Uncovering the American Dream: Inequality and Mobility in Social Security Earnings Data since 1937

Wojciech Kopczuk, Columbia and NBER
Emmanuel Saez, UC Berkeley and NBER
Jae Song, SSA[1,2]

September 15, 2007

[1]Wojciech Kopczuk, Columbia University, E-mail address: wkopczuk@nber.org. Emmanuel Saez, University of California, Department of Economics, 549 Evans Hall #3880, Berkeley CA 94720. E-mail address: saez@econ.berkeley.edu. Jae Song, Social Security Administration, Office of Research, Evaluation, and Statistics, 500 E Street, SW 9th Floor, Washington DC 20254, E-mail address: jae.song@ssa.gov.
We thank Clair Brown, David Card, Jessica Guillory, Russ Hudson, Jennifer Hunt, Lawrence Katz, Alan Krueger, Thomas Lemieux, Michael Leonesio, Joyce Manchester, Robert Margo, David Pattison, Michael Reich, and numerous seminar participants for helpful comments and discussions. We also thank Ed DeMarco, Linda Maxfield, and especially Joyce Manchester for their support, Bill Kearns, Joel Packman, Russ Hudson, Shirley Piazza, Greg Diez, Fred Galeas, Bert Kestenbaum, William Piet, Jay Rossi, Thomas Mattson for help with the data, and Thomas Solomon and Barbara Tyler for computing support. Financial support from the Sloan Foundation and NSF Grant SES-0617737 is gratefully acknowledged.

[2]Series are available at http://www.columbia.edu/∼wk2110/uncovering

Abstract

This paper uses Social Security Administration longitudinal earnings micro data since 1937 to analyze the evolution of inequality and mobility in the United States. Earnings inequality follows a U-shape pattern, decreasing sharply up to 1953 and increasing steadily afterwards. We find that short-term (rank based) mobility among all workers has been quite stable since 1950 (after a temporary surge during World War II). Therefore, the pattern of annual earnings inequality is very close to the pattern of inequality of longer term earnings. Mobility at the top has also been very stable and has not mitigated the dramatic increase in annual earnings concentration since the 1970s. Our series display increases in long-term mobility among all workers. However, the decrease in the gender earnings gap and the substantial increase in upward mobility over a career for women is the driving force behind the increase in long-term mobility measures which mask slight declines in mobility among men. In contrast, overall inequality and mobility patterns are not significantly influenced by the changing size and structure of immigration nor by changes in the black/white earnings gap.

1 Introduction

One of America's most celebrated values is giving its people the opportunity to move up the economic ladder over their lifetimes. This opportunity, often summarized by the "American Dream" expression, is considered as a key building block of the U.S. social fabric. It is seen as the best antidote against the high levels of annual earnings inequality which the free market American economy generates. It also carries the promise that economically disadvantaged groups such as women, ethnic minorities, or immigrants can achieve economic success within their lifetime.[1] Although the concept of the "American Dream" is hotly debated in the press and among policy makers and the broader public, it has never been rigorously measured over long periods of time due to lack of suitable data. In order to understand fully the evolution of economic disparity and opportunity in the United States, it is therefore crucial to combine the analysis of earnings inequality with the analysis of long-term mobility.

A large body of academic work has analyzed earnings inequality and mobility in the United States. A number of key facts on earnings inequality from the pre-World War II years to the present have been established:[2] (1) Earnings inequality decreased substantially during the "Great Compression" of the 1940s (Goldin and Margo, 1992) and remained low over the next two decades, (2) Earnings inequality has increased substantially since the 1970s and especially during the 1980s (Katz and Murphy, 1992; Katz and Autor, 1999), (3) the top of the earnings distribution experienced enormous gains over the last 25 years (Piketty and Saez, 2003), (4) short-term rank-based mobility has remained fairly stable (Gottschalk, 1997) since the 1970s, (5) the gender gap has narrowed substantially since the 1970s (Goldin, 1990; Blau, 1998; Goldin, 2006a). There are, however, important questions that remain open due primarily to lack of homogenous and longitudinal earnings data covering a long period of time.

First, no annual earnings survey data covering most of the US workforce are available before the 1960s so that it is difficult to measure overall earnings inequality on a consistent basis before the 1960s and in particular analyze the timing of the Great Compression during the World

[1] The American Dream expression is also used to refer to intergenerational upward mobility. Our paper focuses only on mobility within a life-time. See Solon (1999) for a survey on intergenerational earnings mobility.

[2] A number of studies have also analyzed inequality in America in earlier periods (see Lindert, 2000, for a survey)

War II decade. Second and as mentioned above, studies of mobility have focused primarily on short term mobility measures due to lack of long and large longitudinal data. Therefore, little is known about earnings mobility across a full career, let alone how career mobility has evolved over time and whether it has contributed to reducing economic disparity across gender and ethnic groups. Third and related, there is a controversial debate on whether the increase in inequality since the 1970s has been offset by increases in earnings mobility. To the extent that individuals can smooth transitory shocks in earnings using savings and credit markets, inequality based on longer periods than a year is a better measure of economic disparity. Two recent findings in the literature suggest that mobility might have mitigated inequality increases. Krueger and Perri (2006) and Slesnick (2001) argue that consumption inequality has not increased despite the increase in income inequality.[3] Kopczuk and Saez (2004), Kennickell (2006), and Scholz (2003) find no major increase in wealth concentration in the 1980s and 1990s in spite of the surge in top income shares.[4]

The goal of this paper is to use the large Social Security Administration (SSA) micro data available since 1937 to make progress on those questions. The SSA data we use combine four key advantages relative to the data that have been used in previous studies on inequality and mobility in the United States. First, the SSA data we use for our research purposes are very large: a 1% sample of the full US covered workforce is available since 1957, and a 0.1% sample since 1937. Second, the SSA data are annual and cover a very long time period of almost 70 years. Third, the SSA data are longitudinal as samples are selected based on the same Social Security Numbers every year. Finally, the earnings data have very little measurement error and are fully uncapped (with no top code) since 1978.

Although Social Security earnings data have been used in a number of previous studies (often matched to survey data such as the Current Population Survey), the data we have assembled for this study overcome three important previous limitations. First, from 1951 to 1977, we use

[3]Those results are challenged by Attanasio et al. (2007) who question the quality of the Consumer Expenditure Survey (CEX) data used in those studies. The earlier study by Cutler and Katz (1991) found an increase in consumption inequality in CEX data.

[4]Edlund and Kopczuk (2007) argue that an increase in *intergenerational* mobility at the top of the distribution explains this pattern.

quarterly earnings information to extrapolate earnings up to 4 times the Social Security annual cap, allowing us to study groups up to the top percentile of the earnings distribution.[5] Coarser quarterly earnings information from 1946 to 1950, and the fact that top code censoring was above the top quintile from 1937 to 1945 allows us to study earnings up to the top quintile over the full period. Second, we can match the data to employers and industry information starting in 1957 allowing us to control for expansions in Social Security coverage which started in the 1950s. Finally, and perhaps surprisingly, the Social Security annual earnings data before 1951 have never been used outside SSA for research purposes.

Few socio-demographic variables are available in the SSA data relative to standard survey data. Date of birth, gender, place of birth (including a foreign birth indicator), and race are available since 1937. Employer information (including geographic location, industry, and size) is available since 1957. Because we do not have information on important variables such as family structure, education, and hours of work, our analysis will focus only on earnings rather than wage rates and will not attempt to explain the links between family structure, education, labor supply and earnings, as many previous studies have done. In contrast to studies relying on income tax returns, the whole analysis is also based on individual rather than family-level data. Furthermore, we focus only on wage earnings and hence exclude self-employment earnings as well as all other forms of income such as capital income, business income, and transfers. We further restrict our analysis to employment earnings from commerce and industry workers which represents about 70% of all US employees as this is the core group always covered by Social Security since 1937.[6]

We construct continuous and homogeneous series of employment earnings inequality and mobility for our core sample of commerce and industry workers. First, we construct inequality measures such as Gini coefficients, and earnings shares of various groups such as quintiles, and

[5]Previous work using SSA data before the 1980s has almost always used data capped at the Social Security annual maximum (which was around the median in the 1960s) making it impossible to study the top half of the distribution.

[6]We can test whether our findings are robust to including all covered workers. Since in recent decades Social Security covers well over 90% of earnings, this robustness check helps in addressing whether transitions in and out of covered employment affect our findings for the recent period. However, we cannot perform such a robustness check for earlier periods when coverage was much less complete.

smaller upper income groups. We construct these measures based on annual incomes but also based on longer measures such 5 or 11 year earnings averages. Second, we construct short-term mobility series showing the probability of moving from one quantile to another quantile from one year to the next. Fourth, we construct two types of long-term mobility series. The first type measures mobility of long term 11 year earnings spans after 10, 15 or 20 years relative to the full work force. The second type controls for age structure by measuring mobility within one's birth cohort. Fifth, we compute cohort-level measures of career long earnings inequality. Finally, in order to analyze the role of women, Blacks and immigrants, we also construct measures of group gaps such as the fraction of Women, Blacks, or immigrants by quantiles as well as series of inequality and mobility specific to those groups.

Our series allow us to uncover three main findings. First, our annual series confirm the U-shape pattern of earnings inequality since the 1930s, decreasing sharply up to 1953 and increasing steadily and continuously afterwards. The U-shape pattern of inequality is also present within each gender group and is more pronounced for men. Uncapped earnings data since 1978 show that earnings shares of all groups except the top 5% have decreased over the last 25 years. Furthermore, the increases within the top 5% have been concentrated among the top 1% and especially the top 0.1%. Therefore the pattern of *individual* top earnings shares is very close to the *family* top earnings shares constructed with tax return data in Piketty and Saez (2003).

Second, we find that short-term and medium-term mobility among all workers has been quite stable since the 1950s.[7] Therefore, the pattern of annual earnings inequality is very close to the pattern of inequality of longer term earnings. Those findings cast doubts on consumption based studies (Krueger and Perri, 2006; Slesnick, 2001) claiming that increases in annual earnings inequality overstate increases in economic welfare inequality. Furthermore, mobility at the top of the earnings distribution, measured by the probability of staying in a top group after 1, 3, or 5 years has also been very stable since 1978 and therefore has not mitigated the dramatic increase in annual earnings concentration.

Third, we find that long term career mobility measures for all workers display visible increases since 1951 either when measured unconditionally or when measured within cohorts. However,

[7]Mobility was unsurprisingly higher during the World War II decade but this was a temporary increase due to the large turnover in the labor market generated by the War.

those increases mask substantial heterogeneity across demographic groups. The decrease of the gender gap in earnings, which started in the late 1960s, has taken place throughout the distribution, including the very top, and has contributed greatly to reducing long-term inequality and increasing long-term mobility across all workers. In particular, upward mobility over a career was much lower for women than for men but this mobility gap has also been reduced significantly in recent decades. Economic progress of women is therefore the driving force behind the increase in overall long-term mobility measures which mask stability or even slight declines in long-term mobility among men. In contrast, overall inequality and mobility patterns are not significantly influenced by the changing size and structure of immigration nor by changes in the black/white earnings gaps. Consistent with previous work (e.g., Smith and Welch, 1989; Donohue and Heckman, 1991), we find a sharp narrowing of the Black vs. White gap exactly during World War II and resuming in the early 1960s but ending abruptly in the late 1970s except within the top percentile of the earnings distribution.

The paper is organized as follows. Section 2 describes the data and our estimation methods. Section 3 presents inequality results based on annual earnings. Section 4 focuses on short-term mobility and its effects on inequality while Section 5 focuses on long-term career mobility and inequality. Section 6 explains how the evolution of gender and ethnic gaps has affected overall patterns of long-term mobility and inequality. Finally, Section 7 offers some concluding remarks. The complete details on the data and our methodology, as well as sensitivity analysis are presented in the appendix of the working paper version (Kopczuk et al., 2007). Complete tabulated results in electronic format are posted online.

2 Data and Methodology

2.1 Social Security Administration Data

• Data

We use primarily datasets constructed in the Social Security Administration for analytical purposes known as the Continuous Work History Sample (CWHS) system. Detailed documentation of these datasets can be found in Panis et al. (2000). These datasets are derived from

the administrative-level data and their primary purpose is to support research and statistical analysis. The annual samples are selected based on a fixed subset of digits of the transformation of the Social Security Number. The same digits are used every year and the sample can be treated as a random sample of the data. We use three main datasets from SSA.

(1) The 1% CWHS file contains information about taxable social security earnings from 1951 to date (2004), basic demographic characteristics such as year of birth, sex and race, type of work (farm or non-farm, wage or self-employment), self-employment taxable income, insurance status for the Social Security Programs, and several other variables. Because Social Security taxes apply up to a maximum level of annual earnings, however, earnings in this dataset are effectively top-coded at the annual cap before 1978. Starting in 1978, the dataset also contains information about full compensation derived from the W2 forms, and hence earnings are no longer top coded. Employment earnings (either FICA employment earnings before 1978 or W2 earnings from 1978 on) include full wage income compensation including all salaries, bonuses, and exercised stock-options exactly as wage income reported on individual income tax returns.

(2) The second file is known as the Employee-Employer file (EE-ER) and we will rely on its longitudinal version (LEED) that covers 1957 to date. While the sampling approach based on the SSN is the same as the 1% CWHS, individual earnings are reported at the employer level so that there is a record for each employer a worker is employed by in a year. This dataset contains basic demographic characteristics, compensation information subject to top-coding at the employer-employee record level (and with no top code after 1978), and information about the employer including geographic information and industry at the three digit (major group and industry group) level. The industry information allows us to control for expansion in coverage overtime (see below).

Importantly, the LEED (and EE-ER) dataset also includes imputed wages above the taxable maximum from 1957 to 1977. The imputation procedure is based on the quarter in which a person reached the taxable maximum and is discussed in more detail in Kestenbaum (1976, his method II). The idea is to use earnings for quarters when they are observed to impute earnings in quarters that are not observed (because the annual taxable maximum has been reached) and to rely on a Pareto interpolation when the taxable maximum is reached in the first quarter. The

number of individuals who were top-coded in the first quarter and whose earnings are imputed based on the Pareto imputation is less than 1% of the sample for almost all years.[8] Consequently, high-quality earnings information is available for the bottom 99% of the sample allowing us to study both inequality and mobility up to the top percentile.

(3) Third, we also have access to the so-called .1% CWHS file (one tenth of one percent) that is constructed as a subset of the 1% file but covers 1937-1977. This file is unique in its covering the "Great Compression" of the 1940s. The .1% file contains quarterly earnings information starting with 1951 (and quarter at which the top code was reached for 1946-1950), thereby extending our ability to deal with top-coding problems (see below).

• **Top Coding Issues**

From 1937 to 1945, no information above the taxable ceiling is available. From 1946 to 1950, the quarter at which the ceiling is reached is available. From 1951 to 1977, quarterly earnings (up to the quarter at which the annual ceiling is reached) are available. Finally, since 1978, the data are fully uncapped. To our knowledge, the quarterly earnings information seems to have been retained only in the LEED 1% sample since 1957 and in the 0.1% CWHS sample since 1951 that we are using in this study. Earnings above the taxable ceiling (from 1937 to 1945) and above 4 times the taxable ceiling (from 1946 to 1977) are imputed based on Pareto distributions calibrated from wage income tax statistics published by the Internal Revenue Service and the top wage income shares series estimated in Piketty and Saez (2003).[9] From 1937 to 1945, the fraction of workers top coded (in our sample of interest defined below) increases from 3.6% in 1937 to 19.5% in 1944 and 17.4% in 1945. The number of top-coded observations increases to 32.9% by 1950, but the quarter when a person reached taxable maximum helps in classifying people into broad income categories. This implies that we cannot study groups smaller than the top 1% from 1951 to 1977 and we cannot study groups smaller than the top quintile from 1937 to 1950. We restrict our mobility series and multi-annual income shares to groups and years where those imputations do not have a significant impact on our series.

[8] The exceptions are 1964 (1.08%) and 1965 (1.17%)

[9] See Kopczuk et al. (2007) for complete details. For 1946-1950, the imputation procedure also uses Pareto distributions and preserves the rank order based on the quarter when the taxable maximum was reached.

• **Changing Coverage Issues**

Initially, Social Security covered only "commerce and industry" employees defined as most private for-profit sector employees and excluding farm and domestic employees as well as self-employed workers. Since 1951, there has been an expansion in the workers covered by Social Security and hence included in the data. An important expansion took place in 1951 when self-employed workers, farm and domestic employees were included. This reform also expanded coverage to some government and non-profit employees (including large parts of education and health care industries), with coverage further significantly increasing in 1954 and then slowly expanding since then. In order to focus on a consistent definition of workers, we include in our sample only commerce and industry employment earnings. Using SIC classification in the LEED, we define commerce and industry as all SIC codes excluding agriculture, forestry and fishing (01-09), hospitals (8060-8069), educational services (82), social service (83), religious organizations and non-classified membership organizations (8660-8699), private households (88), public administration (91-97).

Between 1951 and 1956, we do not have industry information as the LEED starts in 1957. Therefore, we impute "commerce and industry" classification using 1957-1958 industrial classification as well as discontinuities in covered earnings from 1950 to 1951 (see Kopczuk et al. (2007) for complete details). In 2004, commerce and industry employees are about 70% of all employees and this proportion has declined only very modestly since 1937. Using only commerce and industry earnings is a limitation for our study, especially for the analysis of mobility (as workers may move in and out of the commerce and industry sector over a career). Kopczuk et al. (2007) shows that our series for recent decades are robust to including all covered workers. However, we cannot test whether our series would also be robust to that extension in the early decades of the period we are considering.

• **Sample Selection**

For our primary analysis, we are restricting the sample to adult individuals aged 18 and above (by January 1st of the corresponding year) up to age 70 (by January 1st of the corresponding year). This top age restriction allows us to concentrate on the working-age population, while recognizing that some high-income individuals may continue earning very high incomes even

beyond the standard retirement age. Second, we consider for our main sample only workers with annual (commerce and industry) employment earnings above a minimum threshold defined as one-fourth of a full year-full time minimum wage in 2004 ($2575 in 2004), and then indexed by nominal average wage growth for earlier years. From now on, we denote this sample the "core sample". We show in Kopczuk et al. (2007) that virtually all of our results are unaffected if we choose alternative minimum thresholds or if we limit the sample to workers aged 25 to 60.

Figure 0 presents (on the left axis) the average and median real annual earnings for the "core sample". The figure shows that average earnings (expressed in 2004 dollars using the standard CPI deflator) have increased from $15,000 in 1937 to $39,200 in 2004. As is well known, median earnings grew quickly from 1938 to 1973 and have hardly increased over the last 30 years. Figure 0 also displays (on the right axis) the number of workers in our sample. The number of adult covered workers has increased from 27 million to 95 millions over the period (130 million without the commerce and industry restriction).

2.2 Constructing Inequality and Mobility Series

• Dividing Individuals into Groups

The first step of the analysis is to divide individuals into various earnings groups. For this purpose, for each year t from 1937 to 2004, all core sample workers are divided into 10 groups from the bottom quintile P0-20 to the top 0.1% (P99.9-100). The rest of the records for year t (those not yet 18, those above 70, those who are deceased and those who have no earnings or earnings below the minimum threshold) form an 11th group called the Missing group. Such groups are in general defined relative to the full population of interest. Sometimes, we will restrict the population of interest to men or women only, or smaller age or cohort groups. Table 1 displays the level of earnings for each of the groups we consider in 2004.

We will refer to P0-20 and P20-40 (the bottom two quintile) as the bottom groups. The median quintile P40-60 with average earnings of $26,715 will be referred as the moderate income group. P60-80 and P80-90 with average earnings of $41,869 and $63,114 are considered as the middle-class groups. P90-95 and P95-99 with average earnings of $85,304 and $134,639 are considered as upper middle class. Groups within the top percentile (earnings above $219,000)

are considered as top groups.

In order to focus on longer term measures of inequality, we also divide individuals based on earnings averaged over 3, 5, or 11 years. In that case, zeros will be included in the average and the minimum threshold is imposed on earnings in the middle year only. This is done to keep the sample criteria the same for annual earnings and earnings over a number of years. The age restriction is imposed so that individuals are alive and have age between 18 and 70 in *all* years included in the average. Therefore, the only source of the difference between samples averaged over different number of years is due to the age restriction.

• Inequality and Mobility Series

We compute several types of inequality series such as Gini coefficients and shares of total earnings accruing to the earnings groups we have defined. Those inequality series are always defined relative to our core sample of interest (or sometimes specific subsamples for the core sample).

For each year from 1937 to present, we estimate a mobility matrix showing in each cell (a,b) the number of individuals falling in group a in year t and in group b in year $t+1$. Groups are defined as 11 earnings groups (or an aggregated subset of them) above. Conditional mobility series are then estimated as the fraction of individuals in group a in year t who are in group b in year $t+1$ conditional on not being missing in year $t+1$ (due to any reason such as age over 70, earnings below the minimum threshold, or death). We then repeat the same procedure but for mobility between year t and year $t+3$, and $t+5$. Some of those mobility series are computed for specific demographic groups but quantiles are defined relative to the full core sample of workers (unless otherwise stated).

We estimate two types of long term mobility series. The first type is unconditional. We use 11 year earnings spans and estimate mobility matrices between year t and year $t+10$, $t+15$, $t+20$. The sample is selected based on having earnings in the middle year t above the minimum threshold (other years can be zero) and meeting the age restriction 18-70 in all years of the 11 year span. The second is conditional on birth cohort. We estimate mobility matrices from the early career to middle career, middle to late career, and early to late career. Early career is defined as the calendar year the person reaches 25 to the calendar year the person reaches 36.

Middle and later careers are defined similarly from age 37 to 48 and age 49 to 60 respectively. For example, for a person born in 1944, the early career is calendar years 1969-1980, middle career is 1981-1992, and late career is 1993-2004. Those long-term career mobility matrices are always computed conditional on having *average* earnings in each career stage above the minimum threshold. Those mobility matrices are based on cohorts (so that we always compare individuals relative to the individuals born in the same year) and hence will always be presented by year of birth.

3 Cross Sectional Inequality

Figure 1 plots the Gini coefficient from 1937 to 2004 for the core sample of all workers and for various sub-groups separately. The Gini series for all workers follows a U-shape over the period which is consistent with previous work based on decennial Census data (Goldin and Margo, 1992), wage income from tax return data for the top of the distribution (Piketty and Saez, 2003), and CPS data available since the early 1960s (Katz and Autor, 1999). Our series displays a sharp decrease of the Gini from 0.45 in 1938 down to 0.38 in 1953 (the Great Compression) followed by a steady and continuous increase since 1953. The Gini coefficient surpassed the pre-war level in the early 1980s and is highest in 2004 at 0.50. Our series shows that the Great Compression is indeed the period of most dramatic change in inequality since the late 1930s and that it took place in two steps. The Gini coefficient decreased sharply during the war from 1941 to 1944, rebounded partly from 1944 to 1946 and then declined again from 1946 to 1953.[10] The series also shows close to a linear increase in the Gini coefficient over the five decades from 1953 to 2004 which illustrates that changes in inequality were not just limited to an episodic event concentrated primarily in the 1980s.

Figure 1 also shows that the pattern for males and females separately displays the same U-shape pattern. Interestingly, the Great Compression as well as the upward trend in inequality are even more pronounced for men than for all workers. This shows that the rise in the Gini

[10]However, there are substantial changes and turnover in the labor force during (as well as just before and after) the war. Kopczuk et al. (2007) analyze the Great Compression in more detail in order to factor out transitions in and out of the labor force, and show that the Compression is smoother among the workers continuously in the sample during that period.

coefficient since 1970 cannot be attributed to changes in gender composition of the labor force. The Gini for men shows a sharp increase from 1979 to 1988 which is consistent with the CPS evidence extensively discussed in Katz and Autor (1999). On the other hand, stability of the Gini coefficients for men and for women from the late 1950s through the late 1960s highlights that the overall increase in the Gini coefficient in that period has been driven by the changes in the relative earnings of men and women. This provides the first hint of the importance of changes in women's labor market behavior and outcomes, the topic we are going to return to later in the paper.

Finally, Figure 1 shows that the Gini series for the sub-sample of white workers (defined as all workers excluding Black workers) and for native workers are virtually identical to the Gini series for all workers. This shows that changes in the Black-White earnings gaps as well as changes in the size and composition of foreign born workers have had little *direct* impact on overall inequality.[11] We will come back to this in more detail in Section 6.

In order to understand better the mechanisms behind this inverted U-shape pattern, Figure 2 plots the earnings shares for various groups of the earnings distribution. In this Section, we focus on the 1 year, i.e. standard annual, earnings shares. We will discuss the multi-year averages (displayed in lighter grey) in the next Section.

Figure 2A plots the shares of P0-40 and P40-60. The share of the both groups start to increase only after the beginning of World War II, peaks in 1953, and declines steadily afterwards. P0-40 gains most in the post war period while P40-60 gains most during the War. By the early 1980s, all the gains from the "Great Compression" are lost. The earnings share for the bottom group P0-40 stabilizes after the late 1980s while it continues to decline for the moderate income group P40-60.[12] By 2004, the P0-40 and P40-60 shares are at their historical minimum, down by about 25% from their peak level in 1953.

Figure 2B reports the shares of P60-80 and P80-95 since 1951. P60-80 is stable up to the

[11] Foreign born workers can have *indirect* impacts on inequality by lowering the wages of native workers with whom they compete. We do not attempt to analyze such indirect effects in this study (see Borjas (1999) for a survey).

[12] This is consistent with "polarization" of the labor force documented in recent CPS based studies that show that inequality between the 90th percentile and the median widened in the 1990s while it slightly contracted between the median and the 10th percentile (Autor et al., 2007).

mid-1960s but declines steadily afterwards from around 25% to around 21%. P80-95 increases from 1950 to the early 1980s and declines slightly afterwards. This shows that the "middle class" did well in post war decades but lost ground in recent decades. Figure 2C displays the shares of upper income groups, the top 1% (P99-100) and the next 4% (P95-99), since 1951. P95-99 shows a steady increase since 1953. The top percentile decreases slightly till the mid-1960s and then more than doubles from about 6% in the 1960s to almost 14% at the peak in 2000. Therefore, all groups below the top 5% have experienced declines in their earnings share since the 1970s. This implies that not only the bottom quintiles but also the middle class and upper middle class (up to P95) have indeed be squeezed in relative terms by the gains at the top since 1970. Interestingly, the compression in the top part of the distribution lasts for several decades after the war. In contrast, the compression in the lower, middle, and upper middle part of the distribution starts to unravel by the mid 1950s. The different timings of these later changes suggests that different mechanisms took place in the top versus the rest of the distribution.

Finally, Figure 2D uses the uncapped data since 1978 to plot the earnings share for the top 0.1%. The top 0.1% share triples from 1978 to 2004 and this accounts for 60% of the gains accruing to the top 1% during this period. The SSA data confirm the findings of Piketty and Saez (2003) that the increase in top wage income shares is extremely concentrated.[13] The closeness of our SSA based (individual-level) results and the tax return based (family-level) results of Piketty and Saez show that changes in assortative mating played at best a minor role in the surge of family earnings at the top of the earnings distribution.

4 Short Term Mobility and Multi-Year Income Shares

4.1 Mobility at the Top

The longitudinal property of the SSA data can be used to analyze whether this surge in top incomes has been mitigated by an increase in mobility for the high income groups.

Figure 3A shows the probability of staying among the top 0.1% earnings group after 1,

[13] Kopczuk et al. (2007) show indeed that the Gini coefficient excluding the top 1% earners is fairly flat since the 1990s. This shows that results based on survey data such as official Census Bureau inequality statistics which do not measure well the top 1% can give an incomplete view of inequality changes.

3, 5 and 10 years (conditional on staying in our core sample) starting in 1978. The one-year probability is between 60% and 70% and it shows no overall trend. This pattern gives little hope for attributing any part of the increase in earnings share of the top 0.1% over this period to increased short-term fluctuations of incomes at the top. Longer term mobility measures are largely consistent with this conclusion, showing no overall trend in the 1980s and 1990s. Figure 2D confirms this point. It compares the share of earnings of the top 0.1% based on annual data with shares of the top 0.1% defined based on earnings averaged on the individual level over 3 and 5 years (displayed in lighter grey). As mentioned in Section 2, for comparability purposes with standard annual earnings shares, the sample for the 3 and 5 year averages is defined as core sample workers with earnings in the middle year above the minimum threshold. These longer-term measures naturally smooth short-term fluctuations but show the same pattern of robust increase as annual measures do.[14]

Figure 3B analyzes the transition from middle and upper middle class to the top 1%.[15] We consider top 1% income earners in a given year t and estimate in which group did those top 1% income earners belong to 10 years earlier (conditional on being in our core sample). The figure shows that, for top 1% earners in 2004, 38% belonged to the top 1% 10 years earlier (in 1994), about 36% belonged to P95-99, only 15% belonged to the "middle-class" groups P80-95, and a mere 11% belonged to the bottom four quintiles P0-80. Overall, the graph displays a relative stability over the last 50 years. The graph shows that the fraction coming from the top (P99-100 or P95-99) has increased slightly since the late 1970s. At the same time, the fraction coming from the "middle-class" has slightly declined. This is a reverse of the earlier pattern from the 1960s and 1970s where the odds of coming from middle class groups was actually increasing.

These findings suggest that while persistence of staying in the top of the distribution has remained stable, the very top is harder to reach today than 25 years ago unless you start very to close it. This graph provides some support for the notion of the "middle class" squeeze from the

[14]One would expect earnings averaged over a longer time period to be less concentrated than annual earnings because of year to year mobility. However, for multi-year averages, we use the same sample as annual averages. As a result, some earners have zero earnings in years outside the middle year and this increases inequality. Those two effects compensate each other so that the levels across the series displayed on Figure 2D are similar.

[15]Because our data prior to 1978 is top-coded, the top 1% is the smallest group for which we can show longer term patterns.

popular press: as we discussed in Section 3, income earners below the top 5% have experienced declines relative to the average. Meanwhile, top 1% incomes have more than doubled (relative to the average). Thus, at the same time as the gap in earnings between the middle class and the top percentile was drastically widening, it was becoming less likely that an upper middle class earner could reach the top percentile within 10 years.

4.2 Mobility in the rest of the distribution

Figures 2A, 2B, and 2C that we discussed above also display (in lighter grey) income shares averaged over 5 year and 11 year periods. The patterns of annual inequality are virtually identical to the 5 and 11 year patterns. The only notable difference, to which we come back in Sections 5 and 6, is that the drop in the earnings share of P0-40 since 1970 is arguably less pronounced for 5 and especially 11 year averages than for 1 year averages. The surge in the top 1% income share for earnings averaged over 5 or 11 years is virtually identical to the surge for annual earnings. Those results show that year to year mobility has modest effects on the pattern of overall economic inequality. As a result, annual earnings inequality provide a very good proxy for the level and evolution of longer term earnings inequality in the United States. Those findings cast doubts on the findings from Krueger and Perri (2006) and Slesnick (2001) arguing that consumption inequality has not increased. It is difficult to understand how the dispersion of consumption could be stable when long-term earnings (averaged over a 11 year period) disparity displays such a dramatic increase, especially given lack of evidence of significant changes in the wealth distribution since the early 1980s (Kennickell, 2006).

Figure 4A reports the probability of staying in the bottom two quintiles P0-40 or top two quintiles P60-100 after 1 year (again conditional on remaining in the core sample after 1 year). Two basic findings should be noted from those graphs. First, the probability of staying in the top quintiles is higher than the probability of staying in the bottom quintiles, showing that being in the bottom of the distribution in any one year is more a transitory state (on average) than being in the upper part of the distribution.[16] Second, there is certainly no secular increase in mobility over the 70 year period we analyze. After a temporary surge during the War period

[16]This is consistent with previous work based on the PSID such as Moffitt and Gottschalk (1995) or Buchinsky and Hunt (1999).

(which can be explained by the unusually large turnover and changes in the labor force during that period), mobility has been fairly stable since 1950 and if anything has declined slightly so that it is at its lowest in recent years.

Figure 4B examines the probability of downward mobility from P60-100 down to P0-40 and the upward mobility from P0-40 to P60-100. Comparing Figures 4A and 4B shows that downward and upward mobility is unsurprisingly much less likely than stability. Downward mobility captures the notion of earnings instability. The figure shows that it is closely correlated with the business cycle (recession months are displayed in light grey bars, with year tickmarks located in the middle of the year). Spikes in downward mobility are clearly visible during recessions but there is no major long-term trend. Upward mobility was significantly higher in the 1940s (most likely because of significant turnover during those years), and has declined slowly and steadily since the 1950s and appears also to be around its lowest in recent years. Interestingly, the series of downward and upward mobility are negatively correlated (this is most clearly visible in recent decades): in recessions, downward mobility increases while upward mobility decreases and the reverse happens during booms.[17]

In sum, the movements in short-term mobility appear to be much smaller than changes in inequality. As a result, changes in short-term mobility have had no significant impact on inequality patterns in the United States. Those findings are consistent with previous studies for recent decades based on PSID data (see e.g., Gottschalk, 1997, for a summary) as well as the most recent SSA data based analysis of Congressional Budget Office (2007)[18] and the tax return based analysis of Carroll et al. (2007). They are more difficult to reconcile, however, with the findings of Hungerford (1993) and especially Hacker (2006) who find great increases in *family* income variability in recent decades using PSID data.

[17] As (rank-based) upward and downward mobility need to compensate each other by definition, this suggests that recessions create more large losers (those who lose their job) and more small winners (those who can keep their jobs) than other times of the business cycle.

[18] The CBO study focuses on probabilities of large earnings increases (or drops) instead of quantile mobility as we do here.

5 Long-term mobility and Life Time Inequality

The very long span of our data allows us to estimate long-term mobility. Such mobility measures go beyond the issue of transitory earnings analyzed above and describe instead mobility across a full career. Such estimates have not been produced for the United States in any systematic way because of the lack of very long and large panels. Hence, our data can address some of the central questions on the issue of career mobility that are at the center of the debate on the "American Dream": what is the probability of getting toward the top when starting from the bottom within a lifetime? Has this social mobility grown or decreased in the United States over the last 6 decades? How does long-term mobility affect long-term inequality measures such as earnings averaged over a full career?

5.1 Unconditional Long-Term Mobility

We begin with the simplest extension of our previous analysis to a longer-term horizon. We estimate 11 year long average individual earnings. For year t, that means earnings from year $t-5$ to year $t+5$ and we classify individuals in quintiles based on those averages.[19] We report upward and downward mobility measures on Figure 5. Figure 5A displays upward mobility probabilities from P0-40 to P80-100 after 10, 15, and 20 years. If earnings after 10, 15, or 20 years were independent of base earnings, the probability of moving up to the top quintile would be 20%. The probability is substantially lower than this although it reaches about 10% for the 20 year mobility graph for the most recent years. The graph shows increases in upward long-term mobility (especially after 20 years) since the 1950s, with stabilization or even slight decline toward the end of the period. Figure 5B displays downward mobility probabilities from P80-100 to P0-40 after 10, 15, and 20 years. This figure shows that long-term downward mobility is less likely than upward mobility, which is consistent with increasing wage profiles over a career. The figure also shows a clear increase in the probability of downward mobility over time especially since the 1960s which seems to have stabilized in more recent years. Those graph suggests that, in contrast to the stability of short-term mobility documented in Section 4, there was

[19] As above, the sample is selected conditional on the middle year t earnings above the minimum threshold and conditional on being aged 18 to 70 during the full 11 year window.

a noticeable increase in long-term upward and downward mobility among workers in the US economy over the last 5 decades.

5.2 Cohort based Long-Term Mobility

The analysis so far ignored changes in the age structure of the population as well as changes in the wage profiles over a career. In order to control for those effects, we turn to cohort-level analysis. Figure 6 displays long-run mobility series.[20] Figure 6A focuses on upward mobility and reports the probability of moving to the top quintile conditional on being in the bottom two quintiles. Figure 6B focuses on downward mobility (the probability of moving down to P0-40 when starting from P80-100). Each panel reports 3 mobility series: from the early part of the career (age 25 to 36) to the middle career (age 37 to 48), from middle to late career (age 49 to 60), and from early to late. We have also extrapolated in lighter grey the series up to six years.[21]

Two important results should be noted. First, mobility over a life-time is relatively modest. For example, Figure 6A shows that, the probability of moving to the top quintile starting from P0-40 from early to middle is 3-4% and is only 2-3% from middle to late. The same probability from early to late career is on average around 6%. If there were no correlation, those probabilities should be 20%. This shows that there is a quite substantial but not deterministic relationship in earnings across those broad lifetime episodes. Figure 6B shows the probability of downward mobility from the top quintile to the bottom two quintiles is also significantly lower than in the no correlation case.

Second, the pattern of mobility over the period displays modest increases in mobility over the period we analyze. Those changes are most visible in the mobility from early to late career. For example, Figure 6A shows that upward mobility from early to late career increased from

[20] Due to top-coding problems, we restrict attention to quintiles of the distribution and observations that can be constructed using data starting with 1951. Imputations do not have an effect on our results as long as they do not lead to mis-classifying individuals. Since we assign earnings randomly only within the top 1% (in 1951-1977), we can construct longer-term quintiles as long as all individuals in the top 1% stay in the top quintile of the long-term distribution. This is true with probability close to one.

[21] As explained in detail in Kopczuk et al. (2007), those extrapolations are based on series using truncated parts of each career stage.

less than 6% for cohorts born before the Great Depression to over 8% for cohorts born just after World War II. Symmetrically, Figure 6B shows that the probability of downward mobility also increased from less than 10% to over 13%.

Those cohort-based results are consistent with the unconditional long-term mobility results from the previous section. They suggest that, in contrast to the annual inequality and short-term mobility series described above which point to increasing economic disparity, long-term mobility series appear to show modest increases in mobility.

5.3 Long-Term Inequality

Figure 7 reports the top quintile (Figure 7A) and bottom two quintiles (Figure 7B) earnings share in early, middle, and late career by birth cohort. The top quintile earnings shares are consistent with annual inequality and the long-term mobility pattern we have uncovered. Interestingly, the series also show that there is much more income concentration in late career than in middle career, and in middle career than in early career. Coupled with an increasing pattern at all stages, it suggests that overall inequality may further increase as currently young cohorts age.

In contrast, Figure 7B shows that the share of P0-40 has declined for early cohorts but has then increased for cohorts born after 1940. Hence, bottom quintiles are actually doing better when we consider a longer term perspective, especially in the early part of the career. Those results are striking in light of our results from previous sections showing a worsening of the share going to bottom groups either in annual cross-sections or even for multi-year averages. Those results can actually be reconciled once compositional gender effects are understood. We turn to those effects in the next section.

6 The Role of Gender, Racial, and Native-Immigrant Gaps

Economic disparity across groups such as gender, ethnic, and native vs. foreign born groups is widely perceived as one of the central issues in American society, and one that has attracted a lot of attention from scholars. In the context of the analysis of overall inequality and mobility in this paper, we want to examine to what extent the closing (or widening) of economic gaps across those groups has contributed to shaping the patterns we have documented earlier.

6.1 Annual Earnings Gaps

We first document the broad facts on annual earnings gaps and their evolution over time, pointing out which facts were previously known and where the SSA data cast new light. As is well known, the direct comparison of average earnings across different groups of workers can be biased by composition effects such as differential changes in labor force participation[22], or changes in the wage structure.[23] A simple way to get around those composition effects with our data is to consider the fraction of blacks (or women) in each earnings group relative to the fraction of blacks (or women) in the adult population.[24] Those fractions with no adjustment capture the total realized gaps including labor supply decisions. As a result, they combine not only the traditional wage gap among workers but also the labor force participation gap (including the decision to work in the commerce and industry sector rather than other sectors or self-employment). Such measures have rarely been used when analyzing the gender or Black-White gaps[25] because economists have traditionally started by analyzing average wage ratios and then extended that analysis by looking at percentile wage ratios (such as the ratio of medians). However, we believe that the measure of the fraction female in a given group has several advantages relative to percentile wage ratio measures.

First, it is a very transparent measure that is easy to understand and interpret. Second, it is neutral with respect to changes in the wage structure. Indeed, a change in the wage structure can be defined as gender neutral if it leaves the fraction of women in each quantile unchanged. Finally, such measures could easily lend themselves to traditional decompositions in order to analyze the relative contribution of different factors (such as increased education, fertility or marriage decisions, etc.) as this is commonly done in the case of average wage ratios using

[22]For example, if unskilled women start working, this will automatically increase the gender wage gap. Correcting for such selection issues is discussed in the case of the gender gap by Blau (1998).

[23]For example, if Blacks are less skilled than Whites on average, an increase in the skill premium will increase the overall Black-White gap. Juhn et al. (1991) make this point and propose a decomposition. Blau and Kahn (1997) apply this to the gender gap.

[24]In practice, we do not make this adjustment for women as they are approximately 50% of the adult population so that the adjustment would correspond to a straightforward re-scaling of our figures.

[25]Such measures have often been used to measure occupational gaps (such as the fraction of women among CEOs, professors in universities, etc.). See Blau (1998); Blau et al. (2006) for a summary of the literature on such gender occupational gaps.

survey data with richer demographic variables.

• **Gender Gap**

Figures 8A and 8B plot the fraction of women in our core sample and in various upper earnings groups. As adult women aged 18 to 70 are about half of the adult population aged 18 to 70, with no gender differences, those fractions should be approximately 0.5. For comparison purposes, we report on the right y-axis the traditional gender gap measured as average women earnings divided by average men earnings in our core sample (without any adjustment).

Consistent with previous work based on Census and CPS data (Blau et al., 2006), the fraction of women in the core sample of commerce and industry workers has increased from around 27% in 1937 to about 45% in 2004. World War II generated a temporary surge in women labor force participation, two thirds of which was reversed immediately after the war.[26] Women labor force participation has been steadily and continuously increasing since the mid 1950s and has been stable at around 45% since 1990.

Figure 8A shows that the representation of women in upper earnings groups has increased significantly over the last four decades and in a staggered pattern across upper earnings groups.[27] The fraction of women in P60-80 starts to increase in 1965 from around 13% and reaches about 38% in the early 1990s and has remained about stable since then. The fraction of women in the top decile (P90-100) does not really start to increase before the early 1970s. It grows from around 2% in 1973 to almost 22% in 2004 and is still quickly increasing. Figure 8B shows that the representation of women in the top percentile did not really start to increase before the late 1970s. In 2004, the representation of women is still sharply declining as one moves up the earnings distribution.[28]

[26]This is consistent with the analysis of Goldin (1991) who uses a unique micro survey data covering women workforce history from 1940 to 1951.

[27]There was a surge in women in P60-80 during World War II but this was entirely reversed by 1948. Strikingly, women were better represented in upper groups in the late 1930s than in the 1950s.

[28]Before the 1970s, a very large fraction of all college educated working women were teachers (Goldin et al., 2006, Table 5) who are not included in our commerce and industry core sample. Indeed, in the full sample including all industrial groups, the fraction women in the top 10% would double to 4% (instead of 2% in commerce and industry) in 1970. The fraction women in the top 10% in 2004 is 25% (when including all industries) instead of 22% in the commerce and industry core sample. The fraction of women in the top 1% is very close in both 1970

This staggered pattern could be explained by career effects (Goldin, 2004, 2006a): starting in the 1960s, women started entering new careers but it took time before they were able to reach the top of the ladders in their professions. Our findings are consistent with the previous literature (see e.g., Goldin, 1990; Blau and Kahn, 1997; Blau, 1998; Goldin, 2004; Blau and Kahn, 2006; Goldin, 2006b), which finds a narrowing of the gender gap especially during the 1970s and 1980s. It is useful to note that the (uncorrected) ratio of women to men average earnings decreases from 1950 to the early 1970s. Hence, the early gains of women at the top are masked by increased labor force participation of women with low earnings. Over the last 15 years, women's representation in the second to top quintile has stopped growing and is substantially below parity. However, it continues to increase in the top quintile (and especially within the top decile). This is the driving force today in the increase in the women to men average earnings ratio. The analysis of the representation of women by quantile groups has the virtue of showing very saliently where gains for women are taking place and where gains have stopped.

In contrast to the influential study by Albrecht et al. (2003) which does not find an increase in the ratio of percentiles of the distribution of men to those of women toward the top of the earnings distribution using CPS data, our results based on administrative data show that the fraction of women decreases continuously as one moves up the earnings distribution. Correspondingly, in our core sample in 2004 the percentile ratios of men to women increase in the top 10% as well: the log ratio is approximately 0.40 between P40 and P90, increases to 0.44 at P95, 0.75 at P99 and 1.04 at P99.9.[29] This implies that the distribution of earnings among men is not just shifted to the right relative to the distribution of women, but also that the upper tail of the men's distribution is thicker than the upper tail of the women's distribution.

and 2004 in the core sample and in the full sample as very few non commerce and industry workers are in the top 1%. This suggests that the dramatic trend upward in the representation of women at the top should be robust to including all industrial sectors.

[29] Albrecht et al. (2003) do find increasing percentile ratios in the case of Sweden using administrative data. We suspect that the difference between CPS and SSA data is due to top coding and measurement error in the CPS data. Hence, this "Glass Ceiling" phenomenon uncovered by Albrecht et al. (2003) in the case of Sweden seems also to be present in the United States. Such a pattern of increasing percentile ratios could be due to many other factors than "Glass Ceiling" (when "Glass Ceiling" is understood as discrimination preventing women from going above certain positions in their careers).

• **Black-White Gap**

Figures 9A and 9B plot the fraction of Black in our full sample and in various upper income groups relative to the Black share in the adult population (estimated using Census data). With no Black-White differences in the distribution of earnings, those fractions should be around one. For comparison purposes, we also report on the right y-axis the traditional Black-White gap measured as average Black earnings divided by average white earnings without any adjustment.

Figures 9A and 9B show that the Black-White gap has followed a different pattern from the gender gap. Blacks have made progress in the middle class and upper middle class groups during World War II. The average Black to White average earnings ratio display a striking step pattern: it increases sharply exactly during the war years from 1941 to 1945 and is flat afterwards. This is consistent with the census based analysis of Smith and Welch (1989), Donohue and Heckman (1991), and Margo (1995) comparing 1939 to 1949. Such a step pattern is most likely explained by economic migration of Blacks from the South to the North due to labor shortages during the war, and where Blacks remain in their better paid Northern industrial occupations after the war. Unfortunately, the SSA data before 1957 do not provide geographical information (beyond state of birth) allowing us to test this hypothesis in more detail. Such a sudden pattern is harder to reconcile with slow improvement in Blacks' education. In any case, such a sudden and permanent change is perhaps one of the most striking example of the "American Dream" at work for a very disadvantaged group.

After stability in the Black-White gap from the end of World War II to 1960, Blacks made significant progress from the early 1960s. Virtually all of our series on Figure 9A display a clear break starting in the early 1960s, arguably in 1961.[30] Interestingly, Black gains during the 1960s and the 1970s were actually smaller (and certainly much slower) than progress during World War II. Black progress stops around 1980 and is followed by a reversal except at the top of the distribution.[31] Indeed, while the representation of Blacks dropped significantly overall and in

[30] Dating exactly the beginning of Black's earnings gains is important to determine the causes. Donohue and Heckman (1991) emphasize this issue and the difficulty of dating the break point using survey data. Card and Krueger (1993) using matched CPS-SSA (top coded) earnings data date most of the reduction in earnings gap starting after 1965.

[31] A number of studies have tried to account for the lack of progress of Blacks' relative earnings since 1980 (Card and Lemieux, 1994; Juhn et al., 1991).

P60-80 since 1980, it was stable for P80-90 and P90-95, and actually increased significantly in the top percentile. It is also striking to see that, in contrast to women, the fraction Black in top 1% is actually lower than in the top 0.1%. This suggests that the composition of characteristics (such as occupation) of blacks in the top 1% is likely very different for blacks than for the rest of the population or that the labor market environment that blacks face is different from women.

• **Immigrant-Native Gap**

Figure 10 displays the fraction of immigrants (defined as foreign born individuals) in various quantiles in our core sample as well as in the Census adult population (as a comparison).[32] The series display a clear U-shape pattern: the fraction of immigrants in the core sample was above 10% before World War II, fell to around 5% by the 1960s and then increased to almost 15% by 2004. This U-shape series tracks relatively closely the Census based adult population estimates.

More importantly, the fraction of immigrants in the top two quintiles, as well as in the top 1%, is relatively close to the fraction of immigrants in the full core sample and follows the same U-shape pattern over the full period. For example, the fraction of immigrants in the top 1% is 13.6% in 2004 (and 14.9% in the full core sample). This stands in sharp contrast with women and Blacks who are under-represented at the top. Figure 10, nevertheless, shows that the distribution of immigrants has somewhat shifted toward the bottom over the full period: immigrants were over-represented at the top up to the late 1960s and have been slightly under-represented since then.[33] The patterns we find indicate that accounting for immigration is unlikely to make an important difference to measures of overall income distribution and explains why the Gini coefficient series for native workers is virtually identical to overall workers (Figure 1).

The evidence for women and blacks shows that they have made significant economic progress and in part shared the extraordinary gains at the top of the earnings distribution, although

[32]In contrast to the Black-White gap, we do not normalize our series by dividing them by the fraction of foreign born in the adult population because the SSA data we use include only foreign born workers who use Social Security Numbers issued by SSA. Undoubtedly, many undocumented immigrant workers use invalid or no Social Security Numbers at all. The number of foreign-born individuals in our data is actually close to CPS-based estimates, suggesting that problems related to under-counting undocumented immigrants are similar in the two datasets (see Hanson, 2006, for discussion of this problem in CPS).

[33]This is consistent with previous work based on Census data since 1960 (Borjas, 1999).

they are still under-represented at the top of the distribution. This suggests that those groups might have experienced significant upward mobility increases and hence might have contributed significantly to shaping the upward mobility patterns we have described in Section 5.

6.2 Long-Term Earnings Gaps

Figure 11A displays the long-term upward mobility from P0-40 to P80-100 after 20 years for 11 year averages for various groups: all (as in Figure 5A), men, women, Blacks, and foreign-born. The figure shows a striking heterogeneity across groups. First, men have significantly higher levels of upward mobility than overall workers and women and Blacks have significantly lower levels of upward mobility than overall workers. Thus, in addition to the annual earnings gap we documented, there is an upward mobility gap as well across groups. Second, the mobility gap has also been closing overtime: the probability of upward mobility among men has been stable overall since World War II with a slight increase up to the 1960s and declines after the 1970s. In contrast, the probability of upward mobility of women has continuously increased from a very low level of less than 1% in the 1950s to about 7% in the 1980s. The probability of upward mobility for Blacks also started low (around 2-3% in the 1950s) but increased earlier and more sharply than for women. It has however slightly declined since 1965. There is not much difference between upward mobility of foreign-born workers and the rest of the population. The increase in upward mobility for women and Blacks compensate for the stagnation or slight decline in mobility for men so that the overall upward mobility for all workers is slightly increasing.[34] Figure 11A also suggests that the gains in annual earnings made by women and Black documented earlier were in part due to women and Blacks already in the labor force making earnings gains rather than gains entirely due to the entry of new cohorts of women and Blacks with higher earnings.

Figure 11B focuses on career mobility within cohorts (as Figure 6). It displays upward mobility probabilities from early career (age 25-36) to late career (age 49-60) for men, women, and all workers. Similar to Figure 11A, it shows a large upward mobility gap across gender

[34]It is conceivable that upward mobility is lower for women (or Blacks) because even within P0-40, they are more likely to be in the bottom half of P0-40 than men. Kopczuk et al. (2007) show that controlling for those differences leaves the series virtually unchanged. Therefore, controlling for base earnings does not affect our results.

groups that closes overtime: men upward mobility is stable at around 12% while women mobility increases from 1-2% to around 7%. This shows again that the reason for the slight increase in upward career mobility for all workers is entirely due to the gains made by women.[35] It is important to emphasize here that including only commerce and industry earnings is a limitation for our study. For example, if women who start their career as teachers are more likely today (relative to decades ago) to shift to a commerce and industry occupation later in their career, then our series will display higher upward mobility in commerce and industry earnings today than decades ago although upward mobility in total earnings (including covered and uncovered occupations) might not have changed. We can show that upward mobility is very similar in the full sample and in the commerce and industry sample in recent decades. This is re-assuring but it does not rule out enitrely the possibility that mobility patterns across those two groups were different decades ago.[36]

Figure 12 shows that the share P0-40 over various career stages has actually declined sharply when the sample is restricted to men (rather than all workers as in Figure 7B). Interestingly, the drop starts in the early 1970s for each career stage which shows that the worsening of the economic condition for male low earners since the 1970s was a widespread phenomenon that is clearly visible from a long-run perspective. Furthermore, it is possible to show that this worsening economic situation for low earning men was even more pronounced among those men with strong attachment in the labor force (i.e., men working at least 10 years over the 12 year career stages we are considering).

Therefore, the gains of P0-40 displayed on Figure 7B for recent cohorts are due primarily to the increased attachment of women into the labor force. P0-40 used to include a large number of women with very weak labor force attachment and hence very low earnings making the P0-40 share low. The increased labor force attachment of women since the 1960s reduced the number of very low earners in P0-40 and hence drove the P0-40 share up. This effect was actually so strong that it can entirely mask the worsening economic situation of low earning men displayed on Figure 7B.

[35]Those results are also robust to controlling for differences in the distribution of base earnings for men and women within P0-40.

[36]The same critique can be made for movements between employment and self-employment.

Thus, one can say that low income earners have gained modestly in recent cohorts. However, those modest gains are the net effect of great gains experienced by women who work more regularly than before and earn more than before when they work combined with great losses experienced by low earning men. Hence, it appears that women gains were at least partly men's losses, a point that has previously been proposed by Fortin and Lemieux (1998).

Figure 13 displays the fraction of women in the top quintile from a long-term perspective by cohorts at each stage of the career. Three results are worth noting. First, it shows that the period starting after the mid-1960s was favorable to all women (and not only young women): the share of women in the top quintile increases around the 1920 cohort for late career women (aged 49-60), around the 1930 cohort for mid career women (aged 37-48), and around the 1941 cohort for early career women (aged 25-36). This demonstrates that women's progress cannot be entirely due to a change in education, fertility or marriage status, or career decisions of young women. Second, Figure 13 also shows a sharp break in the early and middle career graphs starting with the 1941 cohort. This means that there was also an additional positive effect on women born starting with the 1941 cohort. This is consistent with the sharp breaks found by Goldin (2004, 2006a) in various series such as college graduation of women, fraction women in professional schools, age of first marriage of educated women, or employment expectations of young women.[37] Third, young women representation in the top quintile seems to have stopped growing for cohorts 1965-1974 and the representation of women at the top in mid career is no longer higher than in early career (after the 1960 cohort). This suggests that economic progress of women might well reach an asymptote well before parity is attained. A similar figure for P60-80 shows that the fraction of women in the second to top quintile has stopped growing for early career women born after 1958 and is around 0.39 for cohorts 1958-1974. The fraction of women among all early careers is around 0.45 for those cohorts (Kopczuk et al., 2007). The lack of changes in the top two quintiles for young women born after 1965 is striking in light of the continuous and rapid progress of women relative college graduation rates for cohorts 1965

[37]Goldin and Katz (2002) demonstrate that availability of birth control pills for single women, starting in the late 1960s, had strong effects on marital and educational choices of women. The SSA data show that women start gaining with the 1941 cohort suggesting that factors happening earlier than the pill for single women also likely had a positive impact on women's earnings.

to 1975 (Goldin et al., 2006) and suggests that the reversal of the education gender gap might not be sufficient to close entirely the earnings gender gap.

7 Conclusion and Future Work

Our paper has used U.S. Social Security earnings administrative data to construct series of inequality and mobility in the United States since 1937. The analysis of these data has allowed us to start exploring the evolution of mobility and inequality over a full career as well as complement the more standard analysis of annual inequality and short term mobility in several ways. We found that changes in mobility have not substantially affected the evolution of inequality, so that annual snapshots of the distribution provide a good approximation of the evolution of the longer term measures of inequality.

However, our key finding is that while the overall measures of mobility are fairly stable, they hide heterogeneity by gender groups. Inequality and opportunity among *male* workers has worsened along almost any dimension since the 1950s: our series display sharp increases in annual earnings inequality, slight reductions in short-term mobility, large increases in long-term career wide inequality with slight reduction or stability of long-term mobility. Against those developments stand the very large earning gains achieved by women since the 1950s, due to increases in labor force attachment as well as increases in earnings conditional on working. Those gains have been so great that they more than compensate for the increase in inequality for males when focusing on the bottom of the distribution.

Thus, the weakening of social norms and labor market institutions inherited from the post-war years which favored low skilled white male workers[38] at the expense of women, Blacks, and top talent has had two important and conflicting consequences for earnings inequality in recent decades. It has allowed women to close a large part of the gender gap, hence improving the position of low earners (especially from a lifetime and upward mobility perspective). However, it may have also strengthened pure market forces which have contributed to increasing sharply the pay of top earners in the US economy.

[38] Levy and Temin (2007) describe the earlier set of institutions as the "Treaty of Detroit" and characterize it by strong unions, very progressive taxes, and high minimum wages, and argue that those institutions have been replaced by the "Washington consensus" favoring free markets and deregulation.

References

Albrecht, James, Anders Bjorklund, and Susan Vroman, "Is There a Glass Ceiling in Sweden?," *Journal of Labor Economics*, January 2003, *21* (1), 145–77.

Attanasio, Orazio, Erich Battistin, and Hidehiko Ichimura, "What Really Happened to Consumption Inequality in the US?," in Ernst Berndt and Charles Hulten, eds., *Measurement Issues in Economics - The Paths Ahead. Essays in Honor of Zvi Griliches*, Chicago: University of Chicago Press, 2007.

Autor, David, Lawrence F. Katz, and Melissa Schettini Kearney, "Trends in U.S. Wage Inequality: Re-Assessing the Revisionists," *Review of Economics and Statistics*, 2007.

Blau, Francine D., "Trends in the Well-being of American Women, 1970-1995," *Journal of Economic Literature*, March 1998, *36* (1), 112–165.

_ **and Lawrence M. Kahn**, "Swimming Upstream: Trends in the Gender Wage Differential in 1980s," *Journal of Labor Economics*, 1997, *15* (1), 1–42.

_ **and** _ , "The Gender Pay Gap: Going, Going ... But Not Gone," in Francine D. Blau, Mary C. Brinton, and David B. Grusky, eds., *The Declining Significance of Gender?*, New York: Rusell Sage Foundation, 2006, 37–66.

_ **, Marianne Ferber, and Anne Winkler**, *The Economics of Women, Men and Work*, 4th ed., Prentice-Hall, 2006.

Borjas, George J., "The Economic Analysis of Immigration," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Amsterdam; New York: Elsevier/North Holland, 1999, 1697–1760.

Buchinsky, Moshe and Jennifer Hunt, "Wage Mobility in the United States," *Review of Economics and Statistics*, August 1999, *81*, 351–68.

Card, David and Alan B. Krueger, "Trends in Relative Black-White Earnings Revisited," *American Economic Review Papers and Proceedings*, May 1993, *83* (2), 85–91.

_ **and Thomas Lemieux**, "Changing Wage Structure and Black-White Wage Differentials," *American Economic Review Papers and Proceedings*, 1994, *84* (2), 29–33.

Carroll, Robert, David Joulfaian, and Mark Rider, "Income Mobility: The Recent American Experience," Working Paper 07-18, Andrew Young School of Policy Studies, Georgia State University March 2007.

Congressional Budget Office, "Trends in Earnings Variability Over the Past 20 Years," Letter to the Honorable Charles E. Schumer and the Honorable Jim Webb April 2007. Online at `http://www.cbo.gov/ftpdocs/80xx/doc8007/04-17-EarningsVariability.pdf`.

Cutler, David and Lawrence Katz, "Macroeconomic Performance and the Disadvantaged," *Brookings Papers on Economic Activity*, 1991, *2*, 1–74.

Donohue, John J. and James Heckman, "Continuous Versus Episodic Change: The Impact of Civil Rights Policy on the Economic Status of Blacks," *Journal of Economic Literature*, December 1991, *29* (4), 1603–1643.

Edlund, Lena and Wojciech Kopczuk, "Women, Wealth and Mobility," Working Paper 13162, National Bureau of Economic Research June 2007.

Fortin, Nicole M. and Thomas Lemieux, "Rank Regressions, Wage Distributions, and the Gender Gap," *Journal of Human Resources*, Summer 1998, *33* (3), 610–643.

Goldin, Claudia, *Understanding the gender gap: An economic history of American women* NBER Series on Long-Term Factors in Economic Development, New York; Oxford and Melbourne: Oxford University Press, 1990.

_ , "The Role of World War II in the Rise of Women's Employment," *American Economic Review*, September 1991, *81* (4), 741–56.

_ , "From the Valley to the Summit: The Quiet Revolution that Transformed Women's Work," Working Paper 10035, National Bureau of Economic Research March 2004.

_ , "The Quiet Revolution That Transformed Women's Employment, Education, and Family," *American Economic Review Papers and Proceedings*, May 2006, *96* (2), 1–21.

_ , "The Rising (and then Declining) Significance of Gender," in Francine D. Blau, Mary C. Brinton, and David B. Grusky, eds., *The Declining Significance of Gender?*, New York: Rusell Sage Foundation, 2006, 67–101.

_ **and Lawrence F. Katz**, "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions," *Journal of Political Economy*, August 2002, *110* (4), 730–770.

_ **and Robert A. Margo**, "The Great Compression: The Wage Structure in the United States at Mid-Century," *Quarterly Journal of Economics*, February 1992, *107* (1), 1–34.

_ **, Lawrence F. Katz, and Ilyana Kuziemko**, "The Homecoming of American Women: The Reversal of the College Gender Gap," *Journal of Economic Perspectives*, Fall 2006, *20* (4), 133–156.

Gottschalk, Peter, "Inequality, Income Growth, and Mobility: The Basic Facts," *Journal of Economic Perspectives*, Spring 1997, *11* (2), 21–40.

Hacker, Jacob S, *The Great Risk Shift: The Assault on American Jobs, Families Health Care, and Retirement - And How You Can Fight Back*, Oxford University Press, 2006.

Hanson, Gordon H., "Illegal Immigration from Mexico to the United States," *Journal of Economic Literature*, December 2006, *44* (4), 869–924.

Hungerford, Thomas L, "U.S. Income Mobility in the Seventies and Eighties," *Review of Income and Wealth*, December 1993, *39* (4), 403–417.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce, "Accounting for the Slowdown in Black-White Wage Convergence," in Marvin Kosters, ed., *Workers and Their Wages*, AEI Press, 1991, 107–43.

Katz, Lawrence F. and David Autor, "Changes in the Wage Structure and Earnings Inequality," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Amsterdam; New York: Elsevier/North Holland, 1999.

_ **and Kevin M. Murphy**, "Changes in Relative Wages, 1963-87: Supply and Demand Factors," *Quarterly Journal of Economics*, February 1992, *107* (1), 35–78.

Kennickell, Arthur B., "Currents and Undercurrents: Changes in the Distribution of Wealth, 1989-2004," Discussion Paper 2006-13, Federal Reserve Board 2006.

Kestenbaum, Bert, "Evaluating SSA's Current Procedure for Estimating Untaxed Wages," *American Statistical Association Proceedings of the Social Statistics Section*, 1976, *Part 2*, 461–465.

Kopczuk, Wojciech and Emmanuel Saez, "Top Wealth Shares in the United States, 1916-2000: Evidence from Estate Tax Returns," *National Tax Journal*, June 2004, *57* (2 (part 2)), 445–488.

_ , _ , and Jae Song, "Uncovering the American Dream: Inequality and Mobility in Social Security Earnings Data since 1937," Working Paper 13345, National Bureau of Economic Research August 2007.

Krueger, Dirk and Fabrizio Perri, "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory," *Review of Economic Studies*, January 2006, *73* (1), 163–93.

Levy, Frank and Peter Temin, "Inequality and Institutions in 20th Century America," Working Paper 13106, National Bureau of Economic Research May 2007.

Lindert, Peter, "Three Centuries of Inequality in Britain and America," in Anthony B. Atkinson and Francois Bourguignon, eds., *Handbook of Income Distribution*, Amsterdam; New York: Elsevier/North Holland, 2000, 167–216.

Margo, Robert A., "Explaining Black-White Wage Convergence, 1940-1950," *Industrial and Labor Relations Review*, April 1995, *48* (3), 470–481.

Moffitt, Robert A. and Peter Gottschalk, "Trends in the Autocovariance Structure of Earnings in the U.S.: 1969-1987," 1995. Johns Hopkins University, mimeo.

Panis, Constantijn, Roald Euller, Cynthia Grant, Melissa Bradley, Christine E. Peterson, Randall Hirscher, and Paul Steinberg, *SSA Program Data User's Manual* RAND June 2000. Prepared for the Social Security Administration.

Piketty, Thomas and Emmanuel Saez, "Income Inequality in the United States, 1913-1998," *Quarterly Journal of Economics*, February 2003, *118*, 1–39.

Scholz, John Karl, "Wealth Inequality and the Wealth of Cohorts," 2003. University of Wisconsin, mimeo.

Slesnick, Daniel T., *Consumption and Social Welfare: Living Standards and Their Distribution in the United States*, Cambridge, New York and Melbourne: Cambridge University Press, 2001.

Smith, James and Finis Welch, "Black Economic Progress After Myrdal," *Journal of Economic Literature*, June 1989, *27* (2), 519–564.

Social Security Administration, *Handbook of Old-Age and Survivors Insurance Statistics* (annual), Washington, D.C.: US Government Printing Office, 1937-1952.

_ , *Social Security Bulletin: Annual Statistical Supplement*, Washington, DC: Government Printing Press Office, 1967.

Solon, Gary, "Intergenerational Mobility in the Labor Market," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3A, Amsterdam; New York: Elsevier/North Holland, 1999, 1761–1800.

Figures footnotes

Figure 0: Sample is the core sample defined as all employees in Commerce and Industry with earnings above minimum threshold ($2,575 in 2004 and indexed using average wage for earlier years) and aged 18 to 70 (by January 1st of a given year t). Commerce and Industry is defined as all industrial sectors excluding government employees, agriculture, hospitals, educational services, social services, religious and membership organizations, and private households. Only commerce and industry earnings are included. Self-employment earnings are fully excluded. See Kopczuk et al. (2007) for complete details.

Figure 1: Sample for all workers is the core sample (see Figure 0 footnote). White is defined as all workers excluding Blacks.

Figure 2: Sample is the core sample (see Figure 0 footnote). 3, 5, and 11 year averages earnings shares in year t are estimated for individuals with earnings above the minimum threshold in the middle year t (earnings can be zero in other years) and alive and aged 18 to 70 in the 3, 5, or 11 year window. Earnings over the 3, 5, or 11 year window are averaged using the average wage index.

Figure 3: Sample is core sample (see Figure 0). Probabilities of staying in top 0.1% group in Panel A are *conditional* on being in the top 0.1% group in base year and staying in the core sample after 1, 3, 5, or 10 years. Panel B displays the location in year $t-10$ of top 1% core sample earners in year t *conditional* on being in the core sample in year $t-10$.

Figure 4: Sample is core sample (see Figure 0). Panel A displays the probabilities of staying in group P0-40 (P60-100) after 1 year *conditional* on being in group P0-40 (P60-100) in base year and staying in core sample after 1 year. Panel B displays the probabilities of moving from P60-100 to P0-40 (P0-40 to P60-100) after 1 year *conditional* on being in group P60-100 (P0-40) in base year and staying in core sample after 1 year. Shaded areas denote recession months defined using NBER dates (yearly tick marks correspond to the middle of the year). Because of small sample size, series before 1957 are smoothed using a weighted 3-year moving average with weight of .5 for cohort t and weights of .25 for $t-1$ and $t+1$ (sample size of SSA data is 0.1% up to 1956 and 1% afterwards).

Figure 5: Sample in year t is core sample (see Figure 0) with additional restriction that individual is alive and aged 18 to 70 in the 11 window centered around middle year t. Panel A (Panel B) displays the probability of moving from P0-40 (P80-100) to P80-100 (P0-40) after 10 (15, 20) years *conditional* on being in group P0-40 (P80-100) in base year and staying in core sample (with additional restriction as in base year) after 10 (15, 20) years. Percentile groups in year t $(t+10, t+15, t+20)$ are defined based on 11 year earnings averages centered around middle year t $(t+10, t+15, t+20)$ and using average wage index and conditional on earnings in middle year being above the minimum threshold and conditional on being alive and aged 18 to 70 in the 11 window around middle year. Because of small sample size, series including earnings before 1957 are smoothed using a weighted 3-year moving average with weight of .5 for cohort t and weights of .25 for $t-1$ and $t+1$.

Figure 6: Sample is career sample defined as follows for each career stage and birth cohort: all employees with average Commerce and Industry earnings (using average wage index) over the 12-year career stage above the minimum threshold ($2,575 in 2004 and indexed on average wage for earlier years). Note that earnings can be zero for some years. Quintiles are then defined within each birth cohort based on average (Commerce and Industry) earnings in each career stage. Probability of moving from quintile q_1 to quintile q_1 from a career stage s_1 to career stage s_2 is conditional on being in the career sample in both career stages s_1 and s_2 and conditional on being in quintile q_1 in stage s_1. Because of small sample size, series including earnings before 1957 are smoothed using a weighted 3-year moving average with weight of .5 for cohort t and weights of .25 for $t-1$ and $t+1$. Estimates in lighter grey are imputed based on less than 12 year of earnings (as the career stage is right-censored in 2004), see Kopczuk et al. (2007) for details.

Figure 7: Sample is career sample for each given stage and birth cohort (see Figure 6). Quintile shares are defined within each cohort and career stage. Smoothing as in Figure 6.

Figure 8: Sample is core sample (see Figure 0). Panels A and B display the fraction of women in various groups. Panel A displays the average earnings ratio of women to men on the right axis. Because of top coding in the micro-data, estimates from 1943 to 1950 for P80-90 and P90-100 are estimated using published tabulations in Social Security Administration

(1937-1952) and Social Security Administration (1967) and reported in lighter grey.

Figure 9: Sample is core sample (see Figure 0). Panels A and B display the fraction of Blacks in various groups (relative to the fraction of Blacks in the adult population aged 18 to 70 estimated from decennial Census data and cubic spline interpolation between census years). Panel A displays the average earnings ratio of Blacks to Whites on the right axis.

Figure 10: Sample is core sample (see Figure 0). The Figure displays the fraction of foreign born workers in various groups

Figure 11: Sample in Panel A is core sample with same restriction and smoothing as in Figure 5. Probabilities for each group (such as women) are also conditional on belonging in the group. Sample in Panel B is career sample (as in Figure 6). Probabilities for each group (such as women) are conditional on being in the group (smoothing and imputations beyond 2004 is as in Figure 6).

Figure 12: Sample is career sample for each given stage and birth cohort (see Figure 6). Quintile shares are defined within each cohort and career stage. Quintile shares for men are computed in the career sample restricted to men (smoothing is as in Figure 6).

Figure 13: Sample is career sample for each given stage and birth cohort (see Figure 6). The figure reports the fraction of women in the top quintile for each career stage and birth cohort groups (smoothing and imputations beyond 2004 as in Figure 6).

Figure 0: Aggregate SSA Earnings and Workers in Commerce and Industry Core Sample

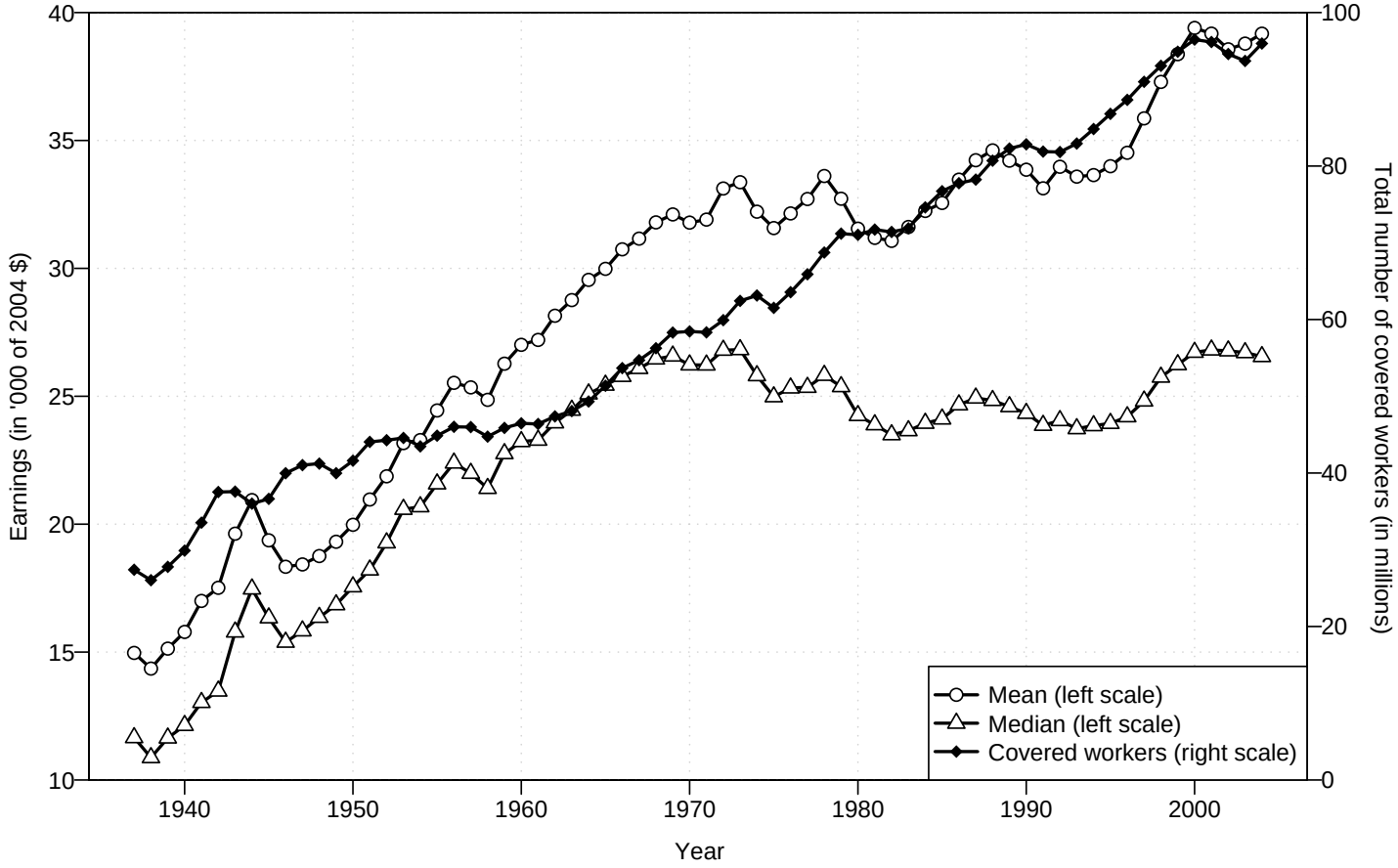


Figure 1: Gini coefficient

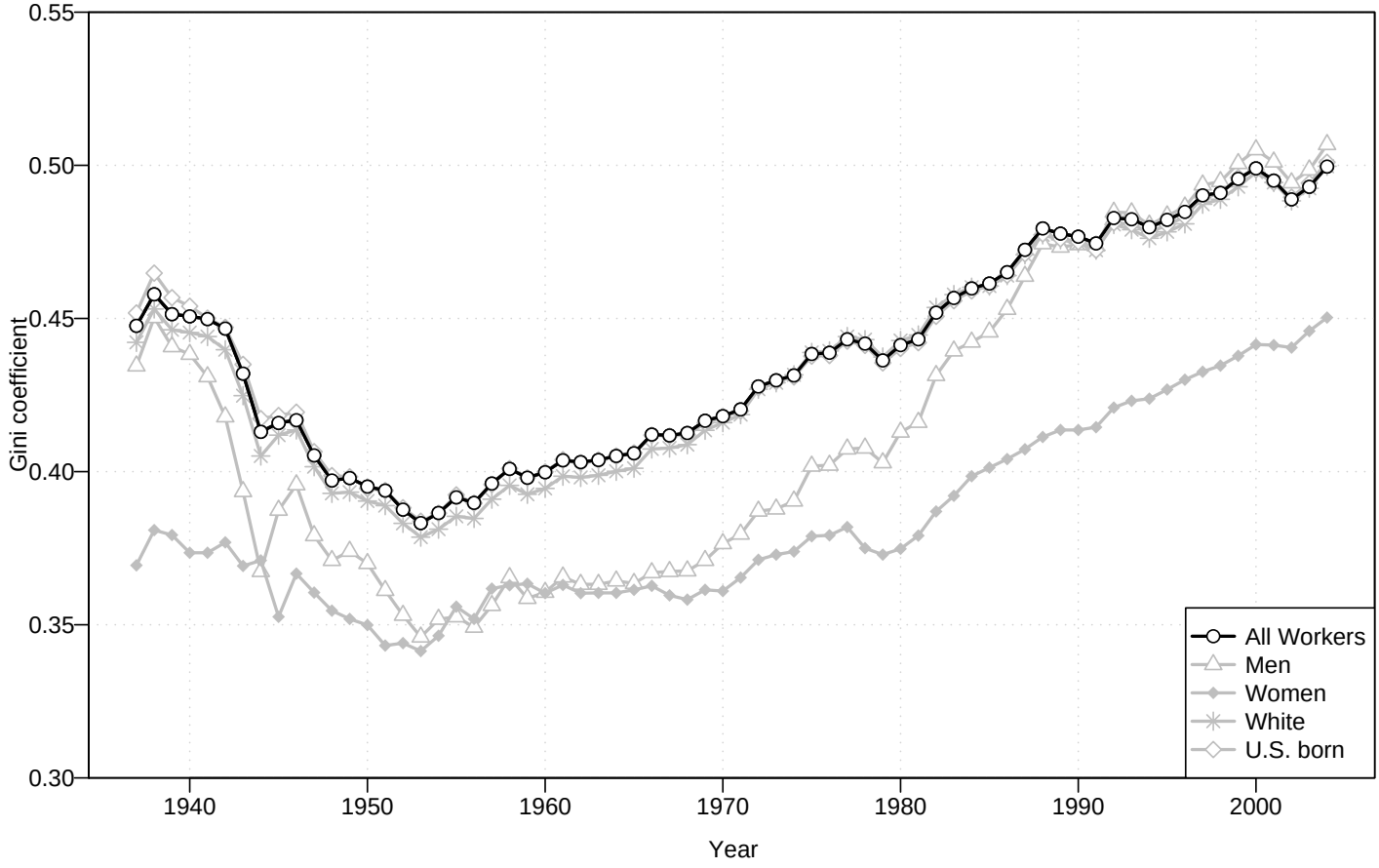


Figure 2A: Bottom Earnings Shares

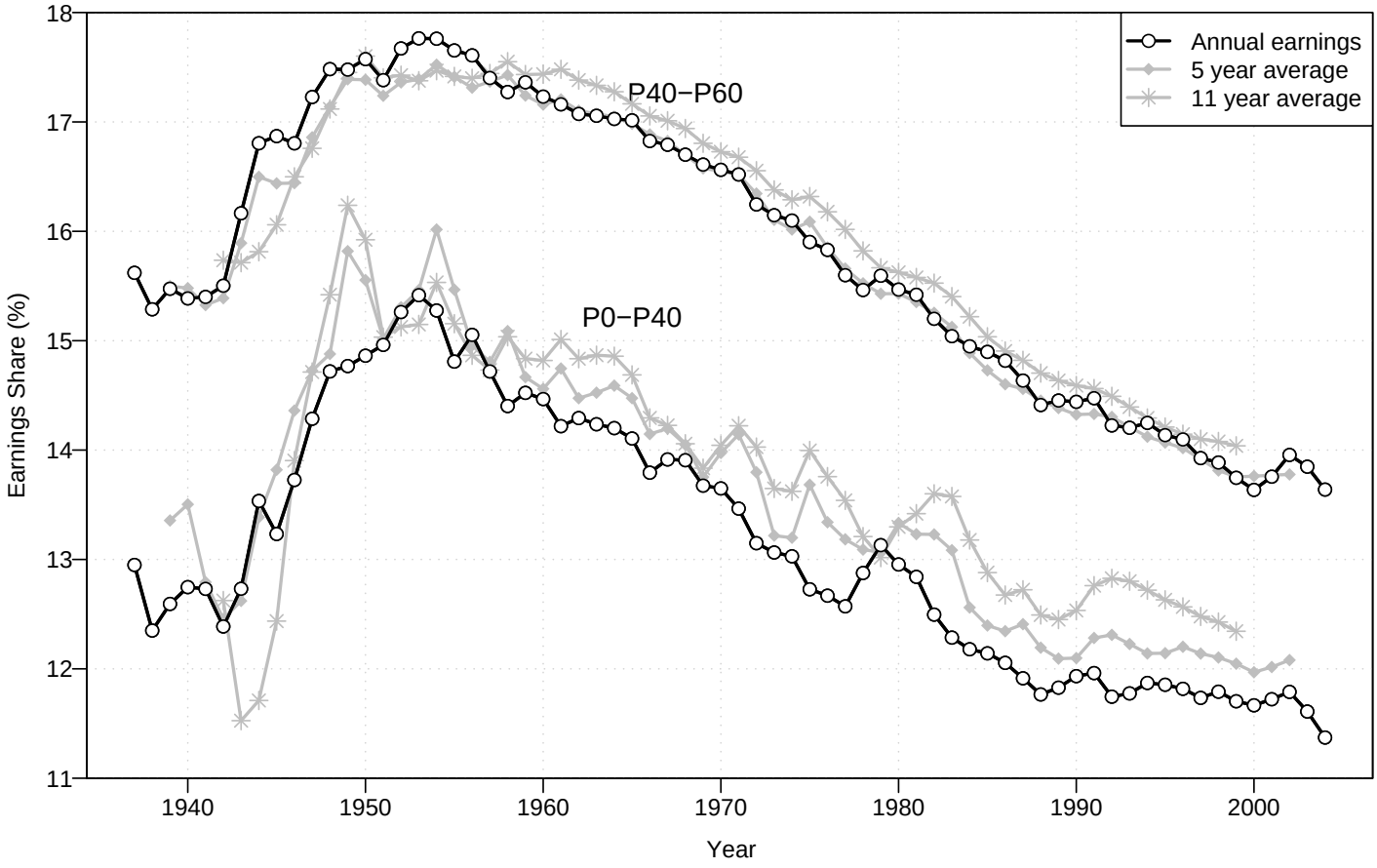
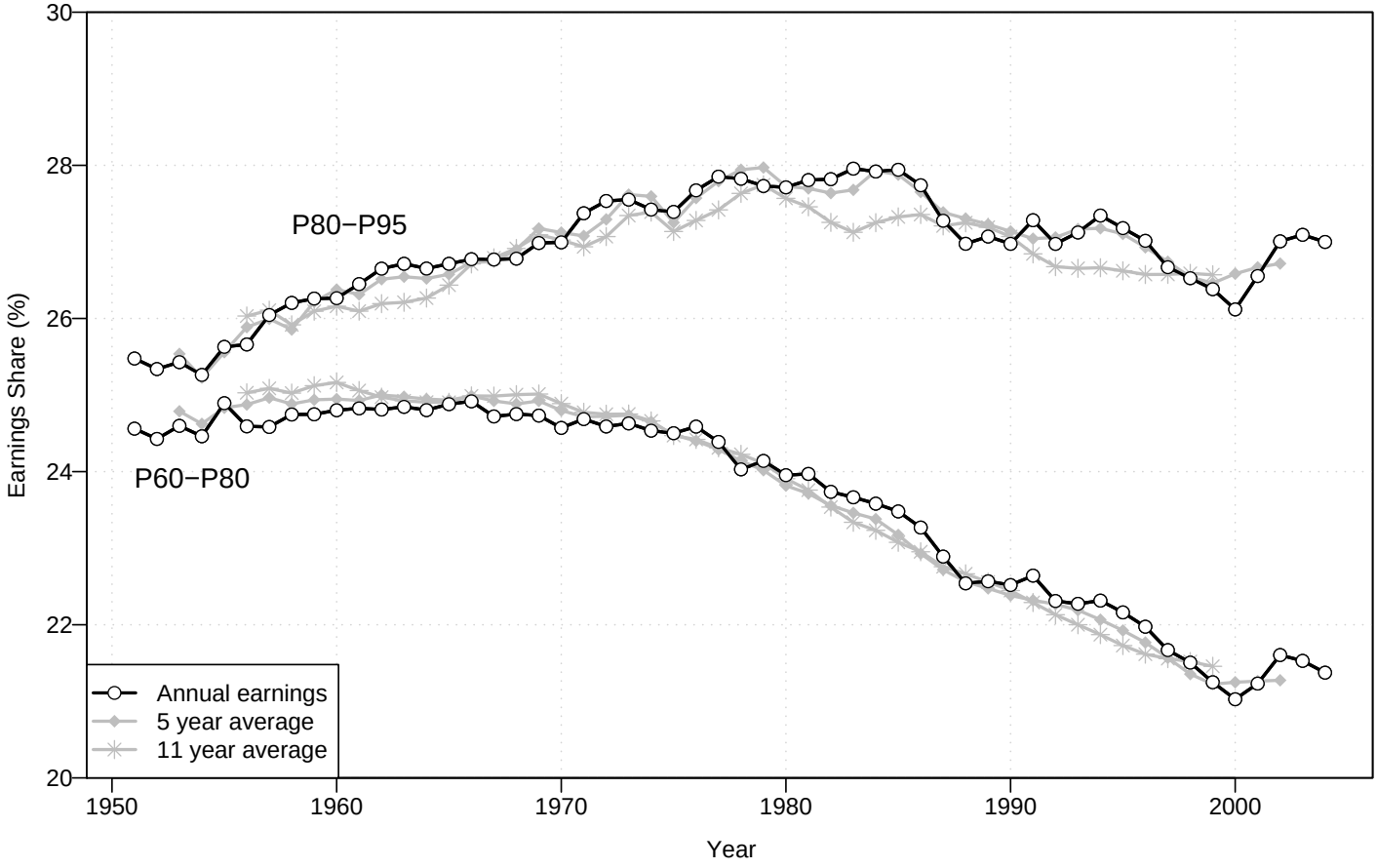


Figure 2B: Upper Earnings Shares



See notes on pages 34-36

Figure 2C: Top Earnings Shares

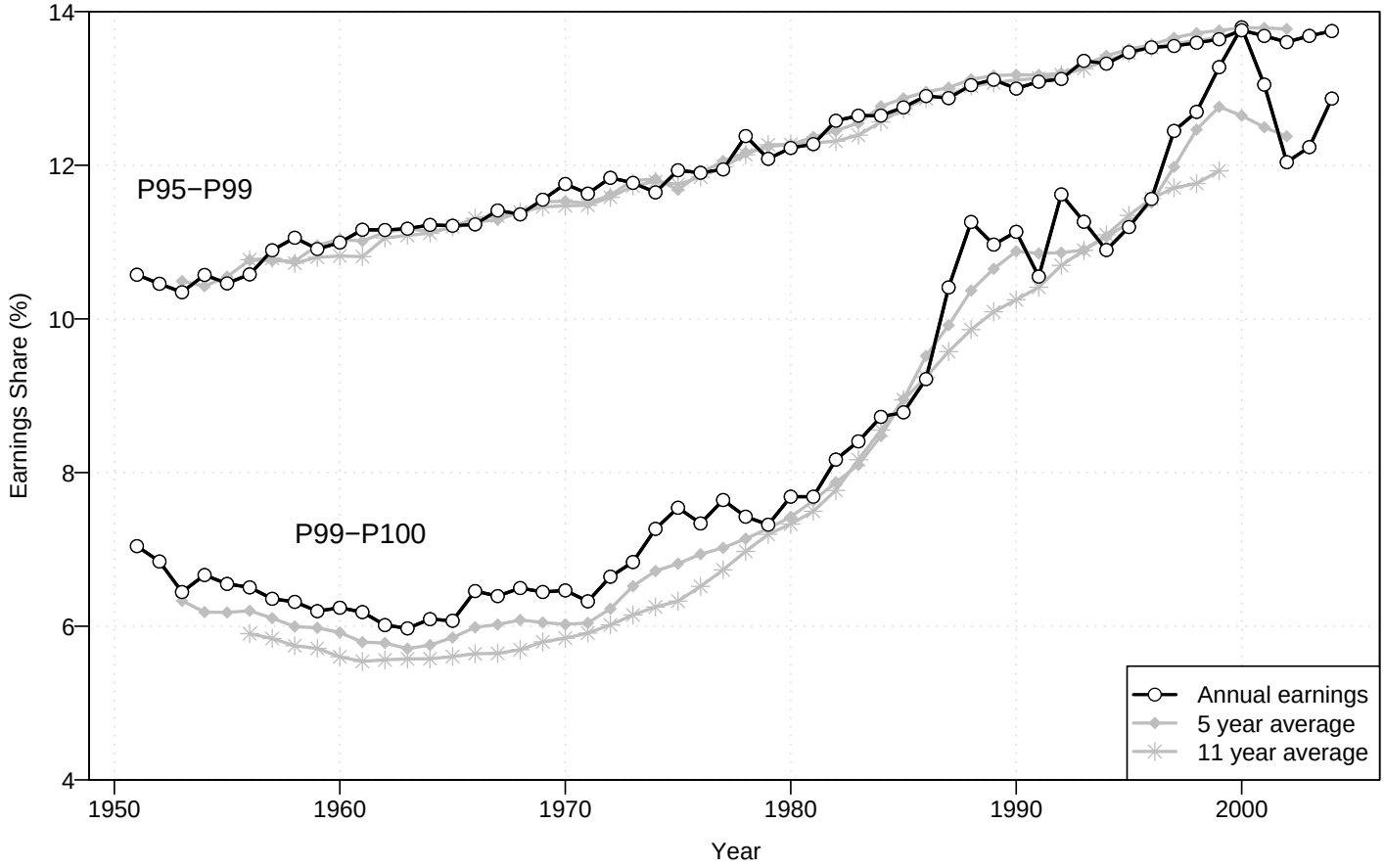
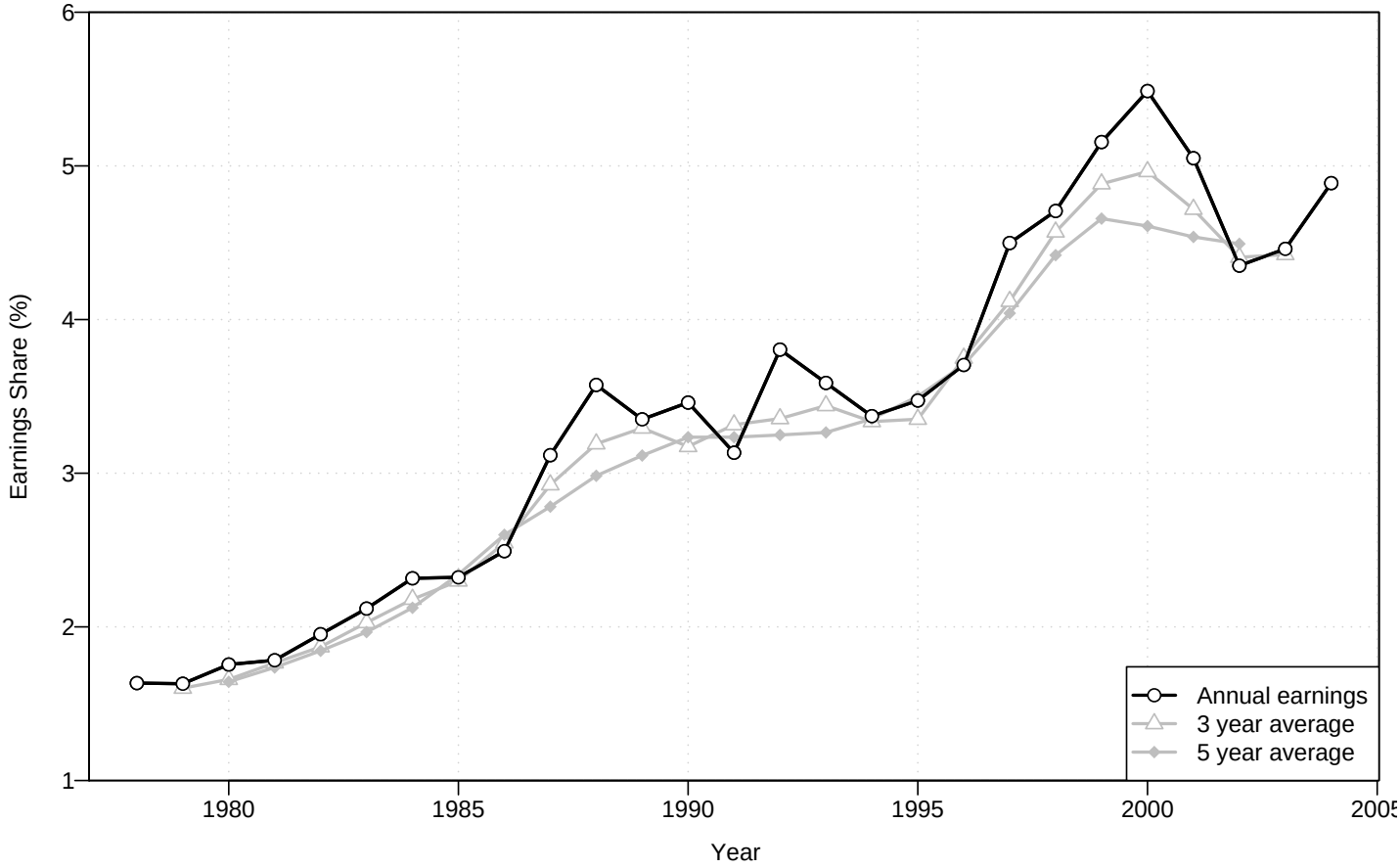


Figure 2D: Top 0.1% Earnings Share



See notes on pages 34-36

Figure 3A: Probability of Staying in Top 0.1%

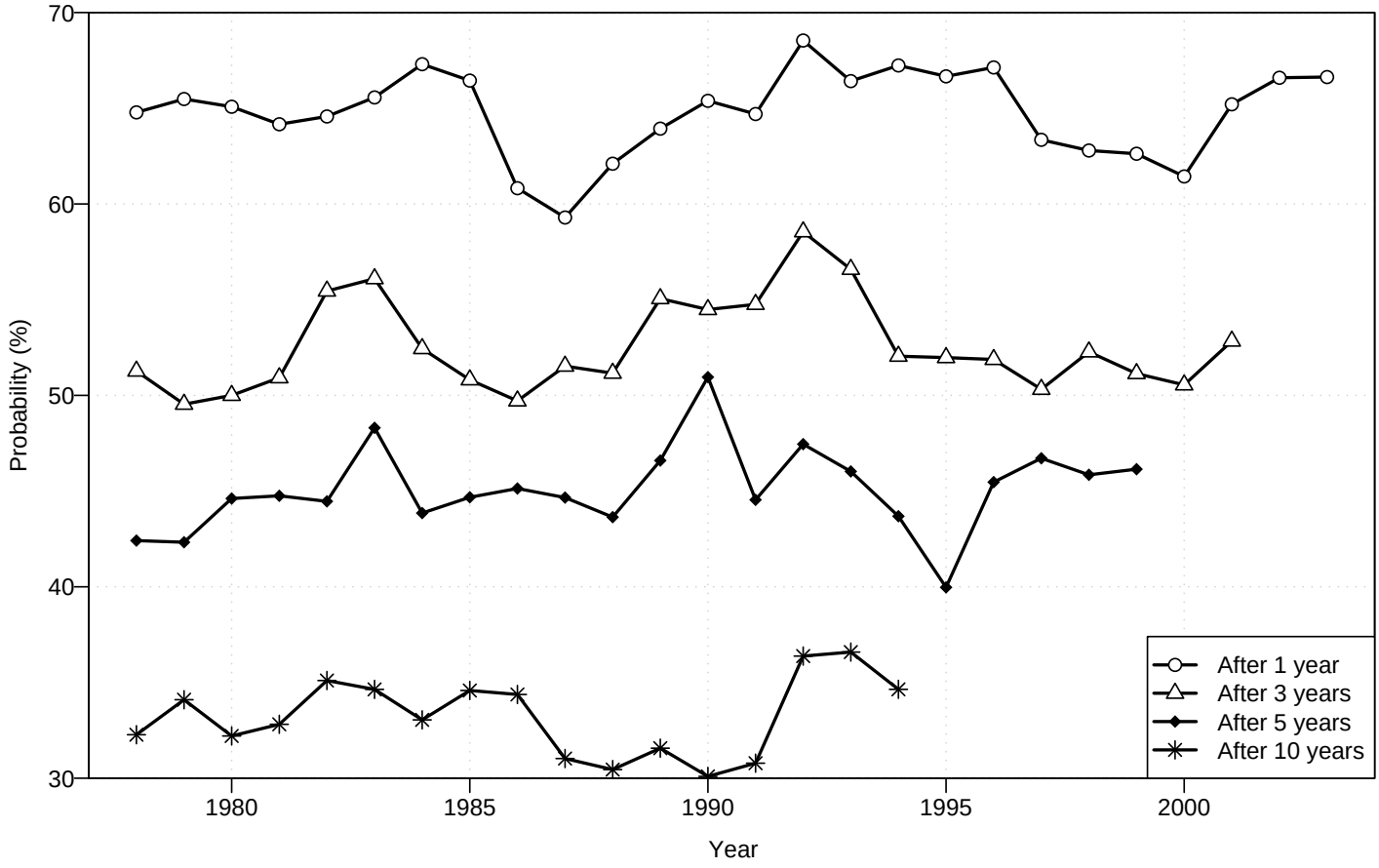
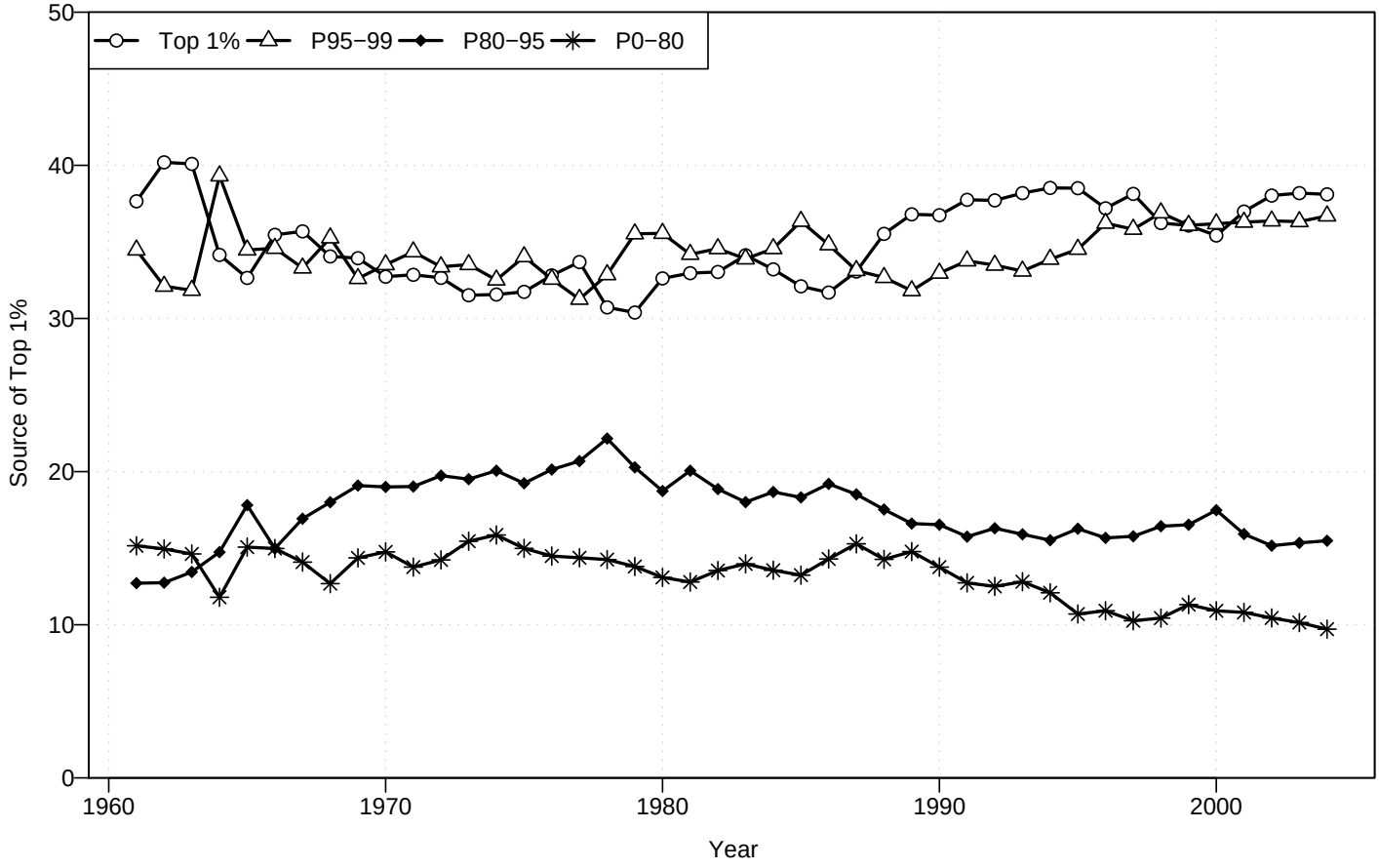


Figure 3B: Source of Top 1% Ten Years Earlier



See notes on pages 34-36

Figure 4A: Probability of Staying in Top and Bottom Groups

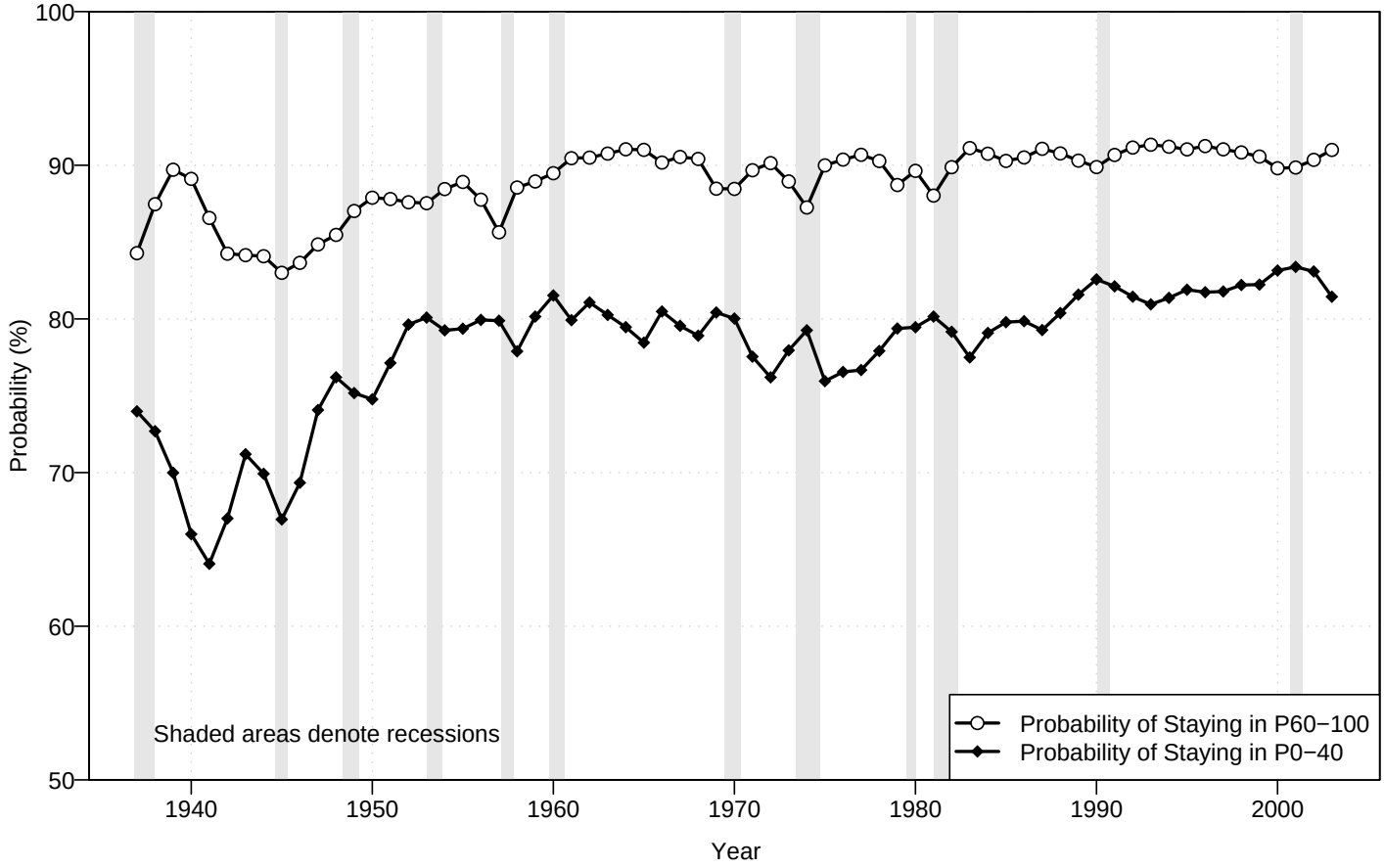
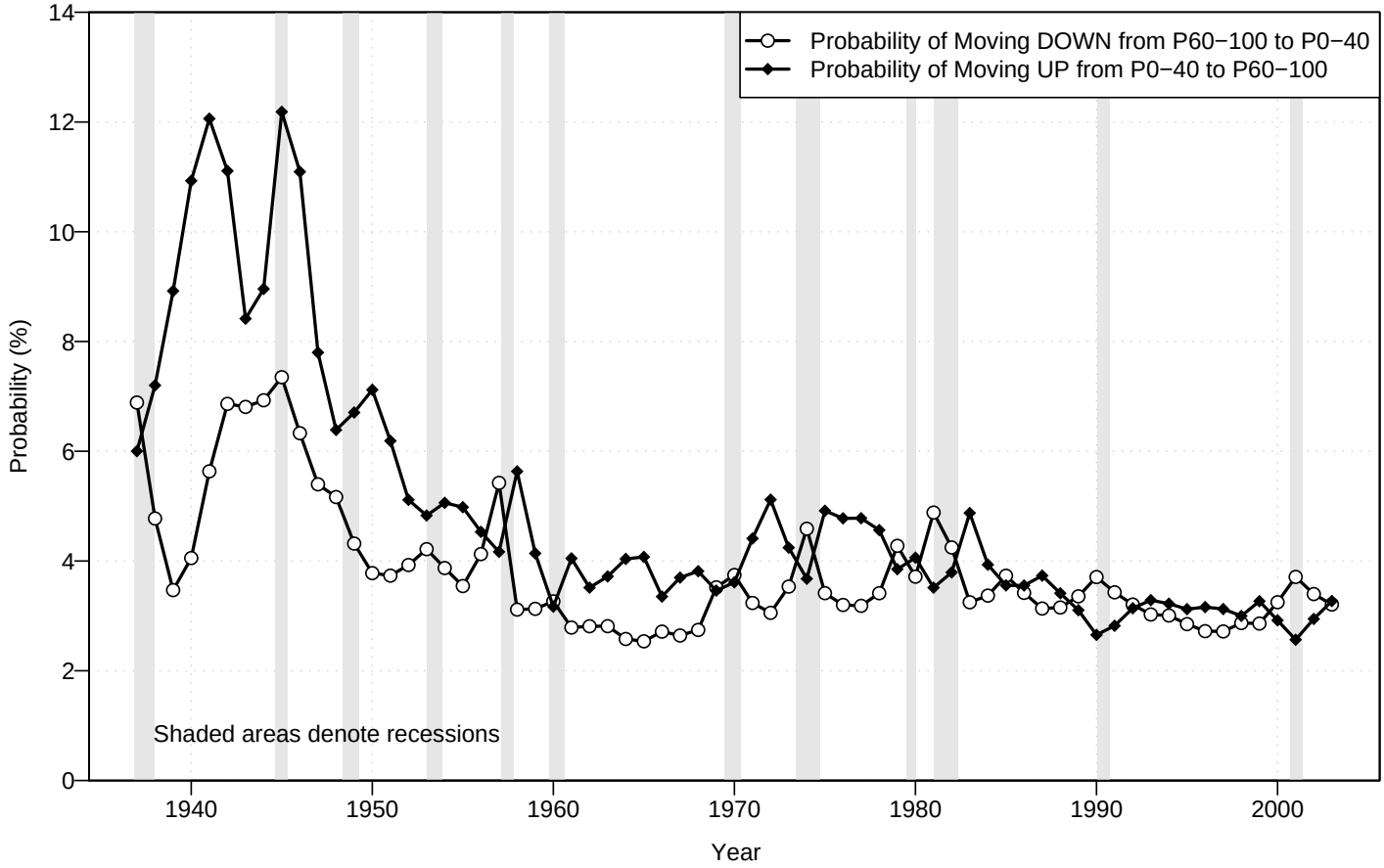


Figure 4B: Downward and Upward Mobility



See notes on pages 34-36

Figure 5A: Probability of Upward Mobility for 11 Year Earnings Average

Figure 5B: Probability of Downward Mobility for 11 Year Earnings Average

See notes on pages 34-36

Figure 6A: Probability of Upward Mobility over a Career

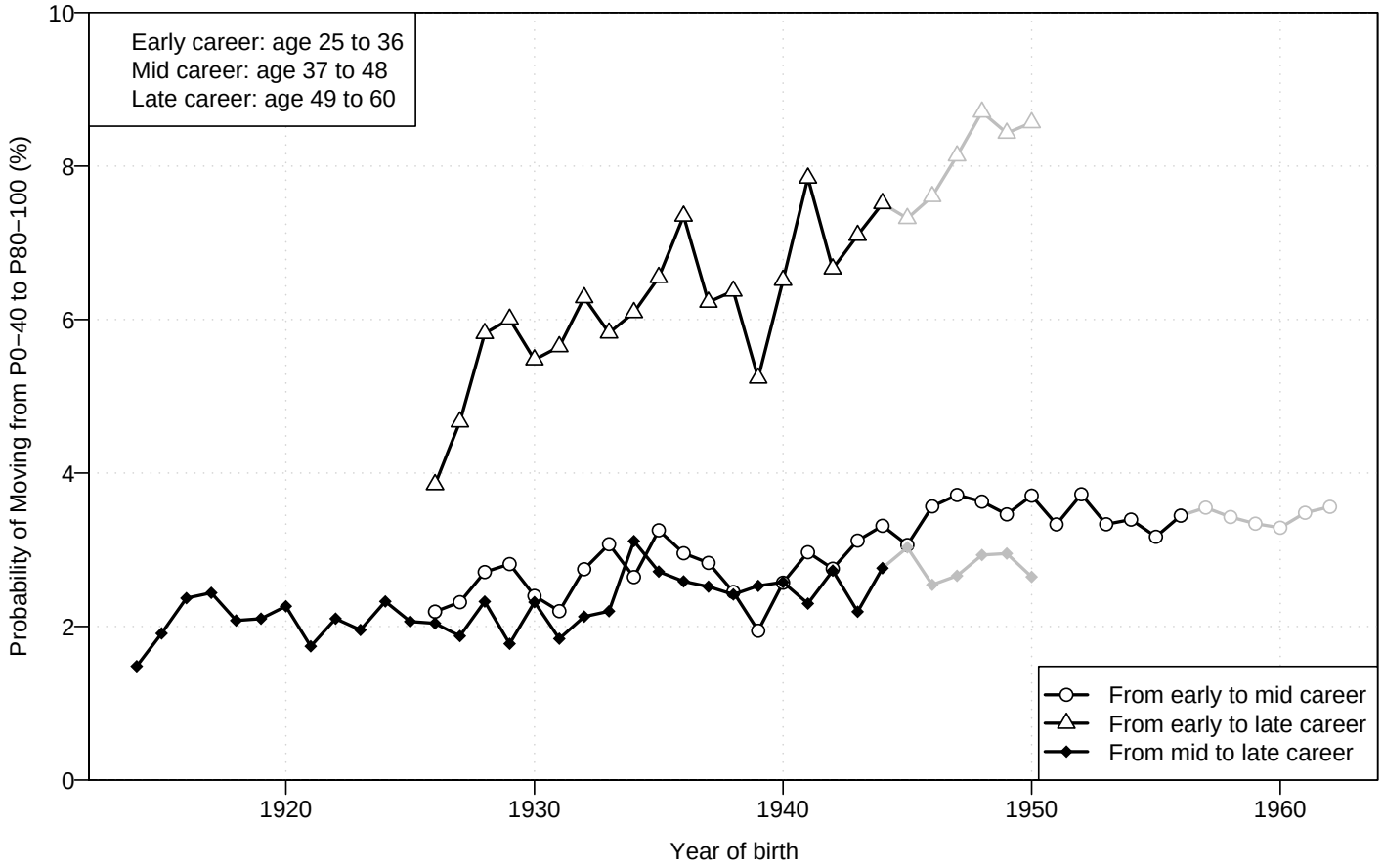
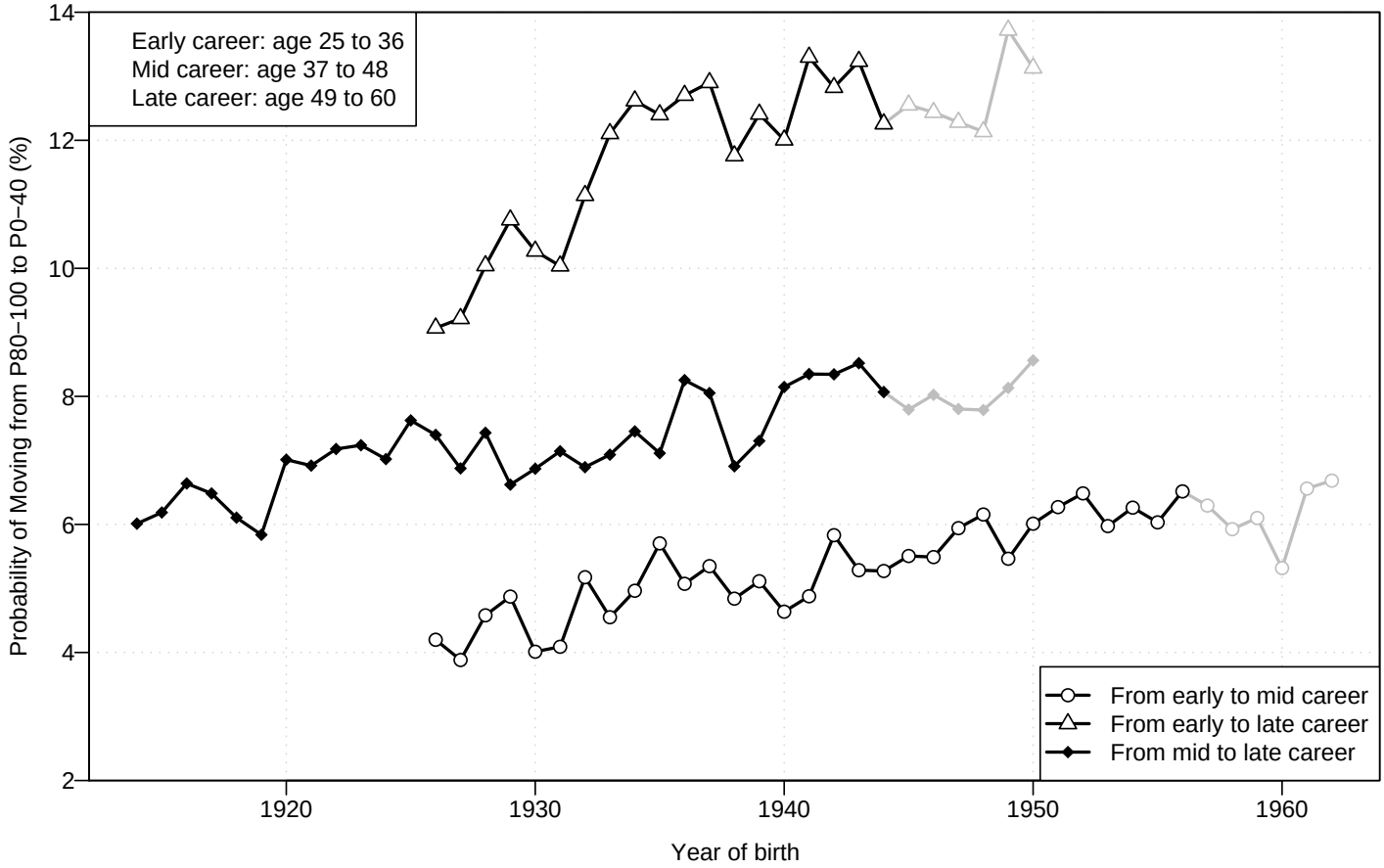


Figure 6B: Probability of Downward Mobility over a Career



See notes on pages 34-36

Figure 7A: Long−Term Top Quintile Share

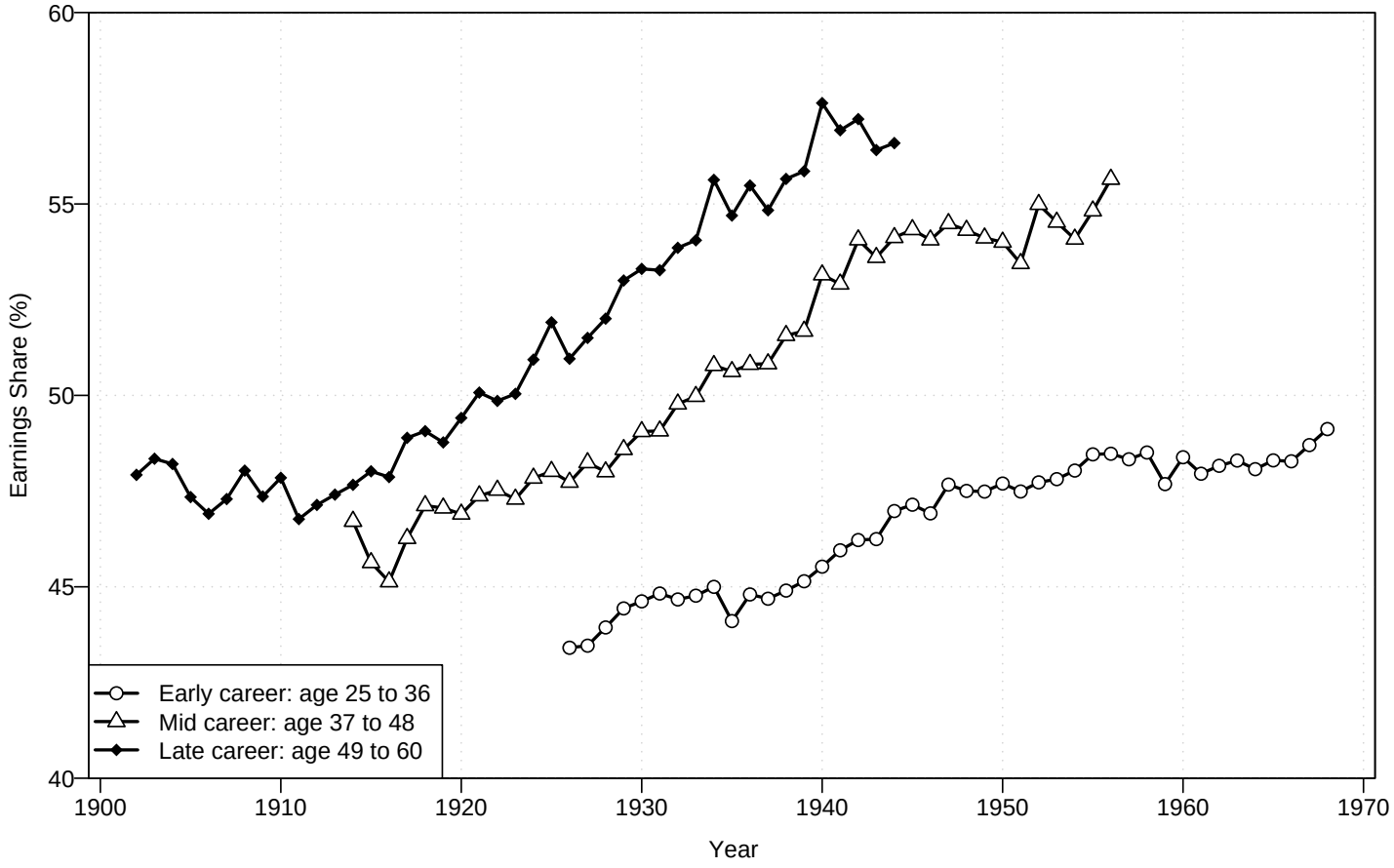


Figure 7B: Long−Term P0−40 Share

See notes on pages 34-36

Figure 8A: Gender Gap in Upper Groups

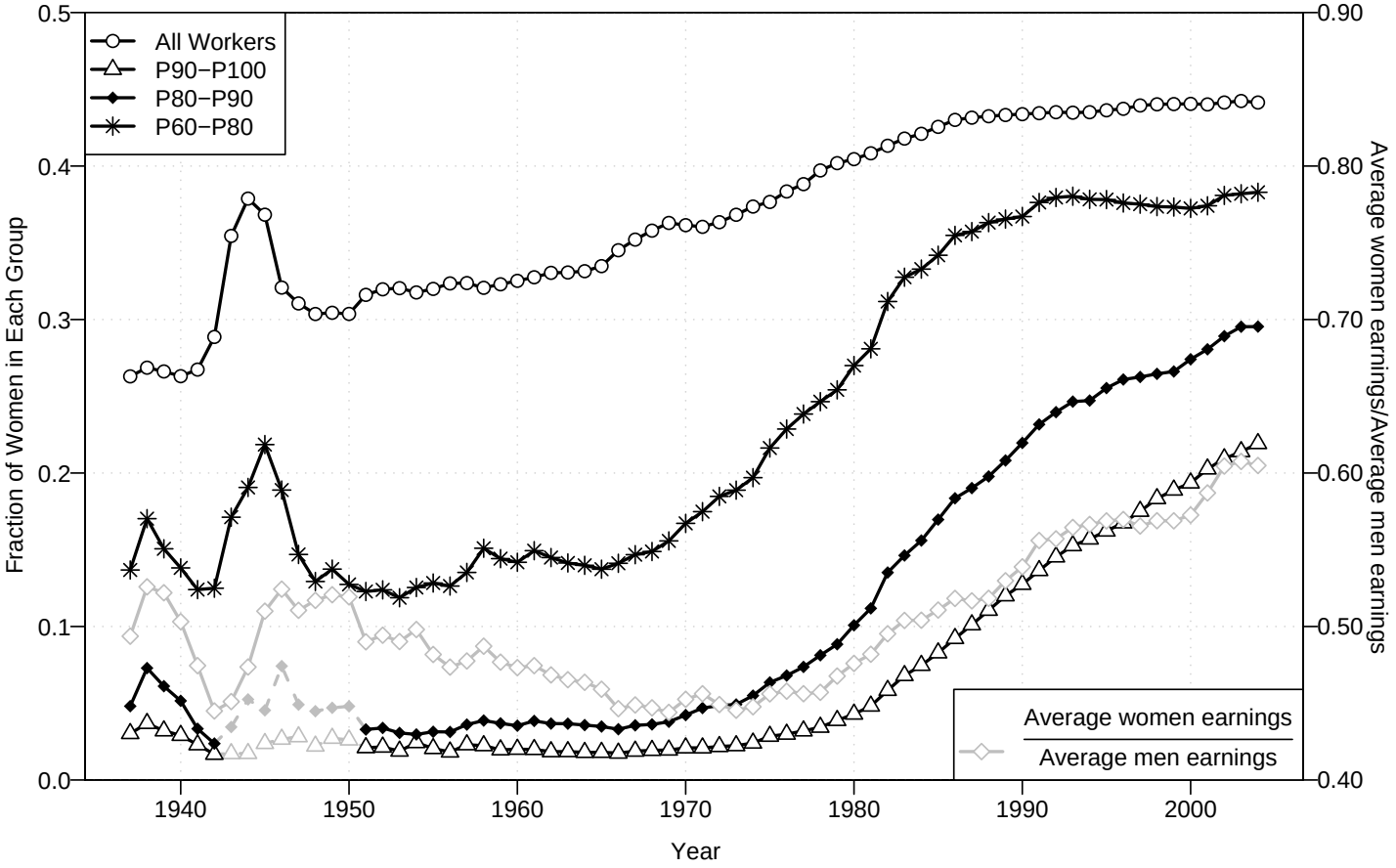
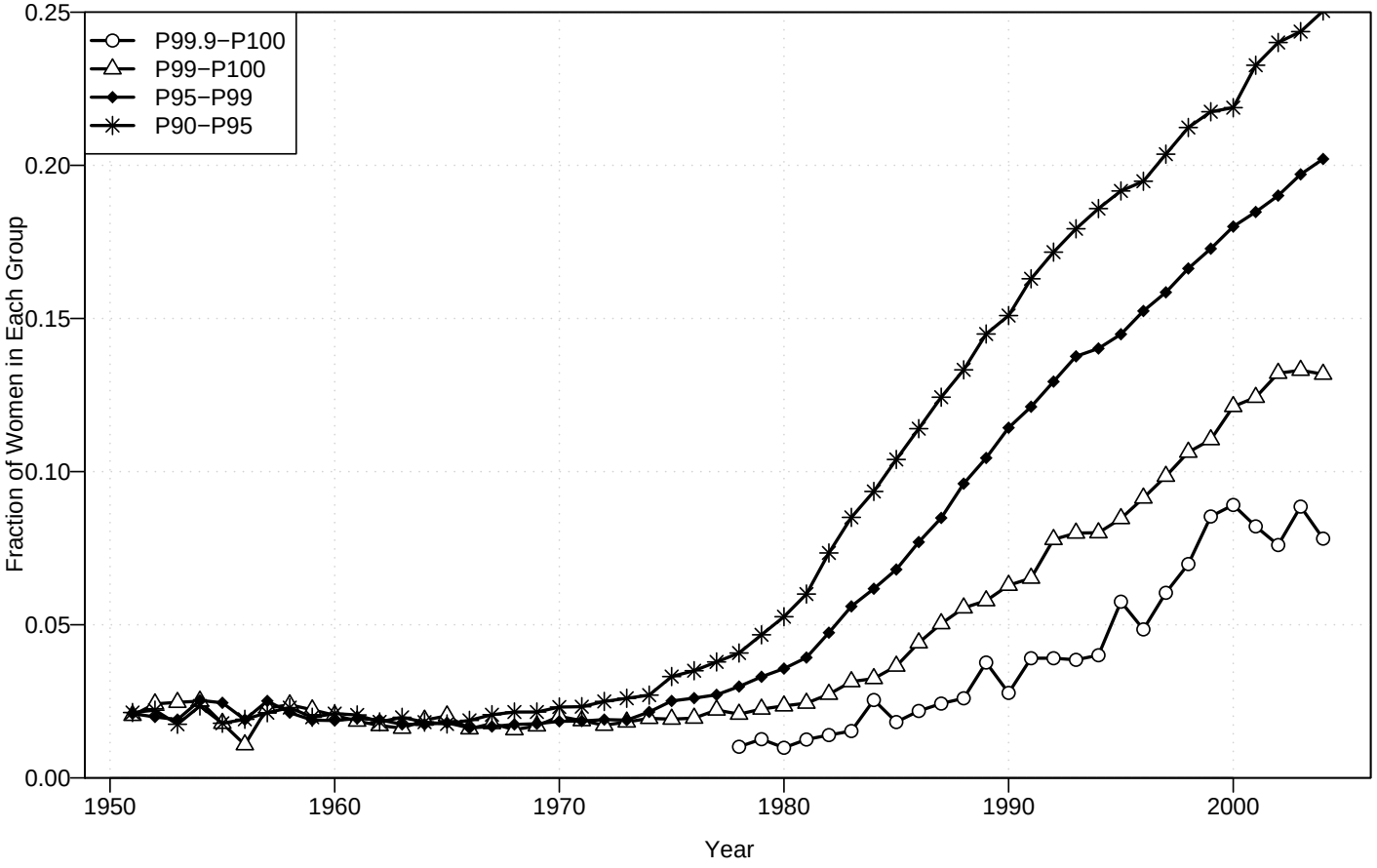


Figure 8B: Gender Gap in Top Groups



See notes on pages 34-36

Figure 9A: Black−White Gap in Upper Groups

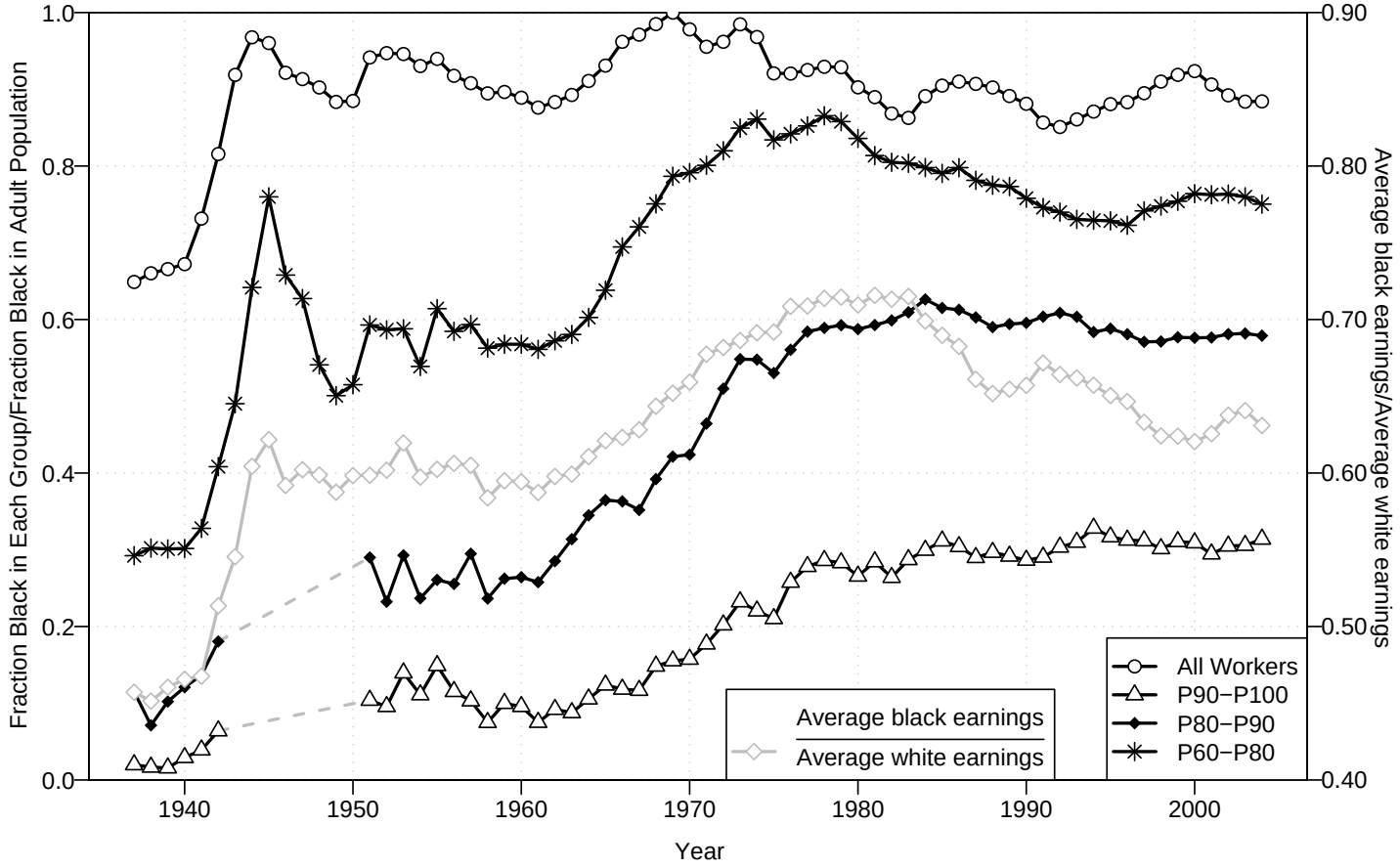


Figure 9B: Black−White Gap in Top Groups

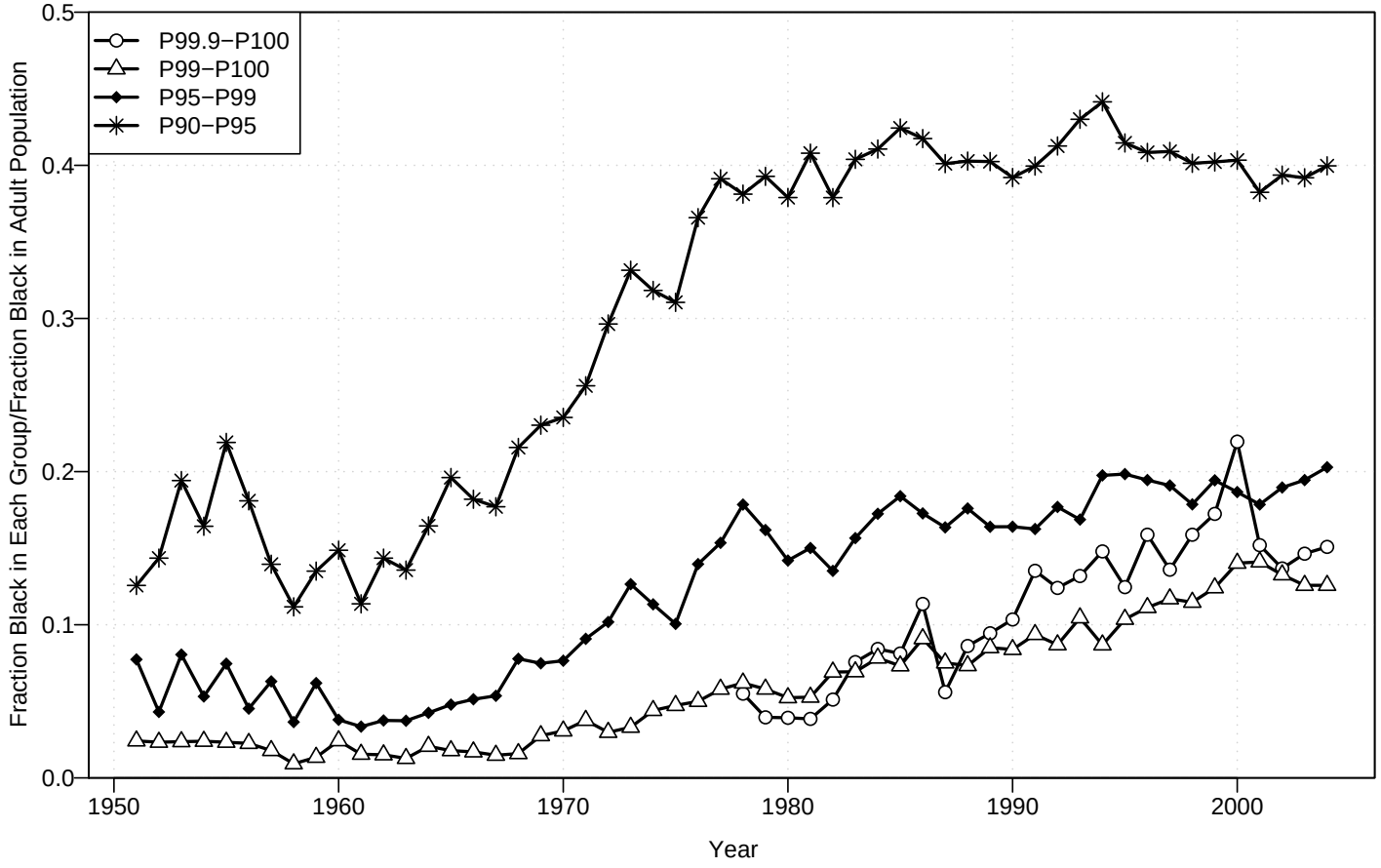


Figure 10: Fraction of Immigrants in Various Groups

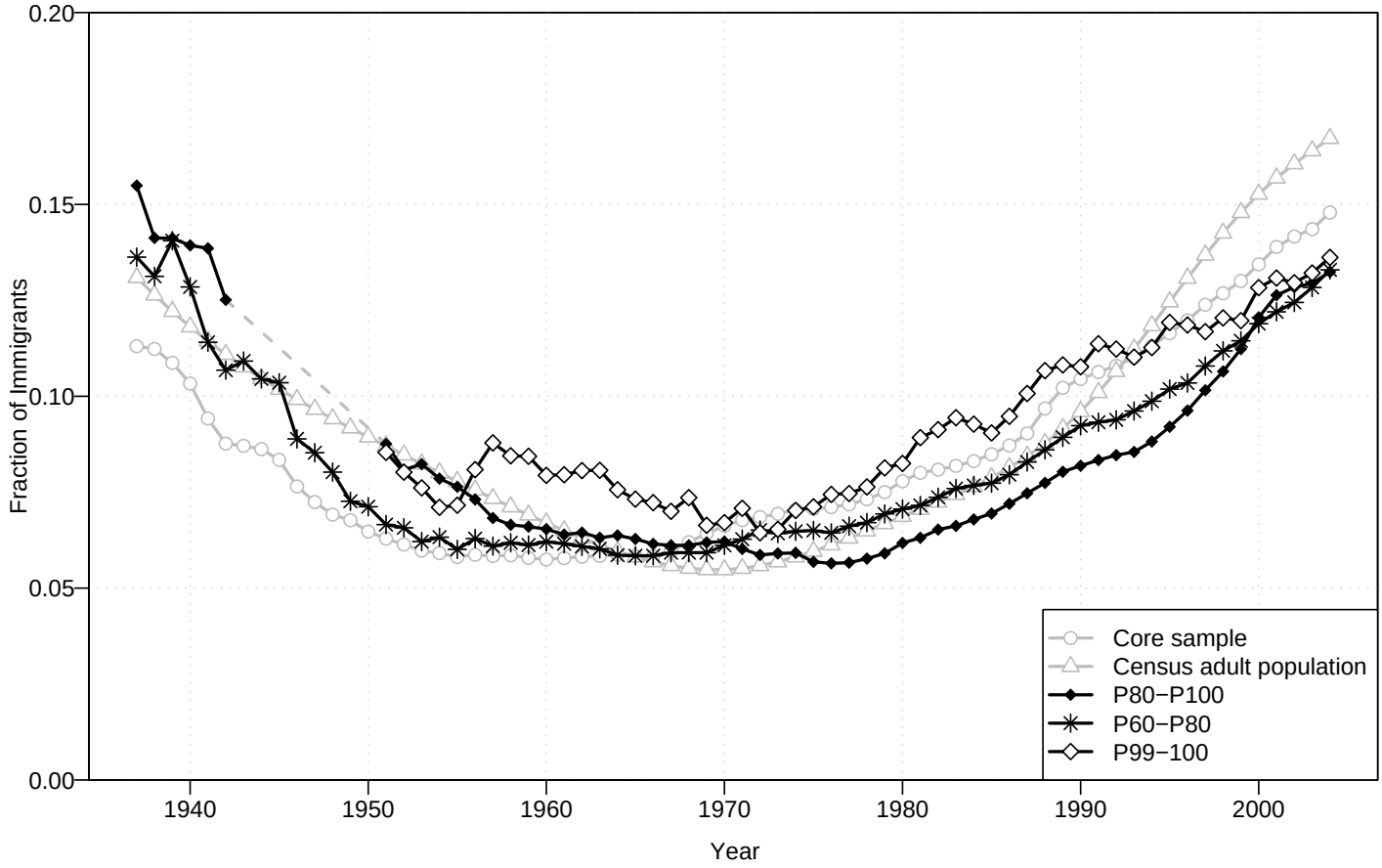
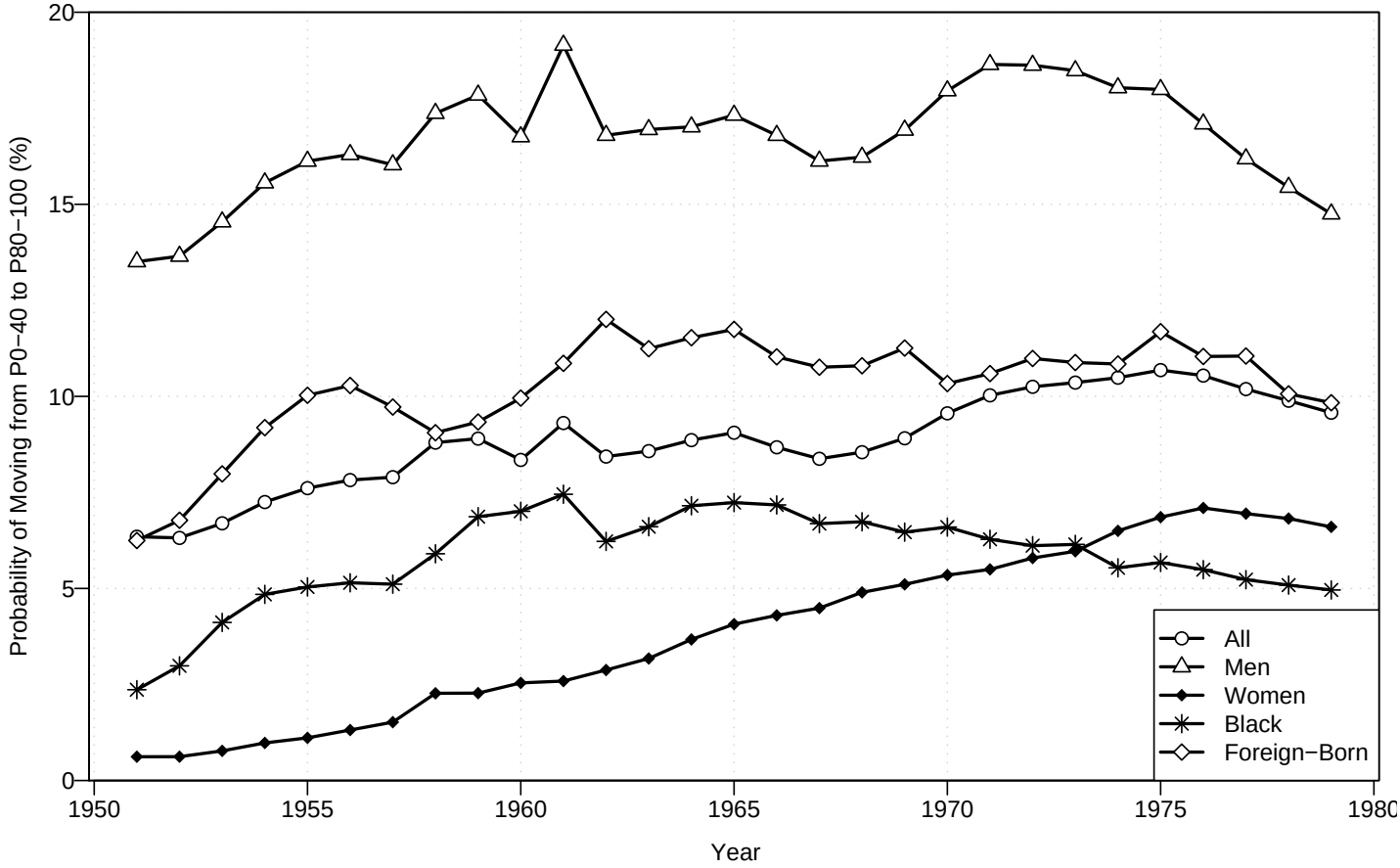


Figure 11A: Upward Mobility After 20 Years (11 year averages), Various Groups



See notes on pages 34-36

Figure 11B: Probability of Upward Mobility over a Career by Gender

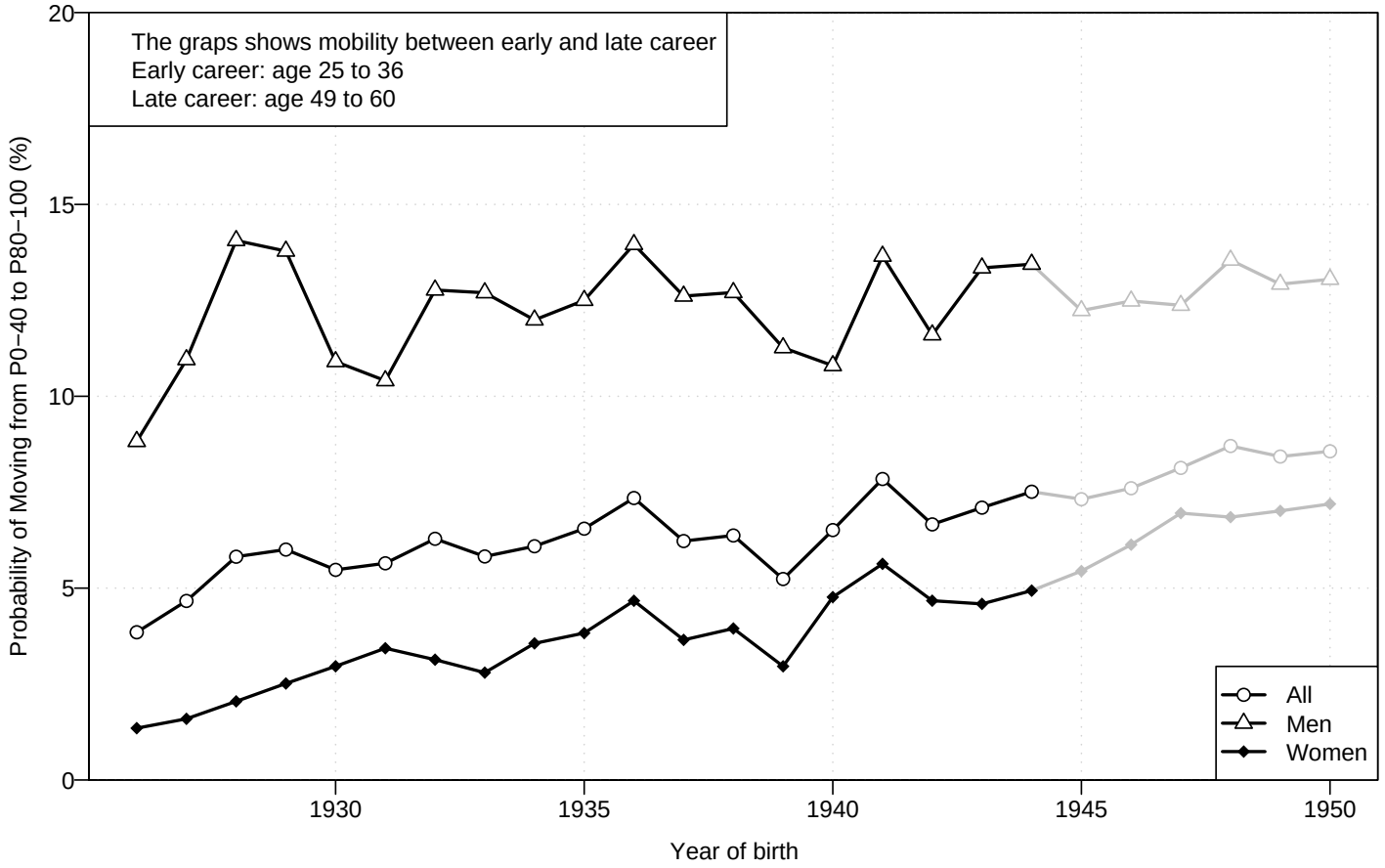
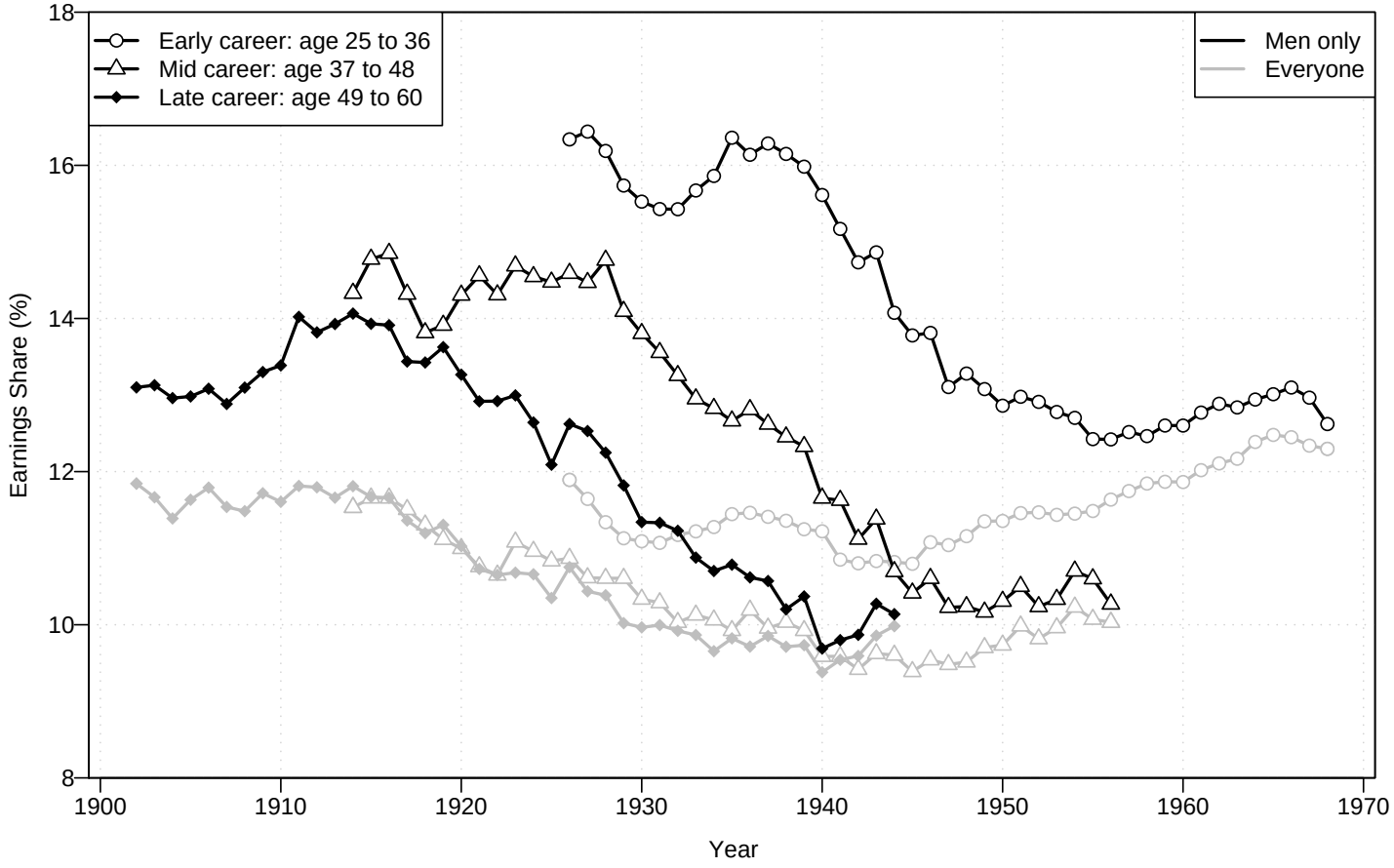


Figure 12: Long−Term P0−40 Share: Men vs. All Workers



See notes on pages 34-36

Figure 13: Women in Top Quintile (Long−Term)

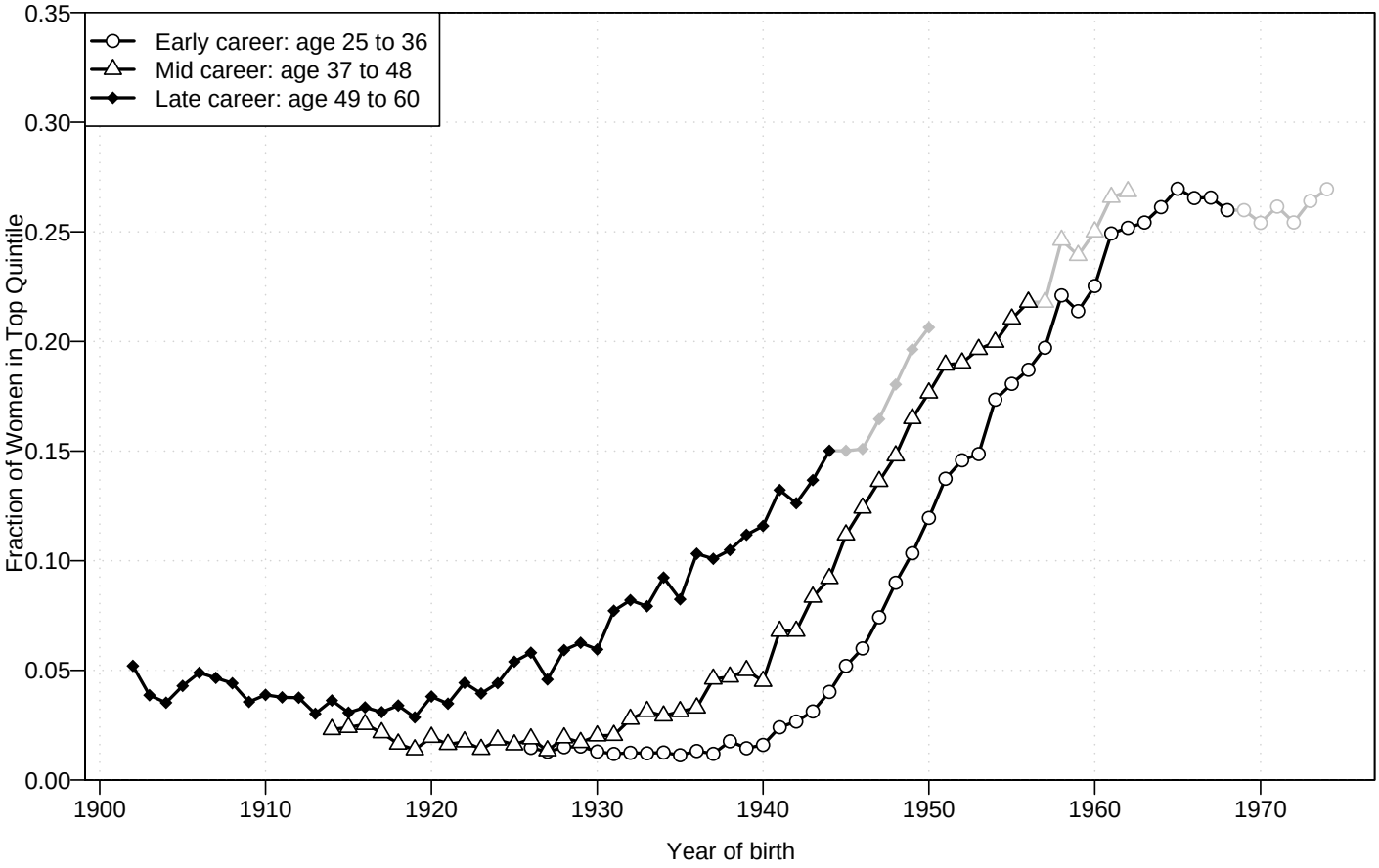


Table 1: Thresholds and Average Earnings by Group in 2004 — Commerce and Industry Sample

Percentile threshold (1)	Earnings threshold (2)	Earnings group (3)	Average earnings (4)	Number of workers (5)
P50	26,553	P0-100	39,176	95,965,600
P0	2,575	P0-20	6,482	19,192,400
P20	10,827	P20-40	15,794	19,193,400
P40	20,966	P40-60	26,715	19,193,500
P60	33,042	P60-80	41,869	19,192,400
P80	53,173	P80-90	63,114	9,596,800
P90	76,211	P90-95	85,304	4,798,800
P95	98,681	P95-99	134,639	3,838,600
P99	219,153	P99-99.5	260,240	479,800
P99.5	319,402	P99.5-99.9	456,234	383,900
P99.9	771,353	Top .1%	1,914,153	96,000