on how individual earnings change over each worker's lifetime. Those lifetime earnings patterns are the key determinants of individual Social Security payroll taxes paid and benefits received.

Earnings in CBOLT are modeled as the sum of predicted earnings for an individual with certain characteristics, an individual-specific differential, and annual permanent and transitory "shocks." The approach to modeling those shocks closely follows the method in Carroll (1992), but it is also informed by the work of other researchers. The earnings patterns that result from these two shocks, with the calibration described below, cause earnings inequality (as measured by the Gini coefficient) to continue to rise through about 2035 (as it has during the past few decades) but to cease rising thereafter.

In this paper, we describe how we estimate permanent and transitory variances of individuals' earnings using 26 years of administrative earnings data provided by the Social Security Administration. We then describe how those variances are used in the microsimulation model to project earnings over time, and we document the resulting patterns of projected earnings.

A Brief Description of CBOLT

CBOLT is a dynamic microsimulation model of the U.S. population, economy, and federal budget.[1] A dynamic microsimulation model starts with individual-level data from a representative sample of the population and projects demographic and economic outcomes for that sample through time. For each individual in the sample, CBOLT simulates annual demographic transitions (for example, fertility, death, immigration and emigration, marital transitions, and marital pairings), economic transitions (for example, labor force participation, hours worked, and earnings), and Social Security-related finances (for example, payroll taxes received and benefits paid). Individuals are also linked to their spouses (current, former, and deceased) and to their children. This allows Congressional Budget Office (CBO) analysts to model, for example, spousal and other auxiliary Social Security benefits.

A complex actuarial framework wraps around the microsimulation model to provide totals for demographic variables as well as additional information in areas where the microsimulation model has not yet been developed. (For example, CBOLT projects Medicare spending, but that spending is projected in aggregate and not at the individual level). That framework also includes a macroeconomic model that projects aggregate economic variables.

The Social Security Administration (SSA) provides an administrative data set that CBO uses to derive the sample of individuals used in CBOLT. That data set contains basic demographic information such as sex and birth year, as well as earnings histories, for a 1 percent random sample of Social Security numbers,

[1] For a description of the model underlying CBO's projections, see Congressional Budget Office (2009).

beginning in 1951.[2] The data also include information on Social Security benefits such as type of claim (retirement, survivors, or disability), date of claim, the primary insurance amount, the monthly benefit amount, and the reason for disability.

Because the core CBOLT administrative data include only earnings, age, sex, and Social Security benefit information, CBO uses data from other sources to expand the demographic characteristics assigned to each person. Additional demographic and economic data come from other data sets such as the Survey of Income and Program Participation (SIPP), the SIPP matched to Social Security administrative earnings and benefits records, the Health and Retirement Survey, and the Current Population Survey (CPS). Information from those data sources is not merged directly to the individual record in CBOLT. Rather, CBO uses these sources to estimate statistical relationships on the basis of individual characteristics that already exist in CBOLT. Those estimates provide the basis for imputing data that are not available in the administrative data.

CBOLT also incorporates a wide range of aggregate data. The model is calibrated to match the Social Security trustees' projections of the total population by age, sex, and marital status, which incorporates the trustees' projections of mortality and fertility for each year of the projection period. CBO then uses its own projections of immigration to adjust the population. For the first 10 years of the projection, CBOLT uses economic projections and projections of federal outlays and revenues produced by other divisions at CBO. In later years, CBOLT uses CBO's longer term economic projections.[3]

How Earnings Are Projected in CBOLT

Annual earnings for each individual in the CBOLT sample is the sum of four components:

1. The predicted value from an earnings equation estimated using CPS data; those predictions amount to average earnings for that person's age, sex, education, birth cohort, and Social Security beneficiary status.

2. The value of the individual's permanent earnings differential (PED), which we define as the gap between that person's earnings and the predicted value from the earnings equation.[4] That gap acts as a measure of individual-specific effects and attempts to capture individual heterogeneity in the model.

3. A permanent shock to earnings, the effect of which accumulates over time.

[2] The data CBO receives do not contain the Social Security numbers (SSNs), The Social Security Administration scrambles the SSNs before the data are transmitted to CBO.

[3] For a discussion of those longer term projections, see Congressional Budget Office (2012), Ch. 2.

[4] Assignment of permanent differentials differs slightly depending on whether the person has earnings as specified in the CWHS data or is a new person "born" in the CBOLT sample (that is, the person is created in the projection period of the microsimulation model). For those with historical earnings, the permanent earnings differential is the difference between predicted and observed earnings; for those "born" in the simulation, a permanent earnings differential is drawn from a pool of observed differentials.

4. An annual transitory shock, which measures any additional (but temporary) variation in a person's earnings.

Predicted Earnings

For each individual in the model, CBOLT projects labor force participation status (that is, whether the individual is in or out of the labor force). For people in the labor force, CBOLT projects whether they work full- or part-time, their hours worked, unemployment spells, and annual earnings.[5] Those projections depend on probabilities estimated using CPS data and are based on several characteristics, including:

- Age,
- Sex,
- Lifetime educational attainment,
- Marital status,
- Number of children under 6 years of age (for women),
- Educational status (that is, whether the person is currently in school),
- Birth cohort (in 10-year ranges), and
- Social Security benefit status.

We use 35 years of CPS data to allow the identification of both long-term economic trends and cohort-specific behavior.[6] The approach in these earnings and labor force equations is to use potential or full-time equivalent earnings (FTE)—the amount that an individual could earn if fully employed throughout the year. In the model, projected full-time equivalent earnings are adjusted for actual hours worked to solve for annual earnings.[7]

---

[5] An in-depth description of the labor force equations can be found in Congressional Budget Office (2006). In both the CPS data and the CBOLT analysis, we use a measure of full-time-equivalent (FTE) earnings. The FTE earnings approach makes it possible to distinguish several factors that independently affect earnings such as age, sex, educational attainment, and birth cohort. FTE earnings for each person in the CPS data are computed by starting with reported earnings and adjusting for the hours a person actually worked during the year. This transformation abstracts from the effect that hours worked have on earnings. Earnings in the administrative data are similarly transformed into FTE earnings based on actual earnings observed in the data and imputed hours worked, which are derived from models based on CPS data.

[6] The CPS has been modified a number of times over its history. For example, categories of educational attainment were changed in the March 1992 survey, and information about people's status as foreign- or native-born was added in the March 1994 survey. Since 1976, the labor force information included in the CPS, such as hours and weeks worked, has been measured more or less consistently. For more details, see Congressional Budget Office (2008), Appendixes A and B.

[7] See Congressional Budget Office (2006) for more details.

Simply using these estimated equations to assign earnings to individuals in CBOLT would result in a single earnings amount within each group (in the case of the CPS earnings regressions, this means within age, sex, education, 10-year birth cohort, and Social Security benefit receipt groups). Incorporating permanent earnings differentials into the earnings assignment adds variance to earnings within (and between) groups in a single year. Including individual permanent and transitory shocks adds longitudinal variation for each person in the model. Incorporating heterogeneity in both the cross-sectional and the longitudinal dimensions generates a more realistic earnings distribution in the microsimulation model.

Permanent Earnings Differentials

To project earnings in the microsimulation model, we calculate a permanent earnings differential (PED), our measure of individual heterogeneity in earnings. For people in the microsimulation model for whom we observe actual earnings in the administrative data, the PED equals the five-year average of the difference between the person's full-time-equivalent earnings (adjusted for growth in nominal wages) in the years prior to the end of the sample period (here, 2005 to 2009) and predicted full-time-equivalent earnings among people with similar characteristics (specifically, age, sex, lifetime educational attainment, 10-year birth cohort, and Social Security benefit receipt) in those same years—as shown in (1):[8]

$$(1) \qquad PED_i = \frac{\sum_{t=1}^{5}(lnE_{it} - ln\hat{E}_{it})}{5}$$

where $\hat{E}$ denotes the predicted value based on demographic characteristics (age, sex, educational attainment, 10-year birth cohort, and Social Security beneficiary status), *i* indexes individuals, and *t* represents time. The differential between observed earnings and predicted earnings is by definition the sum of permanent and transitory differences in a given year. By construction, the expected value of the transitory shocks is zero (more on this below). Averaging the differential across several years should allow us to recover a better measure of the permanent differential than would looking at a single year.

For those persons who do not have earnings in any of the five years in the averaging period or who have all of their earnings projected (e.g., those who were in the labor force and younger than age 25 at the beginning of the simulation or those who are "born" in the simulation period), the PED is drawn randomly from the set of PEDs calculated for those ages 21 to 31 in the historical period.[9] This approach guarantees that the distribution of the shocks for those individuals matches the distribution observed historically for people in that age range.

Permanent and Transitory Shocks

We use a model first developed by Carroll (1992) to estimate variances of permanent and transitory shocks. This model estimates single values of permanent and transitory earnings variances instead of separate estimates by year as other researchers have done (see, for example, Moffitt and Gottschalk, 2012, and Kopczuk et al., 2010). The Carroll approach is useful in the microsimulation context because it does

[8] We have also used a 10-year period for calculation of the PED, but the resulting earnings projections do not match observed projections nearly as well, and requiring 10 years of data reduces the sample for which we are able to calculate PEDs. For details on the calculation of average earnings in each cell, see Congressional Budget Office (2006).

[9] PEDs for individuals ages 21 to 31 are included to ensure that the pool from which to selected PEDs for other people in the model is sufficiently large.

not require projecting the patterns of permanent and transitory variances over time; that is, we do not have to project an annual series for those variances. We simply use the estimates of the two terms in the microsimulation model, as opposed to forcing the distribution to meet certain targets or patterns over time.

Many previous studies in the existing literature employ error-components models to estimate the year-to-year variance in permanent and transitory earnings. The canonical error-components model can be expressed as:

(2) $y_{it} = \rho_i + \tau_{it}$

where y_{it} is log earnings for individual *i* at time *t*, ρ_i is a time-invariant permanent individual component, and τ_{it} is a transitory component for individual *i* that can change over time *t*. With typical assumptions that the two components are uncorrelated, the cross-sectional variance of log earnings is the sum of the two variance components:

(3) $var(y_{it}) = \sigma_\rho^2 + \sigma_\tau^2$

where σ_ρ^2 is the permanent variance and σ_τ^2 is the transitory variance. A simple way to estimate these terms is to identify the permanent variance (by noting that it equals the covariance of earnings between two different time periods) and then compute the transitory variance by subtracting the permanent variance from the total variance. In extensions to this canonical model, researchers have relaxed many of the typical assumptions and have made other extensions to the model. They might, for example, allow the permanent component to vary with both age and time (e.g., ρ_{iat}), or they might vary the process by which the component varies (e.g., by a first-order autoregressive moving-average process), or they might use different estimation methods (e.g., minimum distance).

Implementing the Carroll methodology is straightforward: The first step is to calculate differences in earnings for all pairs of earnings (for example, the difference between earnings in year *t* and year *t+1*, year *t* and year *t+2*, year *t+1* and year *t+2*, and so on). The variances of those earnings pairs are then calculated; the resulting correlations across those variances determine the permanent and transitory variances. More formally, we define the canonical earnings model by defining E_t as earnings in year *t*, V_t as a multiplicative transitory shock in year *t*, and P_t as permanent earnings in year *t*. Thus, in the case where there is no transitory shock ($V_t = 1$), earnings will equal permanent earnings:

(4) $E_t = P_t V_t$, or, in logarithmic terms, $ln(E_t) = ln(P_t) + ln(V_t)$

Now allow *lnP* to follow a random walk process, *G* to equal the growth factor of permanent earnings, and N_t to equal a multiplicative shock to permanent earnings:

(5) $P_{t+1} = GP_t N_{t+1}$, or, in logarithmic terms, $lnP_{t+1} = lnG + \ln P_t + \ln N_{t+1}$

By moving equation (5) ahead in time, defining $g = G - 1 \approx lnG$, and substituting the previous expression of equation (5), we can ultimately compare *m* years in the future to earnings today*:*

(6) $lnP_{t+2} \approx g + \ln P_{t+1} + \ln N_{t+2}$

$$\approx g + (g + \ln P_t + \ln N_{t+1}) + \ln N_{t+2}$$

$$\approx g + g + \ln P_t + \ln N_{t+1} + \ln N_{t+2}$$

…

$$lnP_{t+m} \approx mg + \ln P_t + \ln N_{t+1} + \ln N_{t+2} + \cdots + \ln N_{t+m}$$

From (4) and substituting (6), we get the following evolution of earnings between two periods *t* and *t*+*m*:

(7) $$lnE_{t+m} - lnE_t = (lnP_{t+m} + lnV_{t+m}) - (lnP_t + lnV_t)$$

$$\approx (mg + lnP_t + \ln N_{t+1} + \ln N_{t+2} + \cdots + \ln N_{t+m} + lnV_{t+m}) - (lnP_t + lnV_t)$$

$$\approx mg + (\ln N_{t+1} + \ln N_{t+2} + \cdots + \ln N_{t+m}) + (lnV_{t+m} - lnV_t)$$

Equation (7) therefore shows that, over time, a change in earnings equals the sum of the shocks to permanent earnings plus the change in transitory shocks.

The variance of the change in earnings follows from equation (7):[10]

(8) $$var(r_m) \equiv var(lnE_{t+m} - lnE_t) = m\sigma_{lnN}^2 + 2\sigma_{lnV}^2$$

Because the difference in earnings (*r*) can be calculated for every worker in every pair of years in the data, for any two different values of *m*, the equation can be solved for the permanent and transitory variances. Specifically, in a linear regression of the vector *v*, defined as $[var(r_1), \dots, var(r_n)]$ on a constant term and the vector $[1, \dots, n]$, the coefficient on the slope will capture the permanent variance $(m\sigma_{lnN}^2)$ and the coefficient on the constant term will capture the transitory variance $(2\sigma_{lnV}^2)$.[11] Because the permanent and transitory shocks are lognormal random variables, we use the standard deviations to scale those shocks when we incorporate the two terms into the microsimulation model.

Putting it All Together

To project earnings for each individual in the CBOLT sample, we pull together predicted earnings based on a CPS earnings equation, the permanent earnings differential (PED), and the permanent and transitory standard deviations:

(9) $$lnE_{it} = ln\left(\hat{E}_{it} + PED_i\right) + \sum_{s=1}^{t} \alpha_{is}\, \sigma_{lnN} + \beta_{it}\sigma_{lnV}$$

The permanent and transitory standard deviations, σ_N and σ_V, are used to scale two factors, α_{is} and β_{it}, which are drawn from a random normal distribution and are assigned to each person in each year; thus,

---

[10] This assumes independence between transitory shocks and permanent shocks in different years, independence between transitory and permanent shocks in any pair of years, and shocks that have means of zero.

[11] Carroll (1992) was probably the first to show that the typical assumption that the covariances between the transitory components at different points in time are zero may not hold in general but may hold with "long" autocovariances.

each person in the model is assigned different values of the permanent and transitory shocks in each time period. The third term in equation (9) shows how the permanent shocks accumulate over time: Each person is assigned an initial PED (call it PED_0), which, in the first period, is adjusted using the permanent shock ($\alpha_1\sigma_N$). In the second period, another permanent shock is added to the PED and permanent shock from the first period, and so on:

Period 0: PED_0

Period 1: $PED_1 = PED_0 + \alpha_1\sigma_N$

Period 2: $PED_2 = PED_1 + \alpha_2\sigma_N = (PED_0 + \alpha_1\sigma_N) + \alpha_2\sigma_N$

Period 3: $PED_3 = PED_2 + \alpha_3\sigma_N = (PED_1 + \alpha_2\sigma_N) + \alpha_3\sigma_N$

$$= (PED_0 + \alpha_1\sigma_N) + \alpha_2\sigma_N + \alpha_3\sigma_N$$

Therefore, in period *T*: $PED_T = PED_0 + \sum_{s=1}^{T} \alpha_s\sigma_N$

Thus, in any period, the *PED* for each person is the initial *PED* plus the accumulation of the permanent shocks—in other words, this approach is equivalent to allowing an individual's *PED* to evolve slowly over time.[12]

Estimating the Shocks

To estimate the permanent and transitory variances, we use administrative earnings data from the Social Security Administration, described briefly in section II. In this section, we describe the data in more detail and present some summary statistics. Next, we present the results of the variance estimation.

Data: The Detailed Earnings Record

The earnings data used in the volatility estimation procedure are drawn from the Detailed Earnings Record (DER) data file, which is part of the Continuous Work History Sample (CWHS) extraction system used by the Social Security Administration. The CWHS is a 1:100 sample of all issued Social Security numbers and contains information on each person's birthdate and sex.[13] The DER used for this analysis

[12] This approach is indirectly derived from Deaton and Paxson (1994), who observed an increasing range for earnings within any given birth cohort as its members age. That observation is consistent with continuous, low-variance shocks to permanent earnings differentials.

[13] The input data are drawn from the CWHS, a group of administrative data sets that include the Summary Earnings Record, Detailed Earnings Record, the Numerical Identification System, and the Master Beneficiary Record. The CWHS refers to a 1 percent sampling framework used by the Social Security Administration to extract data from these sources. For CBOLT, we use a 10 percent sample of the CWHS file, which therefore represents a one-in-one-thousand random sample of issued Social Security numbers. In principle, a random sample of Social Security numbers is representative of the U.S. population, although events such as death and immigration are not fully recorded in the data. To adjust the data for some of those potential issues, we extensively re-weight the CWHS data to make sure that the sample drawn is in fact representative along several stratifying dimensions such as age, sex, work status, and beneficiary status. CBO receives only scrambled Social Security numbers.

contains earnings from 1978 through 2009. Because some of the earnings records in the period 1978 – 1980 appear to have some errors (see, Kopzcuk et al., 2010; and Schwabish, 2011), and because prior to 1984 earnings only for workers covered by the Social Security system were included, we restrict our use of the earnings data to the period 1984 – 2009. Earnings for that period—unlike some earlier years—are not capped at the taxable maximum, which allows us to capture the upper tail of the distribution in our estimates.

Compared with survey data, administrative data have both advantages and disadvantages. Administrative data sets tend to have large samples and higher-quality records because they are not subject to the typical errors or biases found in survey data such as nonresponse error, recall error, imputation bias, or topcoding. However, administrative data tend not to have the same sort of demographic information that is typically available in survey data, and they may lack other information better recorded by surveys (for example, earnings from cash-based employment or acquired "under the table"). In general, survey data tend to include more covariates (for example, educational attainment, race, and marital status) but fewer observations than administrative data, and they may suffer from misreported earnings, imputation bias, and other common shortcomings (Schwabish, 2012).

Another advantage of administrative data sets is that the upper tail of those data is longer and thicker than in most survey data sets; the tail also appears to be thicker at the bottom end. At the top end, the lack of topcoding in the administrative data generates a long tail that extends far beyond what most survey data report. At the bottom end, there are two opposing factors that may offset each other:

- A thicker tail may occur because survey respondents may not reliably report small amounts of employment or earnings that are captured on their W-2s.

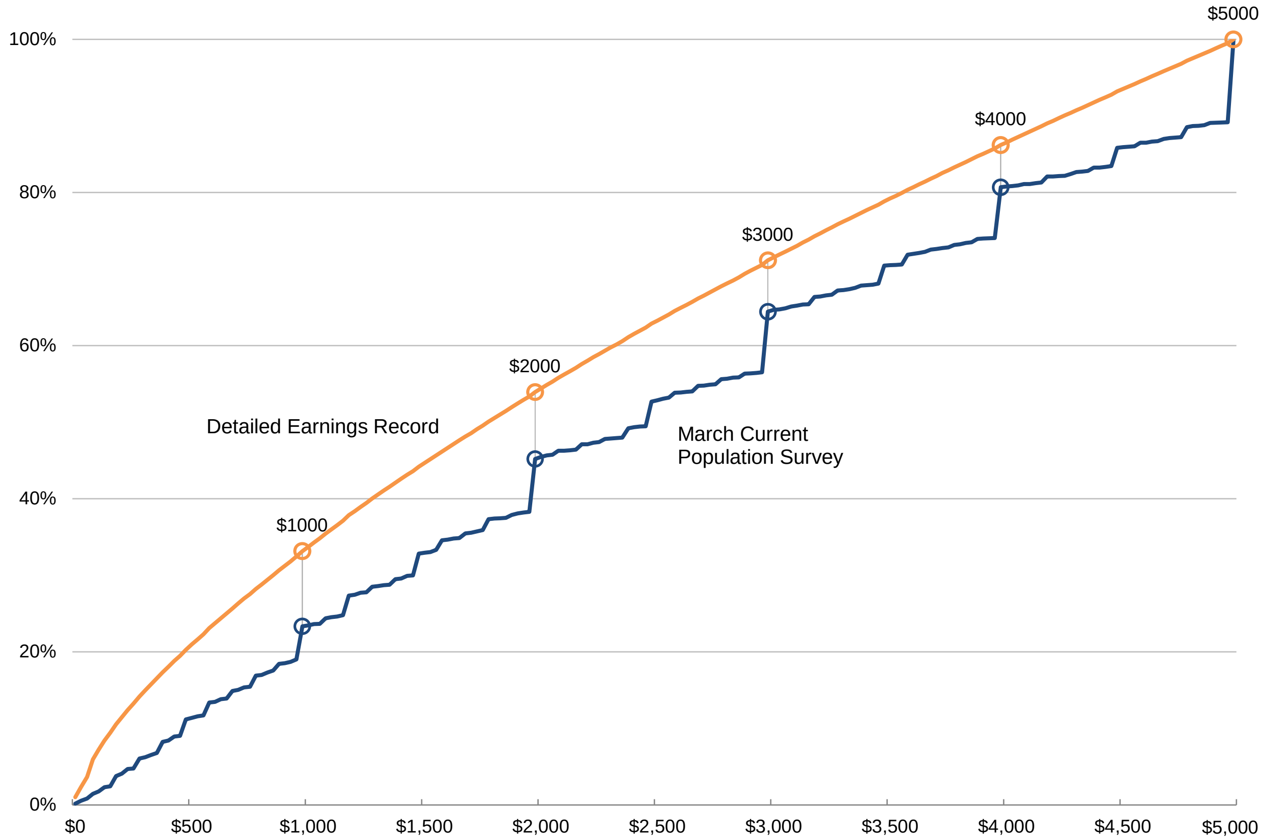
- A thinner tail may occur because people who receive under-the-table earnings may report those earnings to a survey, but those earnings are not captured in the administrative data.

The results shown in Figure 1 suggest the former effect dominates the latter: the figure shows the cumulative distribution of earnings below $5,000 in calendar year 2009 from the CWHS and from the March Current Population Survey (CPS). Two observations from the figure are clear: The share of people at each point in the administrative data (below $5,000) is about 8 to 15 percentage points above the distribution in the survey data, suggesting that cases in which people with low earnings underreport their earnings in a survey (relative to their administrative records) are common, a finding supported in Cristia and Schwabish (2007). The second observation is that CPS survey respondents tend to round reports of their earnings at thousand-dollar increments, indicated by the circled spikes in Figure 1.[14]

Because we are unlikely to produce reliable estimates of full-time equivalent earnings for workers at the extremely low end of the earnings distribution, we include only earnings above a specified minimum amount in the analysis. (The concern about very low earnings is even more of an issue in administrative data because, according to the shares in Figure 1, there appear to be more observations with very low earnings than in survey data.) That minimum is the amount required to earn four quarters of coverage under the Social Security system. (The amount of earnings to earn one quarter of coverage was $1,090 in 2009; an annual amount of $4,360 is used here as the threshold amount of earnings. We have also used

[14] For a discussion see Schwabish (2007, 2012).

Figure 1.
Cumulative Distribution of Earnings for People with Positive Earnings Below $5,000 in the DER and CPS, 2009



the inflation-adjusted value of the minimum wage—at 1,500 hours of work per year—and found some sensitivity to that threshold.)[15] Estimates in previous research have also used minimum earnings cutoffs (e.g., Sabelhaus and Song, 2010) or have trimmed the sample, usually the top and bottom 1 percent and/or 2 percent (e.g., Congressional Budget Office, 2007; Moffitt and Gottschalk, 2012). Earnings are also adjusted for growth in nominal wages.

In CBOLT, we estimate permanent and transitory variances for three age groups (25 to 34, 35 to 44, and 45 to 60) for all workers and for men and women separately.[16] For individuals younger than 25 or older than 60, the variance for the nearest age group is applied. Carroll (1992) "detrended" household income by computing the ratio of each household's income to average household income, and then conditioned

[15] Specifically, both the permanent and transitory variance estimates are smaller resulting in earnings that are lower across most of the earnings distribution. Overall, using the minimum wage threshold in the calculations results in a projected earnings distribution that matches the observed distribution much less well.

[16] We have also estimated the model by position in the earnings distribution and by groups of volatility (measured, for example, as the variance in earnings over some period) but have not yet incorporated all of those estimates into the CBOLT model.

that ratio on the household head's age, occupation, and educational attainment. Because the administrative data have very limited demographic information, however, we are limited in the demographic variation we can exploit.

Including all observations, our estimation sample consists of nearly 115,000 observations, with slightly more men than women (Table 1). About 22 percent of our sample is ages 25 to 34, about 26 percent is ages 35 to 44, and the remainder is ages 45 to 60. Younger workers are excluded because they do not have sufficient years of work to be included in the estimation. For each person, we calculate the difference in earnings between year *t* and year *t+m*, with a maximum value for *m* of 10 years. The worker must have positive earnings in both years (and, in the case where we apply a minimum earnings floor, must have earnings above that floor in both years) and must fall within the broad age bounds for both of the years in the calculation. Thus, a person who has continuously positive earnings for all years between 1984 and 2009 will have 16 ten-year differences, 17 nine-year differences, and so on all the way to 25 one-year differences.

Table 1.
Sample Counts

	Age Group			
	25 to 34	**35 to 44**	**45 to 60**	**Total, 25 to 60**
		Number		
Men	13,259	15,650	31,316	60,225
Women	12,189	13,738	28,771	54,698
Total	25,448	29,388	60,087	114,923
		Percent of Total		
Men	11.5	13.6	27.2	52.4
Women	10.6	12.0	25.0	47.6
Total	22.1	25.6	52.3	100.0

Earnings volatility measures are calculated only for individuals with positive earnings in both pairs of years of the estimation period, resulting in a sample of more than 88,000 observations. For that sample, 13 percent are in the 25–34 age group, about a third are in the 35–44 age group, and the remainder are in the oldest age group.[17] The seemingly unbalanced nature of the age groups reflects the number of workers with positive earnings (people at younger ages are more likely to have a loose connection to the labor force) and the longitudinal patterns in earnings over the entire time period (people must have positive earnings in both pairs of years to be included in the estimation). We follow Carroll and restrict the regression estimation to begin with differences of at least three years; that is, we ignore differences of one and two years.

[17] Because we require that a person have five years of earnings and be at least 25 years old, there is a significantly smaller share of people in the youngest age group.

Estimation: Permanent and Transitory Shocks

Estimates from the administrative data show that variances increase with the number of years between earnings observations.[18] Estimated earnings variances for gaps of different lengths are shown in the three panels of Figure 2 by age group and gender. The first panel shows the variances for all workers and for

the three age groups, the second panel shows the estimates for men, and the third panel shows the estimates for women. The overall pattern is roughly the same across the three: Men exhibit slightly higher variances at all points (gaps) than do women, older men have higher variances at all points than younger men, and younger women have higher variances than older women.

We take the year-gap variances for gaps of three through ten years from Figure 2 and use the Carroll regression approach to back out single values of the permanent and transitory variances for each of the two sex and three age groups. For example, for 35–44 year old men, we take the variances from gaps 3 through 10 (see Figure 3). The best-fit (regression) line through that series generates a slope of 0.024, interpreted as the permanent variance, the estimated intercept is 0.193, half of which (0.097) is interpreted as the transitory variance (see equation 8).

Point estimates for the permanent and transitory variances for the different sex and age groups are shown in the panels of Figure 4. The estimates all show that the permanent variances are much smaller than the transitory variances. For the entire sample, for example, the permanent variance is estimated to be 0.026 and the transitory variance is estimated to be 0.087, or about three times as large. (Recall, however, that in the application of the shocks in the microsimulation model, the effect of the permanent shocks accumulates over time so that when implemented the relative sizes of the estimated shocks in the model differ from those shown in the figure.) The permanent variances vary less than the transitory variances across the three age groups and are generally centered around 0.025. Older male workers have the highest permanent variance; there is less variation across age groups for women. The transitory variances are significantly lower for workers in the oldest age group—for men ages 45 to 60, the transitory variance is estimated to be 0.093, about 87 percent as much as the estimated transitory variance for younger male workers ages 25 to 34. Estimated transitory variances for women tend to be lower for each age group than for their male counterparts, and the differences between the age groups are slightly larger—older women workers have an estimated transitory variance of 0.057, or about 60 percent of the estimated transitory variance for younger women.

The permanent and transitory variances estimated from the SSA administrative data differ from those estimated by Carroll (1992). In his original work, his estimate of the permanent variance (using total family income from the Panel Study of Income Dynamics, PSID, from 1976 to 1985) was 0.016, smaller than our estimate of 0.026. His estimate of the transitory variance was 0.027, about one-third the size of our estimate of 0.087. Carroll's estimates suggest that the permanent variance accounts for about 37 percent of the total variance and the transitory variance accounts for about 63 percent of the total, as opposed to 23 percent and 77 percent in our model. Those differences are probably because of differences in data set and time period covered, as well as the fact that Carroll was using total family income, which is less variable than individual (full-time equivalent) earnings (see Dahl et al., 2011).

[18] To reiterate, we use gaps of up to 10 years in earnings and ignore gaps of one and two years to estimate the permanent and transitory variance.

Figure 2.
Variances of Full-Time Equivalent Earnings Growth Rates, by Sex and Age Group

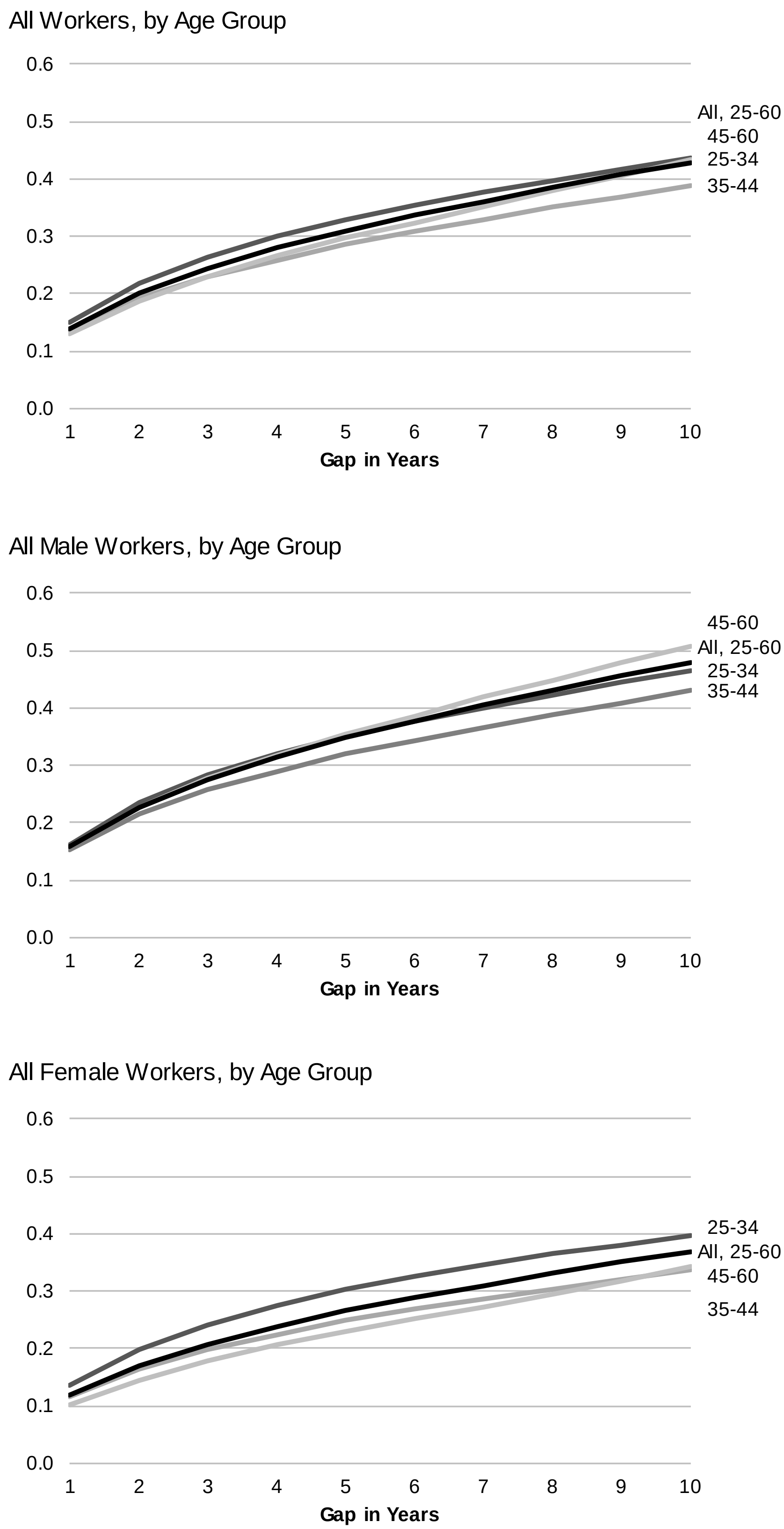
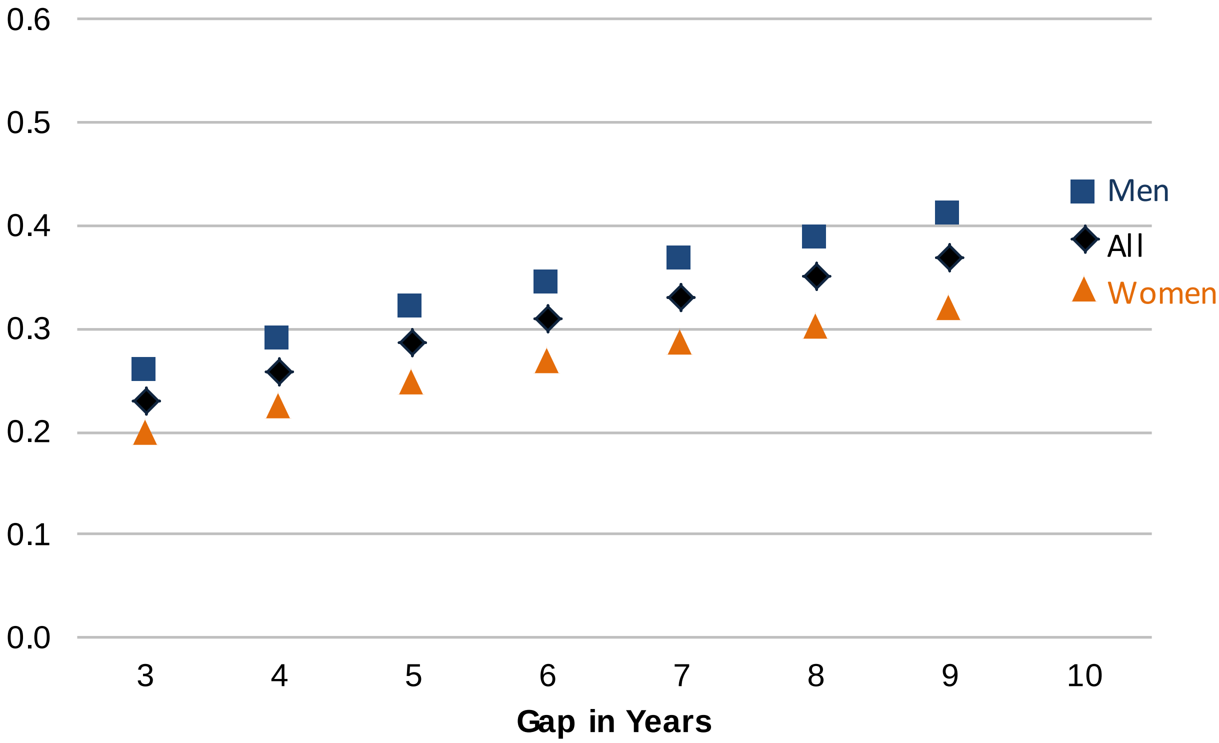


Figure 3.
Applying the Carroll (1992) Model to Variances of Earnings Growth Rates for Workers Ages 35 to 44



Several more recent papers in this literature report permanent and transitory variances that are more similar to our estimates. Sabelhaus and Song (2010), who use data similar to those used here, estimate a transitory variance of about 0.119 and a permanent variance of about 0.027 for men's earnings between 1980 and 2005. Their estimates suggest that the permanent component accounts for about 18 percent of the total variance and the transitory variance accounts for about 82 percent of the total. Moffitt and Gottschalk (2012) use the PSID to estimate an error components model that lets them track trends in earnings shocks over time. They find a permanent variance (among men ages 30 to 59, using the PSID) that averages about 0.016 between about 1980 and 2004 (or about 40 percent of the total variance) and a slightly higher transitory variance that averages about 0.024 over that same period (or about 60 percent of the total variance).[19] DeBacker et al. (2011) use administrative tax return data from the Internal Revenue Service to estimate earnings processes among men and find much larger estimates of both the permanent and transitory components (0.25 and 0.63), but the shares of the total variance are on par with our estimates and suggest that the permanent component accounts for about 28 percent of the total and the transitory component accounts for about 72 percent of the total. Differences between the estimates we present here and other estimates can be attributed to differences in data, available demographic information, and estimation procedure.

[19] Moffitt and Gottschalk's estimates rise between 1970 and 2004; these averages were estimated using a visual inspection of their Figure 3. For purposes of brevity, other relevant papers in this literature are not reviewed here.

Figure 4.
Estimated Permanent and Transitory Variances, by Sex and Age Group

Results from the Microsimlation Model

This section describes the earnings projections that result from putting together all of the moving parts and including them in the CBOLT microsimulation model.[20] The test of any earnings process is how well the resulting earnings distribution matches observed distributions. The combination of the four components (the predicted value, the permanent earnings differential, the permanent, and the transitory shocks) generates earnings for the CBOLT sample that are consistent with cross-sectional (across individuals at a given time) observations in the earnings distribution. The longitudinal earnings patterns (for the same individual over time) generated in the model are not quite as accurate in matching observed patterns.

The earnings patterns in CBOLT that result from implementing the earnings shock process largely continue the historical experience of rising earnings inequality (measured, for example, as changes in selected percentiles; see the discussion below), but inequality more or less stops rising by the end of the projection period shown here (2035). In this section, we show selected percentiles (10th, 25th, 50th, 75th, and 90th) and an entropy measure of inequality (the Gini coefficient) for all workers, for men, and for women over the entire period 1984 to 2035. For the point in time where the model moves from actual earnings in the CWHS data (2009) to projected earnings (2010), we show the percentage change at each percentile for all workers, for men, and for women. This “jump-off” demonstrates the difficulty of capturing movements in earnings over the entire distribution such that the distribution moves in a predictably smooth way. Finally, to compare longitudinal earnings patterns in the historical and projection periods, we calculate earnings transition matrices for all workers, for men, and for women.

Cross-Sectional Earnings Distribution, 1984 to 2035

The panels of Figure 5 show levels of the 10th, 25th, 50th, 75th, and 90th percentiles of full-time-equivalent earnings for all workers, for men, and for women. Earnings are adjusted for average nominal earnings growth before and after 1993. (Growth in average nominal earnings is accounted for separately in CBOLT, and is not discussed here to focus on the distribution of earnings.) Figure 5 also shows the jump-off from history to projection between 2009 and 2010.

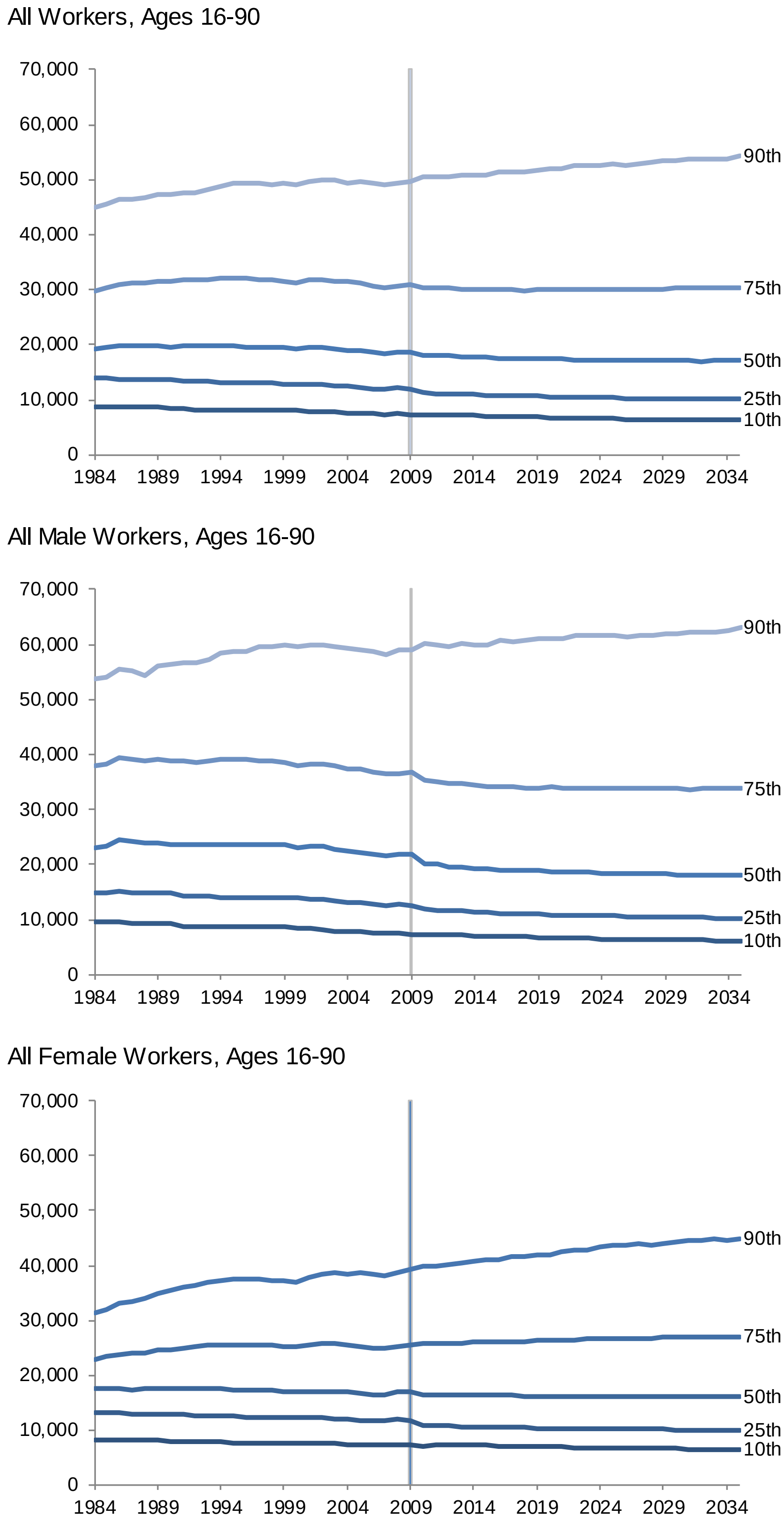
All Earnings. Full-time-equivalent (FTE) earnings (above a threshold equal to four times the earnings amount required to earn a quarter of coverage, or $4,360 in 2009) have increased since 1984 at the 90th percentile; earnings have been flat at the 75th percentile and have fallen slightly at the 10th, 25th, and 50th percentiles. For workers at the 10th percentile, earnings were about $8,800 in 1984; by 2009, the last year of actual earnings from the CWHS, earnings at the 10th percentile had fallen to about $7,200 and are projected to decline to about $6,300 by 2035.

Toward the top of the distribution, earnings at the 90th percentile increased over the past 25 years, rising from about $45,000 in 1984 to almost $50,000 in 2009. By 2035, the projections suggest that earnings at the 90th percentile will reach about $54,000.

Men’s Earnings. Earnings among male workers at the bottom of the distribution declined between 1984 and 2009 from about $10,000 to about $7,000. Between 2010 and 2035, earnings at the 10th percentile of the men’s earnings distribution are projected to decline from about $7,300 to about $6,200.

[20] All estimates presented in this section are from the model used in Congressional Budget Office (2012).

Figure 5.
Earnings Percentiles for All Workers Ages 16–90, Men, and Women, 1984–2035
(Full-time equivalent earnings in 1993 dollars)



At the median, FTE earnings for men have fallen over the past few decades and are projected to continue to fall. Between 1984 and 2009, earnings at the middle of the men's distribution declined slightly from about $23,000 to $22,000. That decline is projected to continue over the following 25 years and to reach about $18,000 by 2035.

In contrast to earnings in the rest of the distribution, which fell more or less steadily from 1984 through 2009, earnings at the top of the men's distribution increased. Between 1984 and about 2009, earnings at the 90th percentile rose from about $54,000 to $59,000. CBOLT projections show continued increases in earnings at the top of the distribution over the longer term. By 2035, earnings at the 90th percentile are projected to reach about $63,000.

Women's Earnings. Women's earnings are typically lower than men's earnings, and patterns of women's earnings growth—both historically and in the projection period—differ from those for men. Earnings at the bottom of the distribution were lower than those for men in 1984 but are projected to be slightly higher by 2035. At the 10th percentile, for example, earnings declined from about $8,000 in 1984 to $7,200 in 2009; men's earnings at the 10th percentile declined by a larger amount. In the projection period, earnings for women at the 10th percentile are projected to decline from just over $7,000 in 2010 to $6,400 in 2035; men's earnings are projected to decline to $6,200.

In the middle of the distribution, women's earnings are projected to be fairly stable (compared with men's earnings) in both history and the projection period. In the historical period, earnings at the 50th percentile of the women's earnings distribution were essentially unchanged around $17,000 between 1984 and 2009; by 2035, median earnings are projected to decline only slightly and to reach about $16,000.

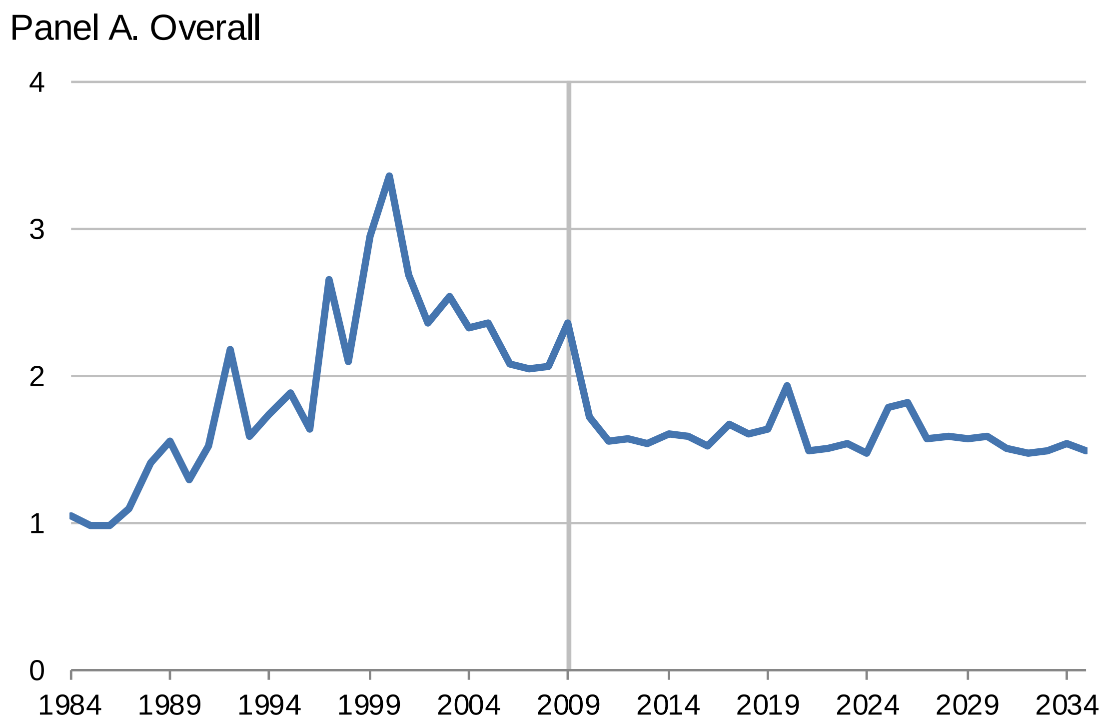
At the top of the earnings distribution, earnings for women have grown more steadily and more quickly than for men. Between 1984 and 2009, earnings at the 90th percentile of the women's earnings distribution rose from $31,000 to about $39,000. Through 2035, earnings at the 90th percentile are projected to continue to increase, and to reach $45,000 by 2035.

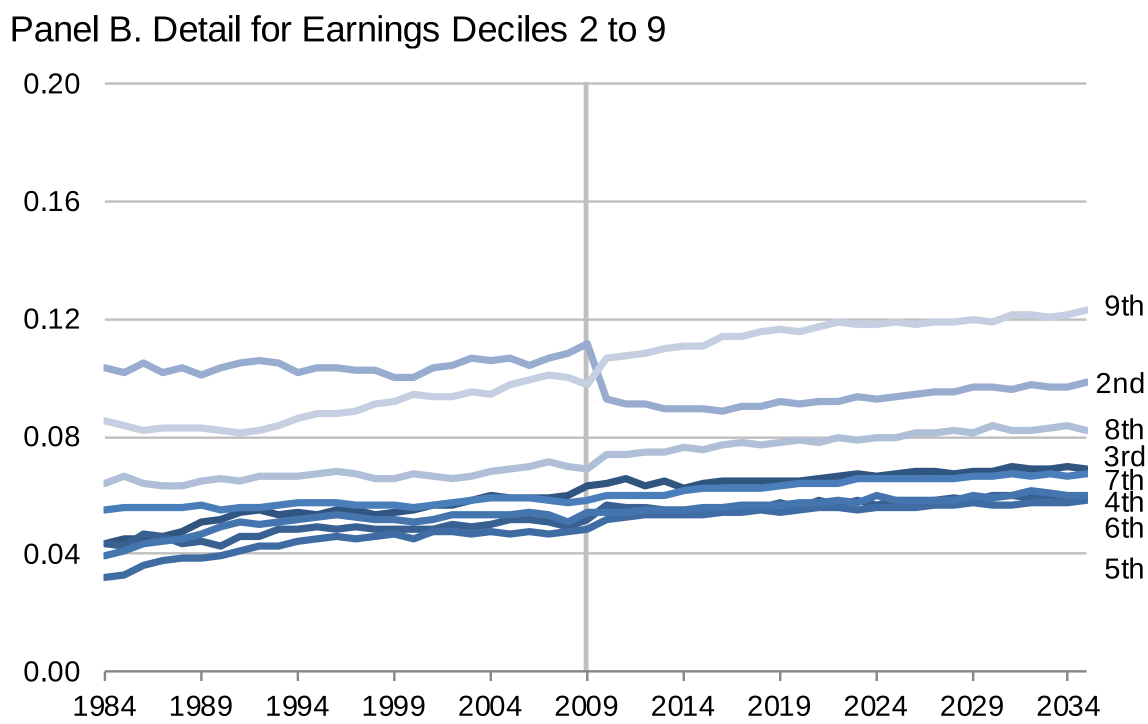
Ratio of Men's to Women's Earnings. Over the past few decades, the ratio of men's earnings to women's earnings has declined because of a combination of factors, including the increasing labor force participation of women, the impact of immigration on labor markets, and technological change favoring skilled rather than unskilled workers that has affected workers with different levels of educational attainment. The decline in the ratio is projected to slow over the next few decades, in part because we assume that future cohorts of men and women have demographic characteristics similar to those observed in the data for workers born in the 1970s rather than assuming continued increases in the labor force participation of women.

Cross-Sectional Variance of Earnings, 1984 to 2035

To assess earnings projections in the microsimulation model, we evaluate the distribution of earnings within each year and compare the projected results to the historical experience. The top panel of Figure 6 shows the coefficient of variation (equal to the standard deviation divided by the mean) for all earnings above the minimum threshold between 1984 and 2035. Overall variability increases from 1984 through about 2000 and then begins a period of decline that continues through the early part of the projection period. Variability then flattens around 2012. The average coefficient of variation of 1.6 for the projection period is only slightly below the average of 1.8 from 1984 through 2000.

Figure 6.
Coefficient of Variation for All Workers with Earnings, Ages 16 to 90, 1984 to 2035





The estimates in the bottom panel of Figure 6 show this metric for the 2nd through 9th deciles of the earnings distribution.[21] In the historical period, the coefficient of variation is flat or rises slightly for each decile pictured. Coefficients of variation in the middle deciles (3rd through 8th) rise the fastest between 1984 and 2009 and are clustered in the bottom portion of Panel B ranging from about 0.05 to about 0.07 in 2009. Over the next 25 years, the variation grows more slowly than in the previous 25 years and ends up about 50 percent higher than in 1984. Toward the ends of the distribution—the 2nd and 9th deciles—the coefficient of variation is higher than in the middle of the distribution; in 2009, the estimate is about 0.1 for both the 2nd and 9th deciles. The second decile demonstrates the largest change at the transition from the historical period to the projection period. The coefficient of variation for the 2nd decile drops sharply between 2009 and 2010, reflecting the difficulty in projecting earnings well in all parts of the earnings distribution. Once past the transition, the rate of growth between 2010 and 2035 is about the same as it was during the historical period (1984 to 2009).[22]

Under alternative experiments in which we modified the estimated permanent and transitory shocks, trends in the coefficient of variation behave as expected. Under simulations in which the permanent variance is doubled, variation rises substantially in the projection period, owing to the accumulation of the permanent shocks. Under simulations in which only the transitory variance is modified (either doubled or halved), the pattern in variation is similar to that shown in the top panel of Figure 6. Finally, under simulations in which the permanent variance is halved, variation declines slightly over the projection period.

Summary Measures of Inequality

Changes in full-time equivalent earnings among both men and women over the past 25 years have resulted in an increase in earnings inequality in the population. Earnings projected using the methods described in this paper continue that trend (as measured by the Gini coefficient and by the share of earnings under the Social Security taxable maximum) through about 2035, but earnings inequality generally ceases to rise thereafter. That result is consistent with the view that earnings inequality is unlikely to rise forever and with the model assumption that many characteristics of future cohorts match those of the 1970s cohort.

Goldin and Katz (2007) find that a majority of the increase in inequality since 1980 can be attributed to rising educational wage differentials. Over the next few decades, they argue, those differentials will probably continue to grow, although potentially at a slower pace as the labor market adjusts to a workforce that has a different level and distribution of education than in the past. Other factors—skill-biased technological change and computer-based technologies, for example—will also play a role in rising inequality as such factors have tended to increase earnings of those at the top of the distribution and push out workers (with lower earnings) at the bottom of the distribution. Other factors may also increase inequality: Changes in social norms (e.g., changes in intrahousehold work arrangements), changes in

[21] The coefficient of variation is calculated overall and then separately for each decile. The coefficient of variation for the second decile, for example, is calculated as the standard deviation of earnings within that decile divided by mean earnings for that decile.

[22] Because earnings in the administrative data are not topcoded, the variance of earnings in the top decile are about ten times the level of the other deciles and are therefore not shown in the figure. The variance of earnings in the bottom decile is about three times as large as the other deciles, most likely because of the extremely low earnings in those deciles.

labor market institutions (e.g., decline of unionization, especially as the baby-boom generation retires and is replaced by non-unionized workers), and globalization all have the potential to result in higher levels of inequality. In addition, demographic shifts in the labor market—namely, the aging and retiring of the baby-boom generation—might result in a higher, but constant, level of inequality in the future as younger workers (for whom the rise in educational attainment is not as rapid as it was for previous cohorts) take their place. However, given the historical pattern of long periods of rising and falling income inequality, some changes in technology, education, workplace management and structure, or social norms will probably result in a slowing or cessation of the growth of inequality in the future.

In CBOLT, patterns of earnings inequality are driven by a combination of factors, such as demographics, timing of the transition from the historical data to the projection period (i.e., the number of years of historical earnings data available in the model), and the use of historical data in the earnings projection methodology and in estimation of the shocks. One particularly important factor influencing the result that inequality ceases to rise by the mid-2030s is the transition over time to having a larger fraction of the population "born" in the simulation and drawing permanent earnings differentials from the same distribution—as that distribution is specified to not change over time. Also, the persistence of the permanent shock heavily influences the shape of CBOLT's pattern of inequality in the projection period. The estimate of that persistence is based on recent historical data and generates rising inequality at a gradually decreasing rate—consistent with the pattern expected based on the economic forces described above. Summary measures of inequality—such as the Gini coefficient and the share of earnings below the Social Security taxable maximum—reflect those various factors in the long term.

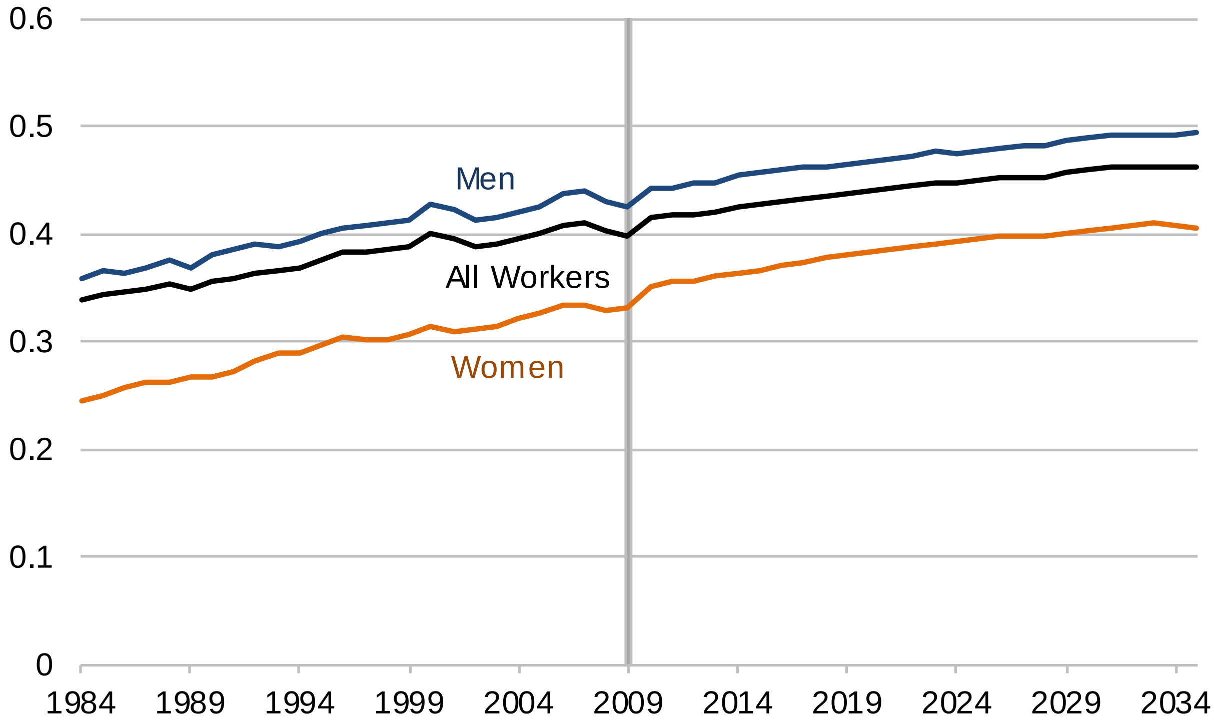
Between 1984 and 2000, the Gini coefficient is estimated to have risen from 0.34 to 0.40, an increase of 18 percent (see Figure 7).[23] Over the next decade, inequality appears to have remained roughly unchanged, and, by 2009, the Gini coefficient was at the same level as in 2000. Through 2035, CBOLT projects that inequality will continue to rise, but at a slower rate. Between 2009 and 2035, the Gini coefficient for all workers is projected to rise by 17 percent, from 0.40 in 2009 to 0.46 in 2035.

Patterns of inequality for men and women are similar, and although the level of inequality for women was about 30 percent lower than it was for men in 1984, inequality is estimated to rise more quickly for women than for men. Between 2009 and 2035, the Gini coefficient for women is projected to increase by 23 percent; it is projected to increase by about 16 percent for men. The ratio of the Gini coefficient for men to the Gini coefficient for women is estimated to have fallen only slightly from 1.5 to 1.3 between 1984 and 2009 and is projected to fall slowly to 1.2 by 2035.

The percentage of earnings under the Social Security taxable maximum is also projected to stabilize by about 2035 (no figure is shown). In 1984, about 88 percent of all covered earnings was below the taxable maximum; that share had fallen to 82 percent by 2007. This decline occurred as earnings at the top of the income distribution increased very quickly during the 1990s. The ratio of taxable payroll to covered earnings reached almost 85 percent in 2009, rising as a result of the recession (Social Security

[23] The Gini coefficient is a measure of concentration (inequality) that ranks the distribution of the variable of interest (in this case, earnings) against a line of perfect equality. That line assumes that each element in the distribution has the same contribution to the total; for example, each person in a given earnings distribution has the same earnings. The Gini coefficient measures deviations from this line and ranges between 0 (no concentration or perfect equality) and 1 (total concentration and perfect inequality).

Figure 7.
Gini Coefficient for All Workers, for Men, and for Women, Ages 16 to 90, 1984 to 2035



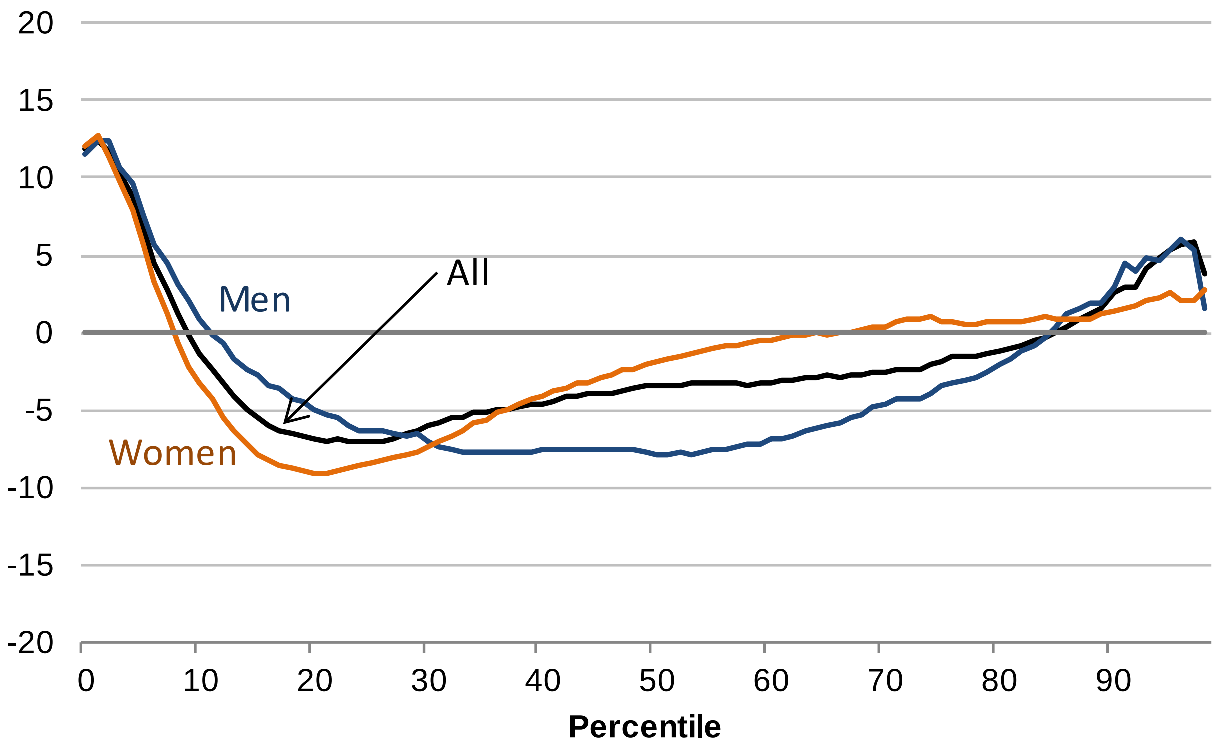
Administration, 2012). CBO projects the ratio will stabilize at about 83 of covered earnings after the economy returns to full employment.

The Earnings Jump-Off: Change in Earnings between 2009 and 2010

A primary challenge of predicting earnings in a microsimulation model is to generate a seamless transition of the earnings distribution from the final year with actual earnings to the first year with projected earnings. The results over the entire projection period (see above) appear to show a relatively smooth transition from the recent historical period, with levels of earnings and inequality evolving smoothly to a new steady state of slightly higher inequality than in the past. However, because of the scale of the previous figures, the distribution is masked when the microsimulation model moves from the final year of actual earnings records (2009) to the first year in which earnings are projected (2010). In reality, that change—which we call the "jump-off"—is larger than other observed year-to-year changes and also differs across the distribution. This is a common problem in microsimulation modeling and reflects the difficulty of accurately projecting both the level and dynamics of the earnings process.

The jump-off is shown for all workers, for men, and for women in Figure 8. We show the percentage change in earnings between 2009 and 2010 for each percentile for those earning more than the minimum earnings floor. For men, the jump-off resembles a U-shaped pattern, with earnings at the tails of the distribution increasing between 2009 and 2010. Earnings for most of the distribution decline between the two years, with declines of about 8 percent between about the 30th and 50th percentiles. Starting at the

Figure 8.
Percent Change in Full-Time Equivalent Earnings Between 2009 and 2010 (the "Jump-Off") for Workers, Ages 16 to 90, by Percentile
(Percent change)



median, the jump-off slopes upwards and reaches about 5 percent at the top of the distribution.[24] For women, the jump-off is larger than the jump-off for men for about the first third of the distribution; the jump-off is closer to zero for women than for men for most of the remainder of the distribution. At the top of the earnings distribution for women, the jump-off never exceeds 3 percent. Overall, the jump-off is -5 percent to 5 percent for about 40 percent of the men's earnings distribution and for about 70 percent of the women's earnings distribution.

Longitudinal Earnings Mobility, 2005 to 2006 and 2024 to 2025

In this section, we calculate transition matrices for two pairs of years. We look at two years in the historical period (2005 and 2006) to capture historical earnings data, and two representative years in the projection period (2024 and 2025).[25] Although incorporating the earnings variance estimation

---

[24] For purposes of projecting Social Security finances, single year changes at the top of the distribution are less important than those at the bottom of the distribution because individuals only pay Social Security taxes up to the taxable maximum. In 2009, the taxable maximum was $106,800, which was higher than the 94th percentile of all workers (90th for men and 96th for women).

[25] There are slight differences in these transitions over time, but the 2005-2006 transition is fairly representative. For example, between 2020 and 2021 about 31 percent of workers did not change in which decile their earnings fell; between 2034 and 2035 that share had risen to 32 percent.

methodology appears to capture the overall earnings cross-sectional distribution well, and the variance of earnings for most of the distribution matches historical experience, the model does less well at capturing individual workers' (longitudinal) transitions from one part of the distribution to another from year to year. In particular, workers in the projection period are much more likely to move from one part of the distribution to another than in the historical period.

The two panels of Figure 9 are contour maps of earnings transition tables calculated by decile of the earnings distribution for all workers.[26] (Each cell in the graph is color-coded to denote the share of workers in that cell; darker colors indicate higher probabilities.) What is apparent from the Panel A, which shows the transitions between 2005 and 2006, is that the percentage of workers along the length of diagonal (i.e., people who do not change deciles between the two years) in the transition matrix is higher than between 2024 and 2025, which is shown in Panel B. For example, the share of workers who stay in the first decile in 2005 and 2006 is 3 percent, the share who stay in the fifth decile in both years is 5 percent, and the share who stay in the tenth decile is 9 percent. Between 2024 and 2025, however, the share of workers who stay in the first decile is 6 percent, the share of workers who stay in the fifth decile is 2 percent, and the share who stay in the tenth decile is 7 percent.

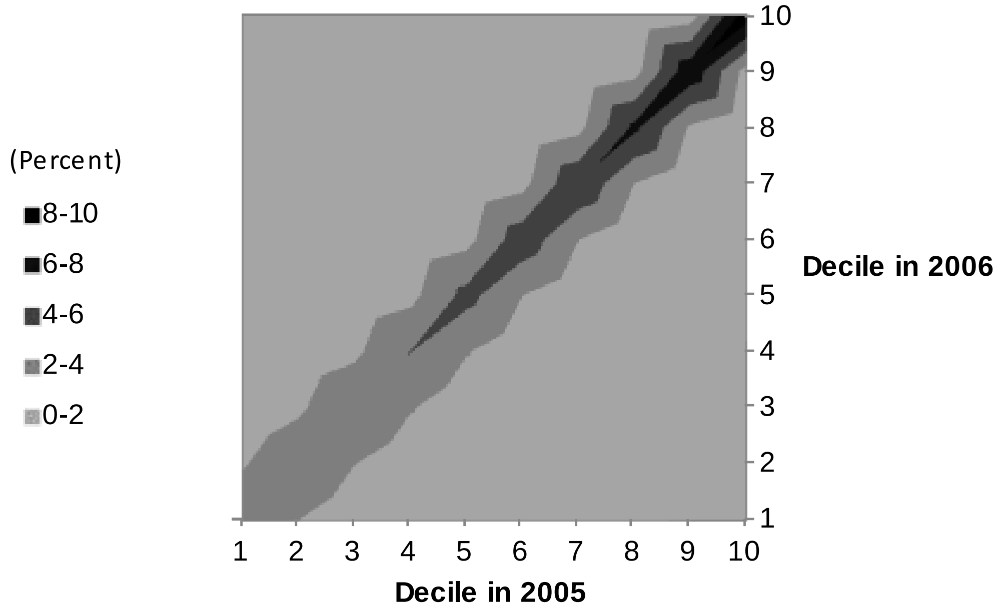
Although on an annual basis more people move across deciles in the projection period than in history, a five-year comparison of average earnings deciles to the current earnings decile yields a pattern in the projection that is more similar to history.[27] (See Figure 10.) That is, although people might shift deciles more often in the projection, those people are about as likely to end up in their average earnings decile in the projection as in history. This is consistent with overall earnings volatility (as measured by the coefficient of variation) not increasing over the projection period.

[26] We measure longitudinal earnings mobility by calculating transition matrices for consecutive pairs of years. A transition matrix is simply a matrix of n rows by n columns, where n is the number of categories (for example, deciles). The intersection of each row and column contains the percentage of people who are in decile x in year t and in the same decile x in some other year, $t+i$. Thus, the diagonal from the lower left to the upper right of the matrix represents the percentages of workers who do not move from one decile to another between years t and year $t+i$; the upper left triangle (the area above the diagonal) represents the percentages of workers who move up from one decile to another; and the lower right triangle (the area below the diagonal) represents the percentages of workers who move down from one decile to another.

[27] For this comparison, we calculate workers' earnings deciles for average earnings over the five years preceding and including the first year (for example, 2002 to 2006). We then compare the decile for five-year average earnings to the first year of that five-year period.

Figure 9.
One-Year Transition Matrices, 2005 to 2006 and 2024 to 2025
(Percent of total)

Panel A. Transition Matrix between 2005 and 2006



Panel B. Transition Matrix between 2024 and 2025

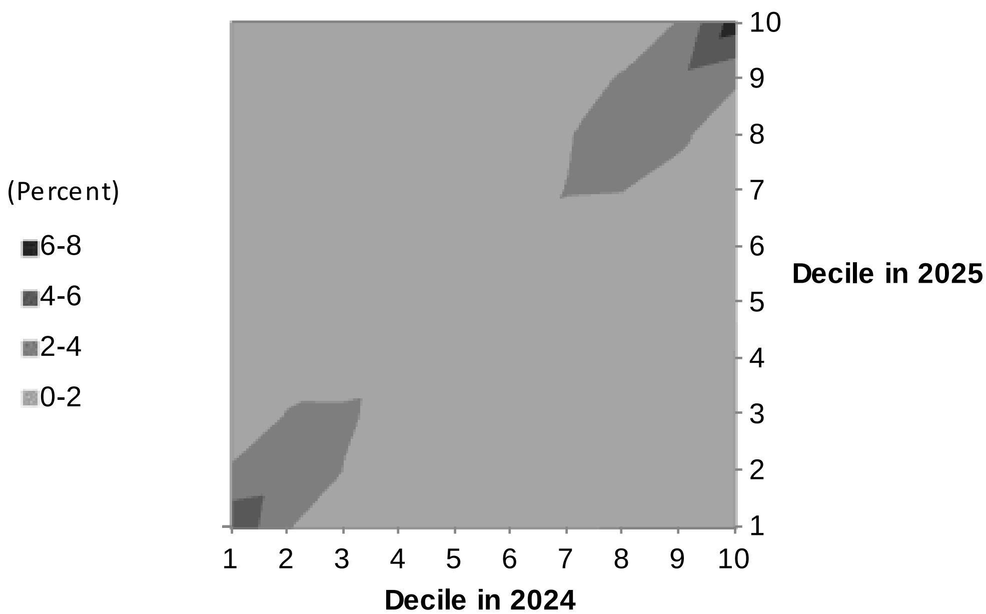
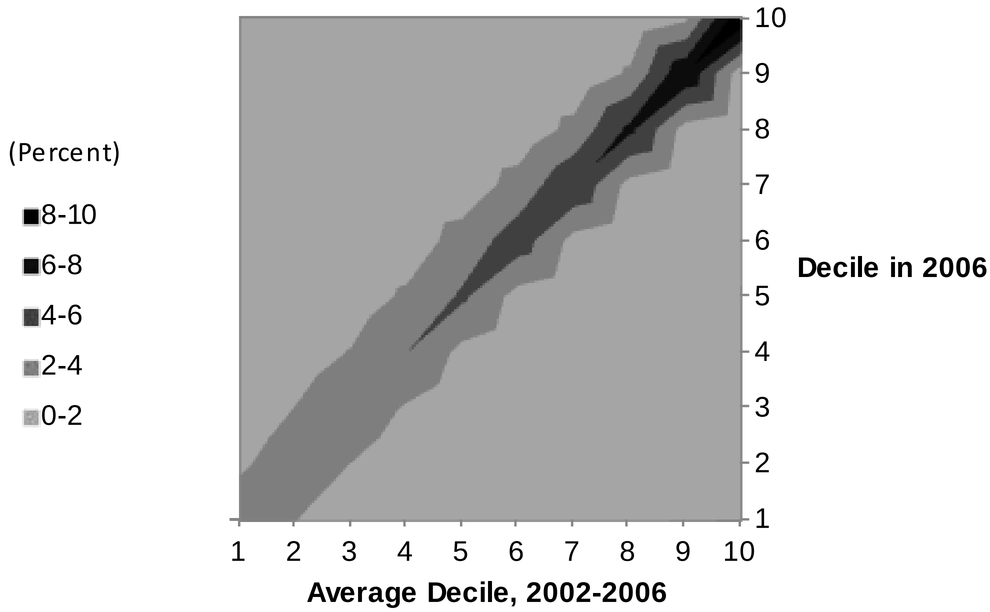
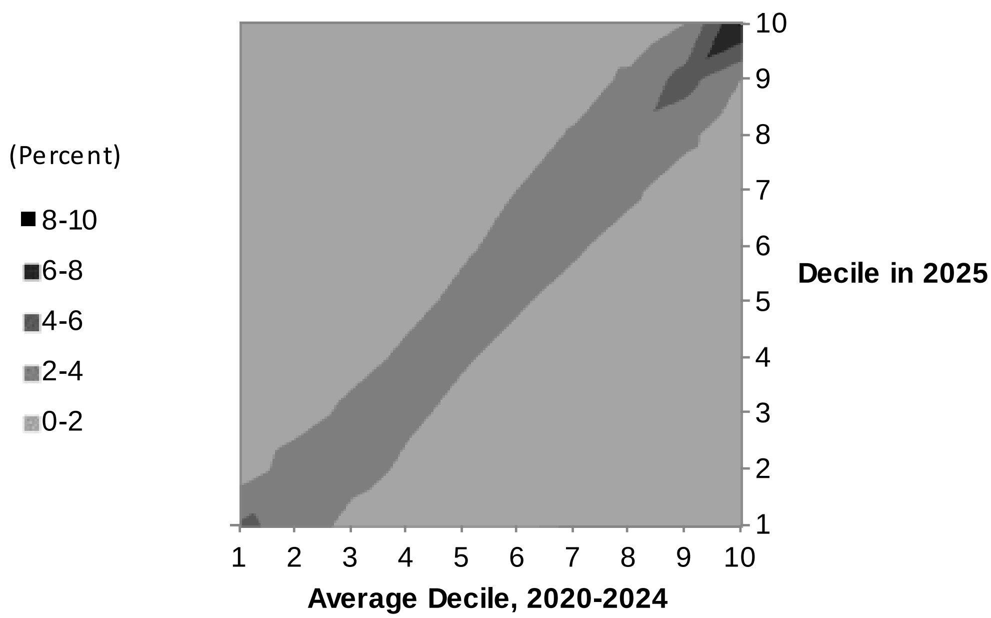


Figure 10.
Five-Year Transition Matrices, 2002 to 2006 versus 2006, 2020 to 2024 versus 2025
(Percent of total)

Panel A. Transition Matrix between 2002-2006 and 2006



Panel C. Transition matrix between 2020-2024 and 2025



Appendix.
How Other Microsimulation Models Project Earnings

Other models have different approaches to projecting earnings in microsimulation models. This section briefly describes how earnings are projected in two models developed by researchers at the Urban Institute—the Dynamic Simulation of Income Model (DYNASIM) and Modeling Income in the Near Term (MINT). Both DYNASIM and MINT are based on Survey of Income and Program Participation (SIPP) data, but the SIPP records used in the MINT model have been matched to administrative data from the Social Security Administration (SSA) on earnings, benefit receipt, and date of death. Both models, like the Congressional Budget Office Long-Term (CBOLT) microsimulation model, have been used extensively to analyze the distributional effects of Social Security.

Like CBOLT, DYNASIM uses permanent and transitory shocks in projecting earnings. DYNASIM uses a minimum distance function to statistically match individuals to similar observations from survey respondents in the Panel Study of Income Dynamics (PSID) and the National Longitudinal Survey of Youth (NLSY). The matching process is based on gender and three-year age bands and is conditioned on a set of observable characteristics, such as earnings, wealth, pension income, race, and ethnicity. To project future earnings, DYNASIM predicts hourly wages and annual hours worked, and uses these values to calculate annual earnings. The error terms from those wage and hours equations are assumed to have a permanent component and a transitory component. The permanent component is assigned from donor observations in the PSID. The transitory component is drawn each year from the PSID donor observations and incorporates a one-year lag (see Favreault and Smith, 2004; Smith, 2012).

MINT forecasts earnings differently from CBOLT. Rather than estimating a model to project earnings, MINT uses observed earnings patterns for workers in older cohorts to predict the earnings of workers in younger cohorts, a process known as "earnings splicing." By using statistical techniques to splice several years of the earnings record of an older worker to the earnings record of a younger worker, the model is able to duplicate the statistical properties of observed earnings patterns. This technique is repeated in five-year intervals to build up lifetime work histories for individuals who have not completed their careers by the last year of available data (see Toder et al., 2002). For individuals ages 55 and over who are not retired, MINT uses a fixed-effects approach to project earnings. In this fixed-effects approach, data from the SIPP are matched with administrative data to estimate earnings.[1]

[1] See Smith et al., 2010

References

Carroll, Christopher D. 1992. "The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence." Brookings Papers on Economic Activity, no. 2: pp. 61 – 156.

Congressional Budget Office. 2006. Projecting Labor Force Participation and Earnings in CBO's Long-Term Microsimulation Model. (October). Washington, D.C.

Congressional Budget Office. 2007. Trends in Earnings Variability over the Past 20 Years. (April). Washington, D.C.

Congressional Budget Office. 2008. Assigning Education Status in CBO's Long-Term Microsimulation Model. (October). Washington, D.C.

Congressional Budget Office. 2009. CBO's Long-Term Model: An Overview. (June). Washington, D.C.

Congressional Budget Office. 2012. The 2012 Long-Term Budget Outlook. (June). Washington, D.C.

Cristia, Julian and Jonathan A. Schwabish. 2007. "Measurement Error in the SIPP: Evidence from Administrative Matched Records." Journal of Economic and Social Measurement, vol. 34, no. 1: pp. 1 – 17.

Dahl, Molly, Thomas DeLeire, and Jonathan A. Schwabish. 2011. "Estimates of Year-to-Year Volatility in Earnings and in Household Incomes from Administrative, Survey, and Matched Data." Journal of Human Resources, vol. 46, no. 4: pp. 750 – 774.

Deaton, Angus and Christina Paxson. 1994. "Intertemporal Choice and Inequality." Journal of Political Economy, vol. 102, no. 3, pp. 437 – 467.

DeBacker, Jason, Bradley Heim, Vasia Panousi, and Ivan Vidangos. 2011. "Rising Inequality: Transitory or Permanent? New Evidence from a Panel of U.S. Tax Returns 1987 – 2006." Research Paper Series, No. 2011-01-01, Indiana School of Public and Environmental Affairs. Bloomington, Indiana.

Favreault, Melissa and Karen Smith. 2004. "A Primer on the Dynamic Simulation of Income Model (DYNASIM3)." The Urban Institute, (February).

Goldin, Claudia & Lawrence F. Katz, 2007. "Long-Run Changes in the Wage Structure: Narrowing, Widening, Polarizing." Brookings Papers on Economic Activity, no. 2, pp. 135 – 168.

Kopczuk, Wojciech, Emmanuel Saez, and Jae Song. 2010. "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data Since 1937." Quarterly Journal of Economics, vol. 125, no. 1: pp. 91 – 128.

Moffitt, Robert A. and Peter Gottschalk. 2012. "Trends in the Transitory Variance of Male Earnings." Journal of Human Resources, vol. 47, no 1: pp. 204 – 236.

Sabelhaus, John, and Jae Song. 2010. "The Great Moderation in Micro Labor Earnings," Journal of Monetary Economics, vol. 57, no. 4: pp. 391 – 403.

Schwabish, Jonathan A. 2007. "Take a Penny, leave a Penny: The Propensity to Round Earnings in Survey Data," Journal of Economic and Social Measurement, vol. 32: pp. 93 – 111.

Schwabish, Jonathan A. 2011. "Identifying Rates of Emigration in the United States Using Administrative Earnings Records," International Journal of Population Research, vol. 2011, article 546201.

Schwabish, Jonathan A. 2012. "Subtle Sources of Bias and Error." In Bad Data Handbook, Q. Ethan McCallum, ed. O'Reilly Media (November), pp. 129 – 142.

Smith, Karen E. 2012. "Projection Methods Used in the Dynamic Simulation of Income Model (DYNASIM3)." The Urban Institute (February).

Smith, Karen E., Melissa M. Favreault, Barbara A. Butrica, and Philip Issa. 2010. "Modeling Income in the Near Term Version 6." The Urban Institute (December).

Social Security Administration. 2012. The 2012 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds. (April). Washington, D.C.

Toder, Eric, Lawrence Thompson, Melissa Favreault, Richard Johnson, Kevin Perese, Caroline Ratcliffe, Karen Smith, Cori Uccello, Timothy Waidmann, Jillian Berk, and Romina Woldemariam. 2002. "Modeling Income in the Near Term: Revised Projections of Retirement Income through 2020 for the 1931–1960 Birth Cohorts." Final Report, SSA Contract No. 600-96-27332, The Urban Institute. Washington, D.C.