

# What Explains Lifetime Earnings Differences across Individuals?

*By José Mustre-del-Río and Emily Pollard*

**E**xpected lifetime earnings are a key factor in many individual and institutional decisions. For example, deciding whether or not to go to college, what kind of occupation to pursue, and when and what kind of house to buy forces individuals to consider not just their earnings today but also their earnings expected over their lifetimes. Additionally, lifetime earnings play a key role in the design of government policies such as Social Security. Thus, given the importance of lifetime earnings for many individual decisions and government policies, understanding the factors that help explain differences in lifetime earnings across individuals is critical.

However, uncovering these factors can be challenging. Some researchers have found that observable characteristics such as age, race, and sex explain only a small portion of the measured differences in wages across individuals at a point in time. Characteristics such as an individual's innate ability or work performance likely play a role, but these characteristics are hard to quantify. Furthermore, the key determinants of lifetime earnings may be different from the determinants of wages at a given point in time.

---

*José Mustre-del-Río is a senior economist at the Federal Reserve Bank of Kansas City. Emily Pollard is an assistant economist at the bank. The analysis in this paper was first performed using the SIPP Synthetic Beta (SSB) on the Synthetic Data Server housed at Cornell University, which is funded by NSF Grant #SES-1042181. Final results were obtained from a validation analysis conducted by Census Bureau staff using the SIPP Completed Gold Standard Files and the programs written by these authors and originally run on the SSB. The validation analysis does not imply endorsement by the Census Bureau of any methods, results, opinions, or views presented in this paper. These data are public use and may be accessed by researchers outside secure Census facilities. For more information, visit <http://www.census.gov/programs-surveys/sipp/methodology/sipp-synthetic-beta-data-product.html>. This article is on the bank's website at [www.KansasCityFed.org](http://www.KansasCityFed.org)*

In this article, we examine what factors help explain lifetime earnings differences across individuals using a novel data set that combines administrative data on earnings with survey data on demographics. Specifically, we use this unique data set to assess how much variation in lifetime earnings across individuals can be explained by observable characteristics. Our results suggest observable characteristics such as sex, race, age, education, and labor market experience explain a little more than half of the lifetime earnings differences we observe across individuals. However, among these characteristics, labor market experience—that is, the fact that some individuals systematically work more years than others—accounts for roughly 40 percent of the difference in earnings. In contrast, standard demographic characteristics such as sex, race, or education alone explain no more than 15 percent of differences in lifetime earnings. Thus, cumulative labor market experience appears to be crucial in explaining lifetime earnings differences across individuals.

Section I describes the data and how lifetime earnings are measured. Section II presents simple statistical and graphical evidence showing how lifetime earnings vary by sex, race, education, marital status, parental status, and labor market experience. Section III shows that in total, these characteristics explain at most 55 percent of lifetime earnings across individuals.

## **I. Defining and Measuring Lifetime Earnings**

Measuring lifetime earnings—as well as identifying what explains differences in lifetime earnings—can be challenging. First, measuring lifetime earnings requires data on entire lifetimes and cannot be proxied by earnings at a point in time (that is, cross-sectional earnings). For example, medical doctors may temporarily have low earnings while in residency but will likely see their earnings rise thereafter. Similarly, individuals raising young children may temporarily work fewer hours but may eventually work more as their children age. Second, examining which individual-level factors help explain earnings differences requires detailed demographic data in addition to comprehensive earnings data.

Because of these two requirements, the Survey of Income and Program Participation Synthetic Beta (SSB) data are ideal for our analysis. These data take respondents from the Survey of Income and Program Participation (SIPP) and match them to their Social Security

(SSA)/Internal Revenue Service (IRS) form W-2 earnings records. These earnings records allow us to construct entire earnings histories for a large sample of individuals. Additionally, because these data are based on a sample of individuals surveyed in the SIPP, they include a host of demographic characteristics (such as race, education, and marital and parental status) that are typically not available in administrative data such as SSA/IRS earnings records.

Although researchers have previously used SSA earnings records to measure lifetime earnings differences, few have related these differences to observable characteristics. For example, using SSA data, Guvenen and others (2018) measure how the distribution of lifetime earnings has changed in the United States since 1957. They find that, over the past six decades, new cohorts of men have seen their median lifetime earnings fall, while new cohorts of women have seen their median lifetime earnings rise. By combining survey and administrative data, we are able to measure lifetime earnings along additional demographic dimensions that have not yet been explored.

### *Sample selection and variable definitions*

To ensure we have a good picture of earnings over a person's entire career, we include only individuals who were age 18–25 in 1978, the start of our sample. We then follow these individuals through 2011, the last year for which data are available, when they were age 51–58. We exclude all individuals who died while in sample, meaning that periods without earnings are due to lack of employment not death. Finally, we restrict our sample to people with a high degree of labor market attachment. In particular, we include only people with at least 17 years of positive earnings—in other words, those who worked for pay during at least half of our 34 year sample.

The key variable of interest, lifetime earnings, is derived from annual earnings data in the SSB. To account for inflation, annual earnings from 1978 to 2011 are converted into real terms using the Consumer Price Index (CPI) and renormalized so that 2018 is the base year. In other words, all dollar amounts presented in this article are directly comparable to wages and prices in 2018. The resulting earnings are then summed up at the individual level to create our final lifetime earnings measure.

To select observable characteristics that might be relevant to lifetime earnings across individuals, we follow the work of Mincer (1958, 1974), which has been the cornerstone of empirical labor economics. Mincer's work suggests earnings at a point in time are related to years of schooling and years of experience in the labor market. Numerous other studies following Mincer's work also suggest we should examine factors such as sex, race, marital status, and parental status.

Thus, the observable characteristics we consider are sex, race, marital status, parental status, educational attainment, and labor market experience (measured as an individual's total number of years with positive earnings). One potential challenge with this broad set of characteristics is that in our sample, these variables are collected at a single point in time. Although sex and race do not change over the lifetime of the individuals in our sample, marital status, educational status, and parental status do. People interviewed early in life are likely to give different answers to these questions than they would have if they had been interviewed later in life. To account for this, we collect the age at which each person was first interviewed in the SIPP. Additional details of our variable construction appear in the appendix.

### *Summary statistics*

Table 1 presents some basic summary statistics for our final sample and shows that it is fairly representative of the entire U.S. population. Rows 1 and 2 show that our sample is nearly equally split between men and women. Rows 3 through 5 show that the majority of individuals in our sample identify as white. Rows 6 and 7 show that the majority of individuals are or have been married, and rows 8 and 9 show that nearly 72 percent of individuals in our sample have had at least one child.

Rows 10 through 15 summarize the educational distribution of our sample. Roughly 33 percent of individuals have a high school degree, while 26 percent have a bachelor's degree or more. In addition to educational attainment, our data allow us to identify majors among college graduates. While the SIPP provides 20 possible major categories, we group college graduates into two categories for simplicity: science, technology, engineering, or mathematics (STEM) majors, and non-STEM majors.<sup>1</sup> Rows 14 and 15 show that one-third of our sample of college graduates obtained a degree in a STEM field. Finally, rows 16 and

*Table 1*  
**Summary Statistics**

Statistic	Value
1. Percent male	50.31
2. Percent female	49.69
3. Percent white	83.85
4. Percent black	11.52
5. Percent other race	4.72
6. Percent ever married	85.09
7. Percent never married	15.06
8. Percent ever had children	71.59
9. Percent never had children	28.42
10. Percent less than high school degree	8.04
11. Percent high school degree	32.92
12. Percent some college	32.92
13. Percent college degree	26.09
14. Percent college degree in STEM field	32.38
15. Percent college degree in non-STEM field	66.67
16. Mean years of positive earnings	30
17. Median years of positive earnings	32

Note: Totals within categories may not sum to 100 due to rounding.  
 Sources: U.S. Census Bureau and authors' calculations.

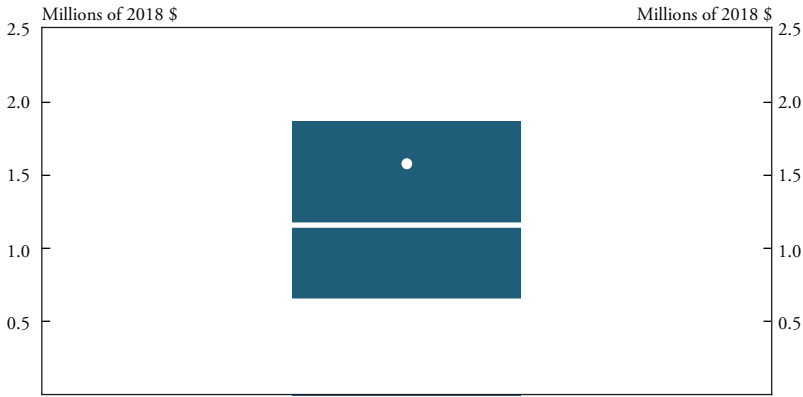
17 show that individuals in our sample are very attached to the labor market. On average, individuals have positive earnings in 30 of the 34 years covered, while the median number of years with positive earnings is higher at 32. That the median is above the mean is not surprising, because the modal or most common outcome is for individuals to have positive earnings in all 34 years covered.

## II. Lifetime Earnings by Observable Characteristics

To gain some initial insight into differences in lifetime earnings, we first examine the overall distribution for our sample. Chart 1 plots this distribution and shows large differences in lifetime earnings between those at the top versus the bottom. The bottom and top of the blue box represent the 25th and 75th percentile of lifetime earnings, respectively. The white dot represents mean lifetime earnings, while the white line represents median lifetime earnings. The dot shows that average lifetime earnings in our sample are slightly over \$1.5 million. However, average lifetime earnings mask important differences across the

Chart 1

## Distribution of Lifetime Earnings Overall



Notes: The bottom and top of the box represent the 25th and 75th percentile of lifetime earnings, respectively. The white line inside the box represents median lifetime earnings, while the white dot represents mean lifetime earnings.  
Sources: U.S. Census Bureau and authors' calculations.

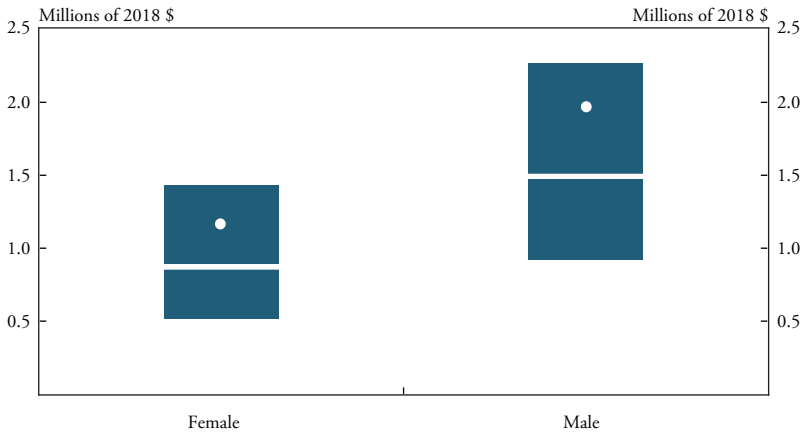
distribution. For example, the top and bottom of the box suggest nearly a three-fold difference between the top and bottom quartiles of the lifetime earnings distribution. Additionally, the median (line) is below the mean (dot), suggesting average lifetime earnings are heavily influenced by outliers at the top of the distribution.

### *The distribution of lifetime earnings by sex and race*

Although Chart 1 shows large differences in lifetime earnings across individuals, it does not show whether these differences are correlated with sex and race. The box plots in Chart 2 show that lifetime earnings differ substantially by sex. For example, comparing the two solid lines in Chart 2 reveals that median lifetime earnings are 70 percent larger for men than for women in our sample. Similarly, comparing the two dots in Chart 2 shows that average lifetime earnings are also about 70 percent larger for men than women in our sample. Interestingly, the largest differences by sex (in percentage terms) occur at the bottom of the lifetime earnings distribution, while the smallest differences occur at the top. For example, men at the bottom quartile of the male-specific earnings distribution earn roughly 80 percent more than women at the bottom quartile of the female-specific earnings distribution. In contrast, men at the top quartile of the male-specific earnings distribution earn

Chart 2

## Distribution of Lifetime Earnings by Sex



Notes: The bottom and top of each box represent the 25th and 75th percentile of lifetime earnings, respectively. The white line inside each box represents median lifetime earnings, while the white dot represents mean lifetime earnings. Sources: U.S. Census Bureau and authors' calculations.

roughly 60 percent more than women at the top quartile of the female-specific earnings distribution.

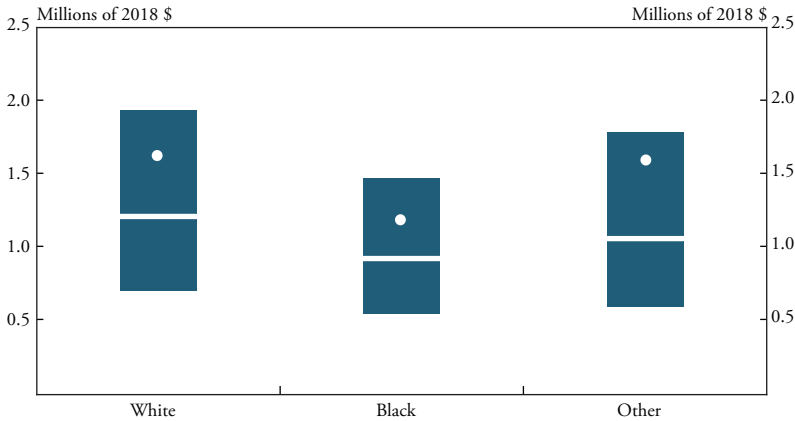
Chart 3 breaks down average lifetime earnings by race and shows systematic differences in lifetime earnings across races. The solid lines in Chart 3 show median earnings for individuals who identify as “white” are higher than median earnings for individuals in the other two race categories the SIPP provides. Specifically, the median earnings of individuals who identify as “white” are 32 percent and 14 percent larger, respectively, than the median earnings of individuals who identify as “black” or “other.” Comparing means or the other quartiles leads to similar results. Indeed, a very robust finding is that the “white” versus “black” lifetime earnings gap is always larger than the “white” versus “other” lifetime earnings gap.

*The distribution of lifetime earnings by marital status, parental status, and education*

Although sex and race do not vary over time in our sample, other observable characteristics are time-variant and may have different associations with lifetime earnings. Importantly, because marital and parental status may have different labor market consequences for men and women, we break down those results by sex.

Chart 3

## Distribution of Lifetime Earnings by Race



Notes: The bottom and top of each box represent the 25th and 75th percentile of lifetime earnings, respectively. The white line inside each box represents median lifetime earnings, while the white dot represents mean lifetime earnings. Sources: U.S. Census Bureau and authors' calculations.

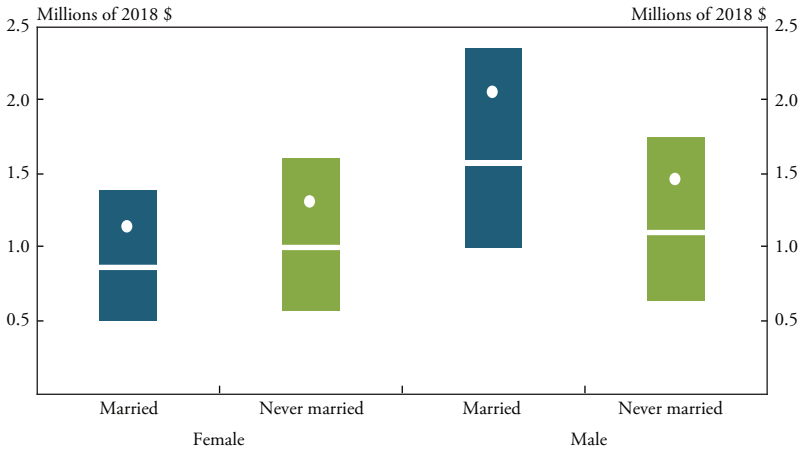
Indeed, Chart 4 shows that the relationship between lifetime earnings and marital status differs substantially across men and women. The first two boxes represent the distribution of lifetime earnings for women who have been married or who have never married, respectively, while the last two boxes represent the distribution of lifetime earnings for men who have been married or who have never married. The first two boxes suggest small differences in lifetime earnings between married and never-married women. Specifically, the white lines show that median earnings for married women are 14 percent less than median earnings for never-married women. In contrast, the last two boxes suggest much larger differences between married and never-married men. Whether comparing means (dots) or medians (lines), married men earn roughly 40 percent more than never-married men.

Chart 5 shows that the relationship between parental status and lifetime earnings also differs measurably between men and women. The first two boxes represent the distributions of lifetime earnings for women who have had children or not, while the last two boxes represent the distributions of earnings for men who have had children or not. Comparing these boxes reveals that having children is associated with lower lifetime earnings for women but higher lifetime earnings for men. For example, the median earnings of women without children are roughly



Chart 4

## Distribution of Lifetime Earnings by Sex and Marital Status



Notes: The bottom and top of each box represent the 25th and 75th percentile of lifetime earnings, respectively. The white line inside each box represents median lifetime earnings, while the white dot represents mean lifetime earnings. Sources: U.S. Census Bureau and authors' calculations.

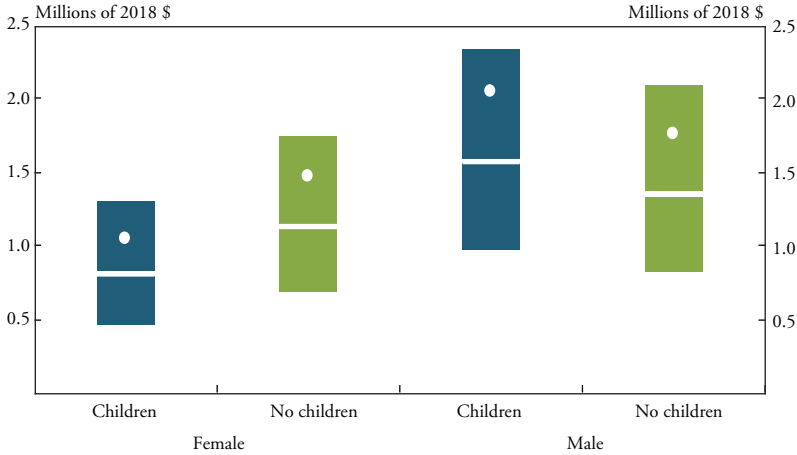
40 percent higher than the median earnings of women with children. In contrast, the median earnings of men without children are 13 percent *lower* than the median earnings of men with children.

Chart 6 shows that higher educational attainment tends to be associated with higher lifetime earnings. For example, comparing the first and last boxes in Chart 6 shows that individuals with a college degree (or more) have higher lifetime earnings than individuals with less than a high school degree regardless of their position within the distribution. More specifically, individuals in the 25th percentile of the college-educated distribution earn more than individuals in the 75th percentile of the less than high school distribution. Comparing the last two boxes in Chart 6 shows that college graduates also typically have higher lifetime earnings than individuals with some college. For example, median lifetime earnings for college graduates are roughly equivalent to earnings at the 75th percentile of lifetime earnings for the some-college distribution. In other words, roughly half of individuals with a college degree have higher lifetime earnings than the most highly compensated individuals in the “some college” distribution.

To assess whether STEM degrees command a premium in the labor market, Chart 7 decomposes the lifetime earnings distribution of

Chart 5

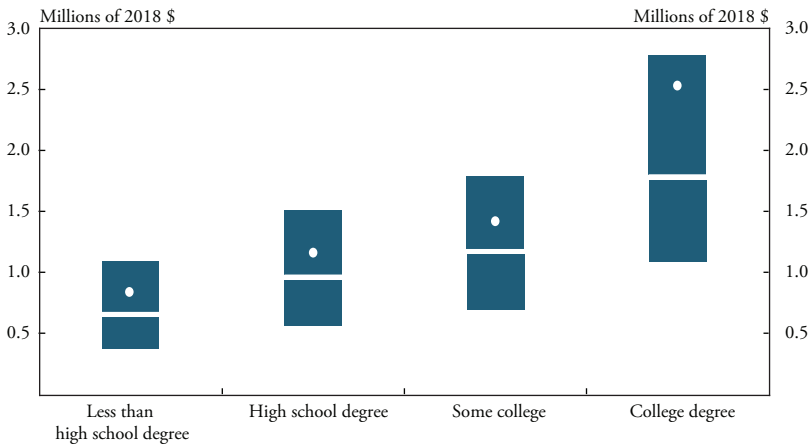
Distribution of Lifetime Earnings by Sex and Parental Status



Notes: The bottom and top of each box represent the 25th and 75th percentile of lifetime earnings, respectively. The white line inside each box represents median lifetime earnings, while the white dot represents mean lifetime earnings. Sources: U.S. Census Bureau and authors' calculations.

Chart 6

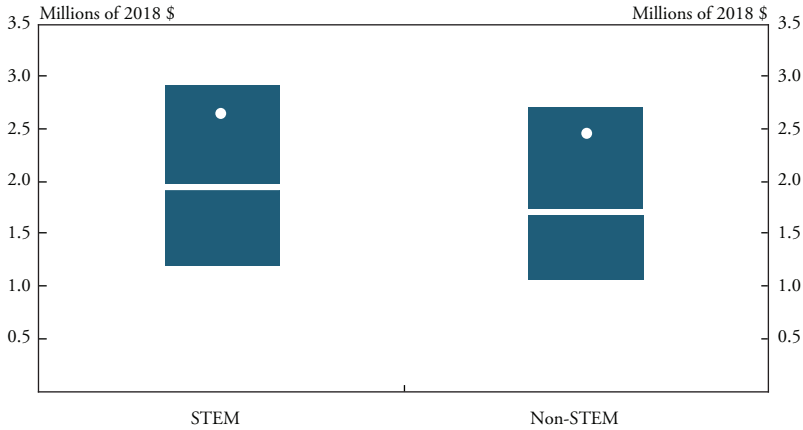
Distribution of Lifetime Earnings by Education



Notes: The bottom and top of each box represent the 25th and 75th percentile of lifetime earnings, respectively. The white line inside each box represents median lifetime earnings, while the white dot represents mean lifetime earnings. Sources: U.S. Census Bureau and authors' calculations.

Chart 7

Distribution of Lifetime Earnings by Field of College Major



Notes: The bottom and top of each box represent the 25th and 75th percentile of lifetime earnings, respectively. The white line inside each box represents median lifetime earnings, while the white dot represents mean lifetime earnings. Sources: U.S. Census Bureau and authors' calculations.

college graduates by whether or not the individual majored in a STEM field. By any metric—means, medians, 25th percentiles, or 75th percentiles—STEM graduates tend to earn about 10 percent more than non-STEM graduates over their lifetimes.

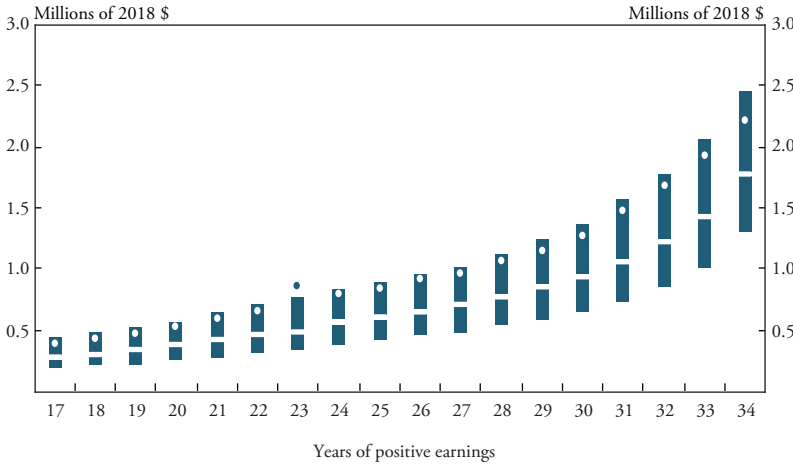
*The distribution of lifetime earnings by total years with positive earnings*

To directly measure labor market experience in line with Mincer’s work, we also examine the distribution by years with positive earnings. Chart 8 shows that individuals who tend to work more over their lifetimes also tend to earn more overall. Indeed, looking at the means, medians, 25th percentiles, and 75th percentiles reveals a systematically increasing relationship between years with positive earnings and total lifetime earnings. For example, the median earnings of individuals in the last bar (34 years with positive earnings) are roughly 6.3 times larger than the median earnings of individuals in the first bar (17 years with positive earnings).

A natural question is whether the differences in Chart 8 are due to working more years or earning more per year. To answer this question, Chart 9 plots the distribution of average earnings *per year* of positive earnings. The chart clearly shows that individuals who work more years also tend to earn more during each year worked. Indeed, comparing the

Chart 8

## Distribution of Lifetime Earnings by Years with Positive Earnings

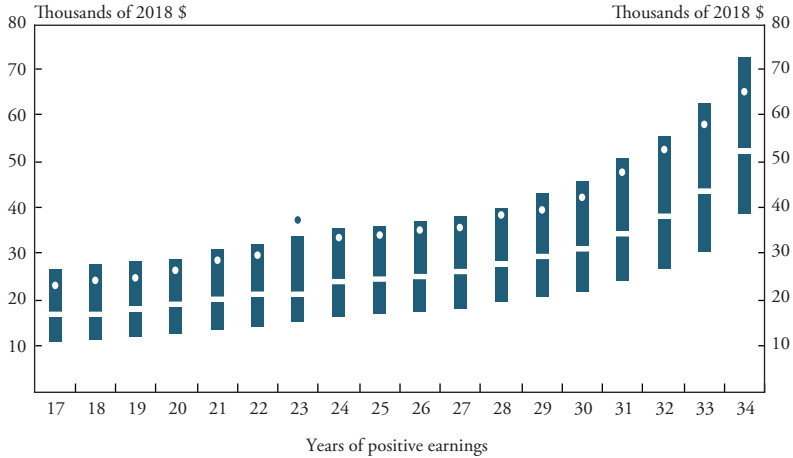


Notes: The bottom and top of each box represent the 25th and 75th percentile of lifetime earnings, respectively. The white line inside each box represents median lifetime earnings, while the dot represents mean lifetime earnings.  
Sources: U.S. Census Bureau and authors' calculations.

median earnings (white lines) in the last and first bars of Chart 9 shows that individuals with a full 34 years of positive earnings earn roughly 3.1 times more per year than individuals with only 17 years of positive earnings. Comparing this difference to the broader 6.3 times difference in lifetime earnings between the two groups suggests that roughly half of the lifetime earnings gap between having 34 versus 17 years of positive earnings remains even after accounting for the fact that the former worked more often than the latter.

The results in this section suggest that individuals of different sexes, races, marital and parental statuses, education levels, and labor market experience levels have significant differences in their lifetime earnings. However, large differences in lifetime earnings remain within groups of individuals with similar characteristics, as evidenced by the length of each bar in Charts 2–9. For example, women at the top quartile of the female-specific earnings distribution earn 2.8 times more than women at the bottom quartile of the female-specific earnings distribution. Similarly, STEM majors at the top quartile of the STEM-specific earnings distribution earn 2.5 times more than STEM-majors at the bottom quartile of the STEM-specific distribution. Lastly, even among individuals with a full 34 years of positive earnings, those at the top

*Chart 9*  
**Distribution of Average Yearly Earnings by Years with Positive Earnings**



Notes: The bottom and top of each box represent the 25th and 75th percentile of lifetime earnings, respectively. The white line inside each box represents median lifetime earnings, while the dot represents mean lifetime earnings.  
 Sources: U.S. Census Bureau and authors' calculations.

quartile earn nearly double what those at the bottom quartile earn. To better understand how much variation in lifetime earnings remains after all of these characteristics are accounted for, we next employ a more formal statistical analysis.

### III. How Much Do Observable Characteristics Account for Lifetime Earnings Differences across Individuals?

To formally assess the quantitative importance of basic observable characteristics such as sex, race, education, labor market experience, and marital and parental status in accounting for lifetime earnings differences across individuals, we estimate a simple linear regression that includes all of these characteristics. Specifically, we estimate the following model:

$$\begin{aligned} \log(LE_{i,p}) = & \alpha + v_p + \gamma age_{i,p} + \beta' X_i + \delta' Z_{i,p} + \varphi'(Z_{i,p} \times age_{i,p}) \\ & + \theta'(sex_i \times \tilde{Z}_{i,p} \times age_{i,p}) + \epsilon_{i,p}, \end{aligned} \tag{1}$$

where  $\log(LE_{i,p})$  represents the natural log of lifetime earnings of person  $i$  who was surveyed in panel  $p$ ,  $\alpha$  is an intercept term common to all

individuals,  $v_p$  is a panel-specific term that captures systematic differences that arise from being surveyed in one panel versus another,  $age_{i,p}$  is the age of person  $i$  at the start of panel  $p$ ,  $X_i$  is a vector of individual-specific characteristics that do not depend on age or the panel in which the person was interviewed (namely, sex, race, and the natural log of total number of years with positive earnings), and  $Z_{i,p}$  is a vector of individual-specific characteristics that can change over time. These characteristics include marital status, parental status, and education. Because these characteristics are only measured at a snapshot in time (rather than over an entire lifetime), we must account for the fact that people report these measures at different ages. We do so by interacting these variables with age at the start of the panel; that is, the term  $Z_{i,p} \times age_{i,p}$ . To account for the fact that marital and parental status may be associated with the lifetime earnings of men and women in different ways, we also include the interaction term  $sex_i \times \tilde{Z}_{i,p} \times age_{i,p}$ , where  $\tilde{Z}_{i,p}$  is a vector including marital and parental status. Lastly, we include an error or residual term,  $\epsilon_{i,p}$ , which captures all the variation in lifetime earnings that cannot be accounted for by the other terms.

Because we are interested in knowing how much of the variation in lifetime earnings observable characteristics can explain, we focus on the adjusted  $R^2$  from equation (1). The adjusted  $R^2$  is bounded between 0 and 1, where a value of 0 signifies that the explanatory variables account for none of the observed variation in lifetime earnings and a value of 1 implies that the explanatory variables account for all of the variation in lifetime earnings. Additionally, the adjusted  $R^2$  penalizes for overfitting or adding explanatory variables to equation (1) that do not help explain variation in lifetime earnings.

Table 2 shows the results from the regression, revealing that observable characteristics account for a little more than half of lifetime earnings differences across individuals. The first row of Table 2 shows that all of our controls together explain about 55 percent of variation in lifetime earnings, leaving 45 percent unexplained by the characteristics we include.

The remaining rows show how each factor contributes individually to this headline number.<sup>2</sup> Rows 2 and 3 show that sex alone explains about 10 percent of variation in lifetime earnings, while race alone explains slightly more than 1 percent. Rows 4 and 5 show that marital

*Table 2*  
Adjusted R<sup>2</sup> for Various Specifications

Specification	R <sup>2</sup>
1. All controls	0.5484
2. Sex	0.1039
3. Race	0.0125
4. Marital status, sex, and age when first interviewed	0.1256
5. Parental status, sex, and age when first interviewed	0.1298
6. Marital status and age when first interviewed	0.0091
7. Parental status and age when first interviewed	0.0102
8. Education and age when first interviewed	0.1486
9. Years of positive earnings	0.4076

Sources: U.S. Census Bureau and authors' calculations.

status and parental status each account for roughly 13 percent of variation in lifetime earnings. However, because these variables are interacted with sex, the reported R<sup>2</sup> statistics inherently include the explanatory power of sex—which, as previously mentioned, accounts for roughly 10 percent of lifetime earnings variation. To show the independent importance of marital status and parental status, rows 6 and 7 show the R<sup>2</sup> statistics for regressions that do not interact marital or parental status, respectively, with sex. Both rows show that marital status or parental status alone explain at most 1 percent of lifetime earnings differences.

Row 8 shows that education alone is an important characteristic when accounting for lifetime earnings differences across individuals. Specifically, educational differences alone account for roughly 15 percent of differences in lifetime earnings. While this number may appear to be small out of context, it is equivalent to roughly one-third of the headline R<sup>2</sup> statistic of 55 percent.

Row 9 shows that the single most important factor in accounting for lifetime earnings differences across individuals is the total number of years with positive earnings. Indeed, nearly 41 percent of lifetime earnings differences—or about three-quarters of the headline R<sup>2</sup> statistic—can be explained by differences in lifetime labor market experience. This finding should be interpreted with some caution, as without additional information, it is hard to discern the direction of causality. For example, individuals who work more over their lifetimes may have traits such as diligence or a strong work ethic that are rewarded in the labor market

with higher compensation. Alternatively, individuals might choose to work more throughout their lifetimes because they are more highly compensated for their time (for example, doctors or lawyers working extra hours may be able to charge higher fees for their services).

#### **IV. Conclusion**

Lifetime earnings can be influenced by characteristics determined at birth, decisions made prior to entering the labor market, and decisions made over one's productive career. We quantify the extent to which several factors explain differences in lifetime earnings and find that overall, observable characteristics account for little more than half of differences in lifetime earnings. Lifetime labor market experience, or the number of years an individual has positive earnings, has the strongest explanatory power among these characteristics. Characteristics such as sex, race, and education explain comparatively less of the variation in lifetime earnings, particularly when viewed in isolation.

Understanding the sources of lifetime earnings differences is critical for the design of social safety net policies such as Social Security and welfare. From a policy perspective, our findings underscore the importance of policies that promote labor market experience or attachment. Indeed, even in our selected sample of fairly attached individuals, our findings show that an additional year of labor market attachment can have potentially profound effects on lifetime earnings. Thus, programs that promote employment may have not only short-term earnings consequences but also long-term earnings consequences.



## Appendix

### Dataset and Variable Creation

This data appendix provides additional details on the data set used in our analysis along with information on the creation of our variables.

#### *The SIPP Synthetic Beta (SSB) version 6.0.2*

Version 6.0.2 of the SSB was released in 2015 and combines nine panels of the Survey of Income and Program Participation (SIPP) with administrative W-2 earnings records and benefit information. Specifically, the SSB includes the 1984, 1990, 1991, 1992, 1993, 1996, 2001, 2004, and 2008 panels. After merging the data sets, the Census fills in all missing values using sequential regression multivariate imputation (SRMI). For increased reliability of results, the Census runs SRMI four times to create four “gold standard” implicates.

From these four implicates, the Census creates entirely synthetic versions of the SSB. The Census runs SRMI on each gold standard implicate, replacing every value in the data set except for sex and links between husbands and wives. The Census runs SRMI on each implicate four times to create a total of 16 synthetic implicates. These 16 implicates are housed for public use on Cornell University’s Virtual Research Data Center. Our analysis code, while constructed and tested using the 16 synthetic implicates, is run on the gold standard implicates by the Census. The results from the four implicates are averaged to create one set of statistics released to us and presented in the paper.

#### *Variable definitions*

While our data set features all of the earnings and demographic data we need for our analysis, we still must create our own variables that match our preferred definitions. This can mean combining variables or recoding the values in variables we already have.

*Lifetime earnings.* The SSB does not include a lifetime earnings variable. However, it does feature three annual earnings variables for every year in our sample: capped earnings from all jobs covered by the Federal Insurance Contributions Act (FICA), total earnings from FICA-covered jobs, and total earnings from all non-FICA jobs. The capped earnings variable has the longest time horizon of 1951 to 2011. The total

earnings variables both cover the period from 1978 to 2011. While the longer time horizon is immediately appealing, accurately calculating lifetime earnings requires understanding what each variable measures. The first step to that is understanding FICA.

FICA taxes are the taxes that fund Social Security and Medicare. However, not all earnings are subject to FICA taxes. Certain jobs and people are exempt. For instance, employees of state and local governments are not always subject to FICA taxes (Social Security Administration 2017). Additionally, individuals can file for religious exemptions (U.S. Department of the Treasury 2018). FICA taxes are also only taxed on the first  $n$  dollars that a person makes. This limit has changed over time, but regardless of time period, many people make much more than the limit (Social Security Administration 2018). This is particularly true from 1951 to 1978. During these years, the percent of covered workers with earnings over the taxable maximum ranged from 15 to 36 percent (Social Security Administration 2015).

Our analysis requires data on all of a person's earnings, both FICA and non-FICA. Thus, we do not use the capped earnings from all FICA-covered jobs variable, as it does not include earnings from jobs not covered by FICA taxes and does not include earnings above the taxable maximum. While it does have a longer time horizon, the variable poorly measures total earnings during those additional years because of the large percentage of people earning more than the taxable maximum. The variables for total earnings from FICA-covered jobs and total earnings from all non-FICA jobs do include earnings above the taxable maximum. While each total earnings variable is missing an important piece of the puzzle (only earnings from FICA jobs, only earnings from non-FICA jobs), they can be combined to create a full earnings history for the people in our sample.

Given these annual earnings variables, we construct a lifetime earnings variable. First, we add together total earnings from FICA-covered jobs and total earnings from non-FICA jobs for each person for each year. We then convert these values into real terms using the CPI. Specifically, we use the seasonally adjusted annual CPI-U all items series. This series is indexed such that the value of the index in 2018 is 100. After converting to real terms, we add up the real annual earnings numbers over the entire sample for each person to generate a lifetime earnings

variable. We also create a variable for years with positive earnings by counting up the number of years in which a person's real annual earnings are greater than zero. This variable gives us an indication of labor market attachment.

*Demographic variables.* Besides lifetime earnings information, our analysis also requires a range of demographic information. We use the sex variable in the SSB. It has two values, male and female. We also use the race variable in the SSB, which has three values: white, black, and other. The variable is derived from the race variable in the original SIPP. However, during the 1984 through 2001 panels of the SIPP, the "other" category was broken down into "American Indian, Eskimo or Aleut" and "Asian or Pacific Islander." Because all people in the SSB have been in the SIPP, SIPP data can give us a good picture of the "other" race category in our sample. Specifically, during the 1990 through 2001 panels, 72 to 83 percent of all people classified as "other" were Asian or Pacific Islander. This means that results concerning the "other" race category mainly reflect the experience of Asians and Pacific Islanders.

Our education definitions require modifications of the variables in the SSB. The education variable in the SSB has five categories: less than a high school diploma, high school diploma, some college, college degree, and graduate degree. A college degree is defined as a Bachelor's degree, while an Associate's or technical degree is considered some college. We combine college degree and graduate degree into a single category. Additionally, we break down this combined college degree or more category by college major. Specifically, we separate college degree holders by whether their degree is in a STEM or non-STEM field. In general, we follow the STEM definition put out by the National Science Foundation. Table A-1 shows how we classify the college major categories present in the SSB. Note that the SSB changed its college major categories at the start of the 1996 panel.

In our analysis, we use binary variables for whether a person has ever been married and whether a person has ever had children. The SSB tracks marital status through the use of flags. These flags indicate new marriages and type of dissolution for up to four marriages. For our marital status variable, we are only interested in whether a person has ever married. Therefore, if the first marital flag indicates a first marriage, we count that person as married regardless of when or if this marriage

*Table A-1*  
College Majors in the SIPP Synthetic Beta by STEM Status

Pre-1996 panels		1996 onward	
STEM	Non-STEM	STEM	Non-STEM
Agriculture/Forestry	Unknown	Agriculture/Forestry	Unknown
Biology	Business/Management	Computer and Information Sciences	Art/Architecture
Economics	Education	Engineering	Business/Management
Engineering	English/Journalism	Health Sciences	Communications
Mathematics/Statistics	Home Economics	Mathematics/Statistics	Education
Medicine	Law	Nature Sciences	English/Literature
Nursing, Pharmacy, Health Technologies	Liberal Arts/Humanities	Psychology	Foreign Language
Physical or Earth Sciences	Police Science or Law Enforcement	Social Sciences/History	Liberal Arts/Humanities
Psychology	Religion/Theology		Philosophy/Religion/Theology
Social Sciences	Vocational or Technical Studies		Pre-Professional
	Other		Other

dissolved. In terms of child information, the SSB features a variable for number of children. We convert this information to a binary variable for whether a person has ever had any children. Admittedly, this variable is specifically for biological children. One might argue that such a variable would overlook the importance of adopted children. However, according to the 2007 National Survey of Adoptive Parents, only about 2 percent of the child population are adopted children. Additionally, less than half of adopted children live in households with no biological children (Vandivere and Malm 2009). Therefore, the number of people who would be mislabeled as never having children is negligible.

## Endnotes

<sup>1</sup>We follow the National Science Foundation definition of STEM fields, which includes majors such as chemistry, computer and information technology science, engineering, geosciences, life sciences, mathematical sciences, physics and astronomy, and social sciences (anthropology, economics, psychology, and sociology).

<sup>2</sup>Due to correlation among explanatory variables, adding up the rows exceeds the headline number of 55 percent. For example, if education and years with positive earnings are positively correlated—that is, if more educated individuals work more years—then adding their respective rows together amounts to double-counting because the education row captures some of the explanatory power of years with positive earnings and vice versa.

## References

- Guvenen, Fatih, Greg Kaplan, Jae Song, and Justin Weidner. 2018. "Lifetime Incomes in the United States over Six Decades." University of Minnesota, working paper, November.
- Mincer, Jacob. 1974. *Schooling, Experience, and Earnings*. New York: Columbia University Press.
- . 1958. "Investments in Human Capital and Personal Income Distribution." *Journal of Political Economy*, vol. 66, no. 4, pp. 281–302.
- Social Security Administration. 2018. "Contribution and Benefit Base." Accessed January 22, 2019. Available at <https://www.ssa.gov/oact/cola/cbb.html>
- . 2017. "How State and Local Government Employees Are Covered by Social Security and Medicare." Publication No. 05-10051.
- . 2015. "Covered Workers with Earnings Over the Taxable Maximum in a Given Year." Accessed January 22, 2019. Available at <https://www.ssa.gov/policy/docs/population-profiles/tax-max-earners-alt.html#chart1>
- U.S. Census Bureau. 2015. *SIPP Synthetic Beta: Version 6.0.2, 1978–2011*. Accessed through Cornell University Synthetic Data Server, Ithaca, NY.
- U.S. Department of the Treasury, Internal Revenue Service. 2018. "Form 4029 Application for Exemption from Social Security and Medicare Taxes and Waiver of Benefits."
- Vandivere, Sharon, and Karin Malm. 2009. *Adoption USA: A Chartbook Based on the 2007 National Survey of Adoptive Parents*. U.S. Department of Health and Human Services.