

Earnings Inequality and Mobility in the United States: Evidence from Social Security Data since 1937

Wojciech Kopczuk, Columbia and NBER

Emmanuel Saez, UC Berkeley and NBER

Jae Song, SSA^{1,2}

September 4, 2008

¹Wojciech 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 Tony Atkinson, Clair Brown, David Card, Jessica Guillory, Russ Hudson, Jennifer Hunt, Markus Jantti, Alan Krueger, David Lee Thomas Lemieux, Michael Leonesio, Joyce Manchester, Robert Margo, David Pattison, Michael Reich, editor Lawrence Katz, four anonymous referees, 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.

²Series 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. Annual earnings inequality follows a U-shape pattern, decreasing sharply up to 1953 and increasing steadily afterwards to the present. Short-term earnings mobility measures have been much more stable over the full period except for a temporary surge during World War II. Therefore, the series of annual earnings inequality and multi-year earnings inequality measures follow close patterns overtime. We find that virtually all of the increase in the variance in annual (log) earnings since 1970 is due to the increase in the variance of permanent earnings (as opposed to transitory earnings). Mobility at the top of the earnings distribution has also been very stable and has not mitigated the dramatic increase in annual earnings concentration since the 1970s. We find increases in long-term mobility series among all workers since the 1950s but slight decreases among men, especially in recent decades. The decrease in the gender earnings gap and the resulting substantial increase in upward mobility over a lifetime for women is the driving force behind the increase in long-term mobility among all workers.

1 Introduction

Market economies are praised for creating macro-economic growth but blamed for economic disparities among individuals they generate. Economic inequality is often measured using annual income. However, market economies also generate substantial mobility in earnings over a working lifetime. As a result, annual earnings inequality might substantially exaggerate the extent of true economic disparity among individuals. To the extent that individuals can smooth changes in earnings using savings and credit markets, inequality based on longer periods than a year is a better measure of economic disparity. Thus, a comprehensive analysis of disparity requires studying both inequality and mobility.

A large body of academic work has indeed analyzed earnings inequality and mobility in the United States. A number of key facts from the pre-World War II years to the present have been established using four main data sources:¹ (1) Decennial Census data show that earnings inequality decreased substantially during the “Great Compression” from 1939 to 1949 (Goldin and Margo, 1992) and remained low over the next two decades, (2) the annual Current Population Surveys (CPS) show that earnings inequality has increased substantially since the 1970s and especially during the 1980s (Katz and Murphy, 1992; Katz and Autor, 1999), (3) income tax statistics show that the top of the annual earnings distribution experienced enormous gains over the last 25 years (Piketty and Saez, 2003), (4) panel survey data (primarily the Panel Survey of Income Dynamics) show that 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, 2006). 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 exact timing of the Great Compression. Second, 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 an entire

¹A number of studies have also analyzed inequality and mobility in America in earlier periods (see Lindert, 2000, for a survey on inequality and Ferrie, 2000, for an analysis of occupational mobility).

working life, let alone how such long term mobility has evolved over time. 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, and whether consumption inequality has increased to the same extent as income inequality.² In particular, the development of performance pay such as bonuses and stock-options for highly compensated employees might have increased substantially variability at the top of earnings distribution so that the trends documented in (Piketty and Saez, 2003) could be misleading.

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 balanced panels as samples are selected based on the same Social Security Number endings 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 quarterly earnings information to extrapolate earnings up to 4 times the Social Security annual cap.³ 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, to our knowledge, the Social Security annual earnings data before 1951 have not been used outside SSA for research purposes since Robert Solow unpublished Ph.D. thesis in 1951 (Solow, 1951).

²See e.g., Cutler and Katz (1991), Krueger and Perri (2006), Slesnick (2001), Attanasio et al. (2007).

³Previous 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.

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 country birth place), 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 employment 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. This is an important limitation when analyzing mobility as individuals may move in and out of the commerce and industry sector.⁴

Our analysis allows us to uncover three main findings. First, our annual series confirm the U-shape pattern of earnings inequality since the 1930s. Inequality decreases sharply up to 1953 and increases steadily and continuously afterwards. The U-shape pattern of inequality is also present within each gender group and is more pronounced for men. Percentile ratio series show that (1) the compression in the upper part of the distribution took place from 1942 to 1950 and was followed a steady and continuous widening ever since the early 1950s, (2) the compression in the lower part of the distribution took place primarily in the post war period from 1946 to the last 1960s and unravelled quickly from 1970 to 1983, especially for men, and has been fairly stable over the last two decades.

Second, we find that short-term mobility measures such as rank correlation measures, Shorrocks indices comparing annual vs. multi-year earnings inequality has been quite stable over the full period except for a temporary surge during World War II.^{5,6} In particular, the short-term mo-

⁴Since in recent decades Social Security covers over 95% of earnings, we show in appendix (XX) that our findings for recent decades are robust to including all covered workers. However, we cannot perform such a robustness check for earlier periods when coverage was much less complete.

⁵Such a surge is not surprising in light of the large turnover in the labor market generated by the War.

⁶An earlier version of this paper (Kopczuk et al., 2007) shows that particular elements of rank transition

bility has been remarkably stable since the 1950s, for a variety of mobility measures and also when restricting the sample to men only. Therefore, the evolution over time of annual earnings inequality is very close to the evolution of inequality of longer term earnings. Furthermore, mobility at the top of the earnings distribution, measured by the probability of staying in the top 0.1% after 1, 3, or 5 years has also been very stable since 1978 (the first year in our data with no top code). Therefore, in contrast to the stock-option scenario mentioned above, the SSA data show very clearly that mobility has not mitigated the dramatic increase in annual earnings concentration.

Third, we find that long term mobility measures among all workers, such as the earnings rank correlations from the early part of a working life to the late part of a working life display significant increases since 1951 either when measured unconditionally or when measured within cohorts. However, those increases mask substantial heterogeneity across demographic groups. Among men, we find stability of long-term mobility over most of the period with a slight decrease in recent decades. 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.

The paper is organized as follows. Section 2 presents the conceptual framework linking inequality and mobility measures, 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. Finally, Section 6 offers some concluding remarks. The complete details on the data and our methodology, as well as sensitivity analysis are presented in an electronic appendix. Complete tabulated results in electronic format are posted online.

matrices have been stable as well.

2 Framework, Data, and Methodology

2.1 Conceptual Framework

Our main goal is to document the evolution of earnings inequality. Inequality can be measured over short-term (such as annual earnings) or over long-term (such as earnings averaged over several years or even a lifetime). With mobility in earnings, long-term inequality will be lower than short-term inequality as moving up and down the distribution of short-term earnings will make the distribution of long-term earnings more equal. Therefore, conceptually, a way to measure mobility (Shorrocks, 1978) is to compare inequality of short-term earnings to inequality of long-term earnings and define mobility as:

$$(1) \text{ Long-term earnings inequality} = \text{Short-term earnings inequality} * (1 - \text{Mobility})$$

Alternatively, one can define mobility directly as changes or “shocks” in earnings. In our framework, such shocks are defined broadly as any deviation from long-term earnings. Those shocks could be indeed real shocks such as unemployment, disability, or an unexpected promotion. Changes could also be the consequence of voluntary choices in response to other economic shocks (such as changes in family structure, health, future job prospects etc.) such as reducing (or increasing) hours of work, voluntarily changing jobs, or obtaining an expected pay raise. Such shocks can be transitory (such as working overtime in response to a temporarily increased demand for employer’s product or a short unemployment spell in the construction industry) or permanent (being laid off from a job in a declining industry). In that framework, both long-term inequality and the extent of shocks contribute to shaping short-term inequality:

$$(2) \text{ Short-term earnings inequality} = \text{Long-term earnings inequality} + \text{Variability in earnings}$$

(1) and (2) are related by the formula:

$$(3) \text{ Variability in earnings} = \text{Short-term earnings inequality} * \text{Mobility} = \text{Long-term earnings inequality} * \text{Mobility} / (1 - \text{Mobility})$$

Thus, equation (3) shows that a change in mobility with no change in long-term inequality is due to an increase in variability in earnings. Conversely, an increase in inequality with no change in mobility implies an increased variability in earnings.

In order to derive those formulas formally, we consider a situation where a fixed group of individuals $i = 1, \dots, I$ have *short-term* earnings $z_{it} > 0$ in each period $t = 1, \dots, K$. For example t can represent a year. We can define *long-term* earnings for individual i as the average across all K periods: $\bar{z}_i = \sum_t z_{it}/K$. We normalize earnings so that average earnings (across individuals) are the same in each period.⁷

From a vector of individual earnings $\mathbf{z} = (z_1, \dots, z_I)$, an inequality index can be defined as $G(\mathbf{z})$, where $G(\cdot)$ is convex in \mathbf{z} . We also assume that $G(\cdot)$ is homogeneous of degree zero (multiplying all earnings by a given factor does not change inequality). For example, $G(\cdot)$ can be the Gini index or the variance of log earnings. Shorrocks (1978) shows that (Theorem 1a, p. XX):

$$G(\bar{\mathbf{z}}) \leq \sum_{t=1}^K G(\mathbf{z}_t)/K,$$

where \mathbf{z}_t is the vector of earnings in period t and $\bar{\mathbf{z}}$ the vector of average (across periods) earnings. This inequality captures the idea that movements in individual earnings up and down the distribution reduces long-term inequality (relative to short-term inequality). Hence we can define a related Shorrocks mobility index $0 \leq M \leq 1$ as:

$$1 - M = \frac{G(\bar{\mathbf{z}})}{\sum_{t=1}^K G(\mathbf{z}_t)/K},$$

which is a formalization of equation (1) above.

It is also possible to define direct mobility indices such as the rank correlation in earnings from year t to year $t + p$ (or quintile mobility matrices from year t to year $t + p$). Such mobility indices are likely to be closely related to the Shorrocks indices as re-ranking from one period to another is precisely what creates a wedge between long-term inequality and (the average of) short-term inequality. The advantage of direct mobility indices is that they are more concrete and transparent than Shorrocks indices. In our paper, we will therefore use both and show that they evolve very similarly overtime.

Alternatively, introducing $y_{it} = \log z_{it}$ and $\bar{y}_i = \sum_t \log z_{it}/K$, we can define deviations in (log) earnings as:

$$\varepsilon_{it} = y_{it} - \bar{y}_i.$$

⁷In our empirical analysis, earnings will be indexed to the nominal average earnings index.

It is important to note that ε_{it} incorporates both transitory earnings shocks (such as an iid process) and permanent earnings shocks (such as a Brownian motion). The deviation ε_{it} could either be uncertain ex-ante from the individual perspective, or predictable.⁸

The Shorrocks theorem applied to the inequality index variance of log-earnings implies that

$$\text{var}_i(\bar{y}_i) \leq \text{var}_{it}(y_{it}),$$

where the variance $\text{var}_{it}(y_{it})$ is taken over both $i = 1, \dots, I$ and $K = 1, \dots, t$. If, for illustration, we make the statistical assumption that $\varepsilon_{it} \perp \bar{y}_i$ and $\text{var}(\varepsilon_{it}) = \sigma_\varepsilon^2$, then we have:

$$\text{var}_{it}(y_{it}) = \text{var}_i(\bar{y}_i) + \sigma_\varepsilon^2,$$

which is a formalization of equation (2) above. The Shorrocks inequality index in that case is

$$M = \sigma_\varepsilon^2 / \text{var}_{it}(y_{it}) = \sigma_\varepsilon^2 / (\text{var}_i(\bar{y}_i) + \sigma_\varepsilon^2).$$

This shows that short-term earnings variance can increase because of an increase in long-term earnings variance or an increase in the variance of earnings deviations. Alternatively and equivalently, short-term inequality can increase while long-term inequality is stable if mobility increases. This simple framework can help us reconcile the findings from the previous literature on earnings mobility in the United States. Rank based mobility measures (such as year-to-year rank correlation or quintile mobility matrices) are stable over time (XX cite) while there has been an increase in both the variance of transitory earnings (XX cite). Obviously, such findings can be reconciled if the disparity in permanent earnings has widened while rank-based mobility of earnings are remained stable so that earnings changes are larger on average in levels.

The theoretical framework we just described critically assumed that we consider the same set of individuals across the K short-term periods. In practice, because individuals leave or enter the labor force (or the “commerce and industry” sector we will be focusing on), the set of individuals with positive earnings varies across periods. As the number of periods K becomes large, the sample will become smaller. Therefore, we will only consider relatively small values

⁸Uncertainty is important conceptually as individuals facing no credit constraints can fully smooth predictable shocks while uncertain (large) shocks can only be smoothed with insurance. We do not pursue this distinction in our analysis as we cannot observe the degree of uncertainty in the empirical earnings shocks.

of K such as $K = 3$ or $K = 5$. When a period is a year, that allows us to analyze short term mobility. When a period is a longer period of time such as 12 consecutive years, with $K = 3$, we cover 36 years which is almost a full lifetime of work. This allows us to analyze long term mobility, i.e., mobility over a full working life.

Our analysis will focus on the time series of various inequality and mobility statistics. The framework we have considered can be seen as an analysis at a given point in time s . We can recompute those statistics for a various points in time in order to create time series.

2.2 Data and Methodology

• Social Security Administration Data

We use primarily datasets constructed in the Social Security Administration for research and statistical analysis known as the Continuous Work History Sample (CWHS) system.⁹ 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 so that the sample is a balanced panel and can be treated as a random sample of the full population 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, employment 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

⁹Detailed documentation of these datasets can be found in Panis et al. (2000)

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 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 quarterly earnings information from 1957 to 1977 which allows us to impute earnings above the top code (see below).

(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 the same demographic variables as well as 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.¹⁰ Finally, since 1978, the data are fully uncapped.

For individuals with earnings above the taxable ceiling (from 1937 to 1945) or who reach the taxable ceiling in the first quarter (from 1946 to 1977), we impute earnings assuming a Pareto distribution above the top code (1937-1945) or four times the top code (1946-1977). The Pareto distribution is calibrated from wage income tax statistics published by the Internal Revenue Service to match the top wage income shares series estimated in Piketty and Saez (2003). From 1951 to 1977, we use earnings for quarters when they are observed to impute earnings in quarters

¹⁰To 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.

that are not observed (when the taxable ceiling is reached after the first quarter). This method is discussed in more detail in Kestenbaum (1976, his method II).¹¹

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 virtually all years after 1951. 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. 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.

In order to assess the sensitivity of our *mobility* and multi-year inequality estimates with respect to top code imputation, we use two Pareto imputation methods (see appendix). In the first or main method, the Pareto imputation is based on draws from a uniform distribution that are independent across individuals but also time periods. This generates an upper bound on mobility within top coded individuals. The alternative method, the uniform distribution draw is independent across individuals but fixed over time for a given individual. This generates a lower bound on mobility. We always test that the two methods generate virtually the same series (see appendix XX for examples). It is not too surprising because, starting with 1951, imputations matter for just the top 1% of the sample and mobility measures for the full population are not very sensitive to what happens within the very top group.

• 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

¹¹For 1946-1950, the imputation procedure Kestenbaum (1976, his method I) also uses Pareto distributions and preserves the rank order based on the quarter when the taxable maximum was reached.

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). We show in appendix XX 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 25 and above (by January 1st of the corresponding year) up to age 60 (by January 1st of the corresponding year). This top age restriction allows us to concentrate on the working-age population.¹² 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.¹³ From now on, we denote this sample the “core sample”

¹²Kopczuk et al. (2007) used a wider age group from 18 and 70 and found the same overall time patterns for all the series.

¹³We show in appendix XX that our results are not sensitive to choosing a higher minimum threshold such as

3 Annual Earnings Inequality

Figure 1 plots the annual Gini coefficient from 1937 to 2004 for the core sample of all workers, and for men and women separately in lighter grey. 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). The series displays a sharp decrease of the Gini coefficient from 0.44 in 1938 down to 0.36 in 1953 (the Great Compression) followed by a steady increase since 1953 which accelerates in the 1970s and especially the 1980s. The Gini coefficient surpassed the pre-war level in the late 1980s and is highest in 2004 at 0.47.

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 1942 to 1944, rebounded partly from 1944 to 1946 and then declined again from 1946 to 1953. Among all workers, the increase in the Gini coefficient over the five decades from 1953 to 2004 is close to linear which suggests that changes in overall inequality were not just limited to an episodic event in the 1980s.

Figure 1 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 much more pronounced for men than for all workers. This shows that the rise in the Gini coefficient since 1970 cannot be attributed to changes in gender composition of the labor force. The Gini for men shows a dramatic increase in a short time period from 0.35 in 1979 to 0.43 in 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 1950s through the late 1960s highlights that the overall increase in the Gini coefficient in that period has been driven by a widening of the gender gap in earnings (i.e., the between rather than within group component). Strikingly, there is more earnings inequality among women than among men in the 1950s and 1960s while the reverse is true before the Great compression and since the late 1970s.

a full year-full time minimum wage.

Finally, the increase in the Gini coefficient has slowed since the late 1980s in the overall sample. It is interesting to note that a large part of the 3.5 points increase in the Gini from 1990 to 2004 is due to a surge in earnings within the top percentile of the distribution. The series of Gini coefficients estimated excluding the top 1% increases by less than 2 points since 1990 (see appendix XX).¹⁴ It should also be noted that, since the late 1980s, the Gini coefficient has increased faster for men and women separately than for all workers. This has been driven by an increase in the earnings of women relative to men, especially at the top of the distribution as we shall see.

Most previous work has focused on gender specific measures of inequality. As men and women share a single labor market, it is also valuable to analyze the overall inequality generated in the labor market (in the “commerce and industry” sector in our analysis). Our analysis for all workers and by gender provides clear evidence of the importance of changes in women’s labor market behavior and outcomes for understanding overall changes in inequality, a topic we will return to.

In order to understand where in the distribution the changes in inequality displayed on Figure 1 are occurring, Figure 2 displays the (log) percentile annual earnings ratios P80/P50—measuring inequality in the upper half of the distribution—and P50/P20—measuring inequality in the lower half of the distribution. We also depict the series for men and women only separately in lighter grey.¹⁵

The P80/P50 series also displays a U-shape over the period with a brief but substantial “Great Compression” from 1942 to 1947 and a steady increase starting in 1951 which accelerates in the 1970s. Interestingly, P80/P50 is virtually constant from 1985 to 2000 showing that the gains at the top of the distribution occurred well above P80. The series for men is similar except that P80/P50 increases sharply in the 1980s and continues to increase in the 1990s.

P50/P20 displays a fairly different time pattern than P80/P50. First, the compression

¹⁴This 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 even when using global indices such as the Gini.

¹⁵We choose P80 (instead of the more usual P90) in order to avoid top coding issues before 1951 and P20 (instead of the more usual P10) so that our low percentile estimate is not too closely driven by the average wage indexed and very low minimum we have chosen (\$2575 in 2004).

happens primarily in the post war period from 1946 to 1953. There are large swings in P50/P20 during the war, especially for men, as many young low income earners leave and enter the labor force because of the war, but P50/P20 is virtually the same in 1941 and 1946 or 1947.¹⁶ After the Great Compression, P50/P20 for all workers remains fairly stable to the present alternating periods of increase and decrease. In particular, it decreases smoothly from the mid-1980s to 2000 implying that inequality in the bottom half shrunk in the last two decades although it started increasing after 2000. The pattern for men only is quite different and displays an overall U-shape with a sharper great compression which extends well into the post-war period with an absolute minimum in 1969 followed by a sharp increase up to 1983 and relative stability since then.

Table 1 summarizes the annual earnings inequality trends for all (Panel A), men (Panel B), women (Panel C) with various inequality measures for selective years (1937-1939, 1960, 1980, and 2004).

XX summarize key findings XX

4 The Effects of Short Term Mobility on Earnings Inequality

In this section, we use our theoretical framework from Section 2.1 to analyze multi-year inequality and relate it to annual earnings inequality analyzed in Section 3. In the theoretical framework, we will consider each period to be a year and the longer period will be 5 years ($K = 5$).¹⁷ We will compare inequality based on annual earnings and earnings averaged over 5 years. We will then derive the implied Shorrocks mobility index, and decompose annual inequality into multi-year inequality and the variability in earnings. We will also examine some direct measures of mobility such as rank correlations.

Figure 3 plots the Gini coefficient series for earnings averaged over 5 years¹⁸ and annual

¹⁶In the working paper version (Kopczuk et al., 2007) we show that compositional changes during the war are strongly influencing the bottom of the distribution during the early 1940s and find some suggestive evidence that the “Great Compression” at the bottom of the distribution may have taken place in the early 1940s as well.

¹⁷Computing series based on 3 year averages instead of 5 year generates very similar patterns. Increasing K beyond 5 would reduce substantially sample size as we require earnings to be above the minimum in each of the 5 years as described below.

¹⁸The average is taken after indexing annual earnings by the average wage index.

earnings. For a given year t , the sample for both the five year Gini and the annual Gini is defined as all individuals with “Commerce and Industry” earnings above the minimum threshold in all 5 years, $t-2, t-1, t, t+1, t+2$ (and aged 25 to 60 in the middle year t). We show the average of the five annual Gini coefficients between $t-2$ and $t+2$ as our measure of annual Gini coefficient, because it matches the Shorrocks’ approach. Because the sample is the same for both series, Shorrocks’ theorem implies that the five year Gini is always smaller than the average of the annual Gini (over the corresponding 5 years) as indeed displayed on the Figure. We also display the same series for men only (in lighter grey).¹⁹ The annual Gini displays the same overall pattern as in Figure 1. The level is lower as there is naturally less inequality in the group of individuals with positive earnings for 5 consecutive years. The Gini coefficient estimated for 5 year annual average follows a very similar pattern and is actually extremely close to the annual Gini, especially in recent decades.

Interestingly, in this sample, the Great Compression takes place primarily during the war from 1940 to 1944. The war compression is followed by a much more modest decline till 1952. This suggests that the post war compression observed in annual earnings in Figure 1 was likely due to entry (of young men in the middle of the distribution) and exit (likely of war working women in the lower part of the distribution). The 5 year Gini displays an accelerated increase during the 1970s and especially the 1980s.

Figure 4 displays two measures of mobility (in black for all workers and in lighter grey for men only). The first one is the Shorrock measure defined as the ratio of five year Gini to the (average of) annual Gini. This measure is very close to one after a temporary dip during the war. In particular, it has steadily increased since the early 1970s from 0.945 to 0.967 in 2004. This small change further confirms, as we expected from Figure 3, short-term mobility has played a minor role in the surge in annual earnings inequality documented on Figure 1. The second measure is the straight rank correlation in earnings between year t and year $t+1$ (in the sample of individuals present in our core sample in both years t and $t+1$). Rank mobility follows the same overall pattern as the Shorrock mobility index: a temporary dip during the war followed

¹⁹Alternatively, we could have defined the sample as all individuals with earnings above the minimum threshold in any of the 5 years, $t-2, t-1, t, t+1, t+2$. Those series follow the same pattern and are presented in appendix XX. We prefer to use the positive earnings in all 5 years criterion because this is a necessity when analyzing variability in log-earnings.

a slight increase. Over the last two decades, the rank correlation in year-to-year earnings has been very stable and very high around 0.9.

Figure 5 displays the average of variance of annual log earnings over $t-2$ to $t+2$ (defined on the stable sample as before), the variance of five year average log-earnings, $\text{var}\left(\frac{\sum_{s=t-2}^{t+2} \log z_{is}}{5}\right)$, and the variance of log earnings deviations estimated as

$$D_t = \text{var}\left(\log(z_{it}) - \frac{\sum_{s=t-2}^{t+2} \log z_{is}}{5}\right),$$

where the variance is taken across all individuals i with earnings above the minimum threshold in all 5 years $t-2, \dots, t+2$. As the previous two mobility measures, those series, displayed in black for all workers and in grey for men only, show a temporary surge in the variance of transitory earnings during the war, followed by a fairly stable pattern after 1960. In particular, it is striking that we do not observe an increased earnings variability over the last 20 years so that all the increase in the log-earnings variance can be attributed to the increase in the variance of permanent (five year average) log-earnings. Our results stand in some contrast to Gottschalk and Moffitt (1994) using PSID data found that over one third of increase in the variance of log-earnings from the 1970s to the 1980s was due to an increase in transitory earnings (Table 1, row 1, p. 223). It should be pointed out though that a part of an autoregressive process (which is a part of the G-M model) would show up in our permanent component and that innovations to G-M permanent component occurring over the five-year window would show up in the variance of our transitory component. The small magnitude of our temporary component suggests that most of the variance is due to a persistent component — be it random walk or a relatively persistent component of a mean-reverting process.²⁰

The absence of top code since 1978 allows us to zoom on top earnings which, as we showed in Table 1, have surged in recent decades. Figure 6A uses the uncapped data since 1978 to plot the annual earnings share for the top 1% (those with earnings above \$236,000 in 2004). The top 1% annual earnings share doubles from 6.5% in 1978 to 13% in 2004.²¹ Figure 6A then compares the

²⁰Our results could differ from Gottschalk and Moffitt (1994) for many other reasons such as measurement error and earnings definition consistency issues in the PSID or the sample definition. Gottschalk and Moffitt focus exclusively on white males, use a different age cut-off, take out age-profile, and include earnings from all industrial sectors. Gottschalk and Moffitt use 9 year earnings periods (instead of 5 as we do) and include all years with positive annual earnings years (instead of requiring positive earnings in all 9 years as we do).

²¹The closeness of our SSA based (individual-level) results and the tax return based (family-level) results

share of earnings of the top 1% based on annual data with shares of the top 1% defined based on earnings averaged on the individual level over 5 years (displayed in lighter grey). The 5 year average earnings share series naturally smoothes short-term fluctuations but shows the same pattern of robust increase as the annual measure.²² This shows that the surge in top earnings is not due to increased mobility at the top. This finding is confirmed in Figure 6B which shows the probability of staying in the top 1% earnings group after 1, 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. Therefore, our analysis shows that the dramatic surge in top earnings has not been accompanied by a similar surge in mobility in and out top earnings groups. Hence, annual earnings concentration measures provide a very good approximation to longer-term earnings concentration measures.²³

Table 2 summarizes the key short-term mobility trends for all (Panel A) and men (Panel B) with various mobility measures for selective years (1939, 1960, 1980, and 2002).

XX summarize key findings XX

In sum, the movements in short-term mobility series appear to be much smaller than changes in inequality over time. As a result, changes in short-term mobility have had no significant impact on inequality trends 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)²⁴ 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. Our finding of stable transitory earnings variance is also at odds with the findings of Gottschalk and Moffitt (1994) of Piketty and Saez (2003) show that changes in assortative mating played at best a minor role in the surge of family employment earnings at the top of the earnings distribution.

²²In order to follow the framework from Section 2.1 (applied in this case to the top 1% earnings share measure of inequality), we have computed such shares (in year t) on the sample of all individuals with minimum earnings in all 5 years, $t - 2, \dots, t + 2$.

²³In particular, the development of bonuses and stock-options (whose exercise profits are included in our earnings measure) paid to highly compensated employees does not seem to have increased dramatically mobility.

²⁴The CBO study focuses on probabilities of large earnings increases (or drops) instead of quantile mobility as we do here.

who decompose transitory and permanent variance in log-earnings using PSID data and show an increase in both components. Our decomposition using SSA data shows that only permanent earnings variance has increased in recent decades.

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 working life. Such estimates have not yet been produced for the United States in any systematic way because of the lack of very long and large panels.

5.1 Unconditional Long-Term Inequality and Mobility

We begin with the simplest extension of our previous analysis to a longer-term horizon. In the context of the theoretical framework from Section 2.1, we now assume that a period is 11 consecutive years. We define the “core long-term sample” in year t as all individuals aged 25-60 in year t with average earnings (using the standard wage indexation) from year $t - 5$ to year $t + 5$ above the minimum threshold. Hence, our sample includes individuals with zeros in some years as long as average earnings are above the threshold.²⁵

Figure 7 displays the Gini coefficients for all workers, and men and women separately based on those 11 year average earnings from 1942 to 1999. The overall picture is actually strikingly similar to our annual Figure 1. The Gini coefficient series for all workers displays an overall U-shape with a Great compression from 1942 to 1953, an absolute minimum in 1953, followed by a steady increase which accelerates in the 1970s and 1980s and slows down in the 1990s. The U-shape pattern is also much more pronounced for men than for women and shows that, for men, the inequality increase was concentrated in the 1970s and 1980s.²⁶

After exploring base inequality over those 11 year spells, we turn to long-term mobility. Figure 8 displays the rank correlation between 11 year earnings spell centered in year t and 11 year earnings spell after T years (i.e., centered in year $t + T$ respectively) in the same sample

²⁵This allows us to analyze large and representative samples as the number of individuals with positive “commerce and industry” earnings in 11 consecutive years is only between 35 and 50% of the core annual samples.

²⁶We show in appendix XX that those results are robust to using a higher minimum threshold.

of individuals present in the “long-term core sample” in both year t and year $t + T$. The figure presents such correlations for three choices of T : 10 years, 15 years, and 20 years. Given our 25-60 age restriction (which applies in both year t and $t + T$), for $T = 20$, in sample in year t is aged 25 to 40 (and the sample in year $t + 20$ is aged 45 to 60). Thus, this measure captures mobility from early career to late career. The figure also displays the same series for men only in lighter grey. Three points are worth noting.

First, the correlation is unsurprisingly lower as T increases but it is striking to note that even after 20 years, the correlation is still substantial (in the vicinity of 0.5). Second, the series for all workers show that rank correlation has actually significantly decreased over time: from example the rank correlation between 1950s and 1970s earnings was around 0.57 but it is only 0.49 between 1970s and 1990s earnings. This shows that long-term mobility has *increased* significantly over the last five decades. This result stands in contrast to our short-term mobility results displaying substantial stability. Third however, Figure 8 shows that this increase in long-term mobility disappears in the sample of men. The series for men displays a slight decrease in rank correlation in the first part of the period followed by an increase in the last part of the period. On net, the series for men display a slight overall increase in rank correlation and hence a slight decrease in long-term mobility over the full period.

5.2 Cohort based Long-Term Inequality and 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. We divide working lifetimes from age 25 to 60 into three stages: 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. For a given year of birth cohort, we define the “core early career sample” as all individuals with average “commerce and industry” earnings (using again the standard wage indexation) above the minimum threshold in the early career. We define similarly the “core mid career” and “core late career” are defined similarly for each birth cohort. The earnings in early, mid, or late career are defined as average “commerce and

industry” earnings during the corresponding stage (always using the average wage index).

Figure 9 reports the Gini coefficient series by year of birth for early, mid, and late career. The Gini coefficients for men only are also displayed in lighter grey. The cohort based Gini coefficients are consistent with our previous findings and display a U-shape over the full period. Two points are worth noting. First, there is much more inequality in late career than in middle career, and in middle career than in early career showing that long-term inequality fans out over the course of a working life. Second, the Gini series show that long-term inequality has been stable for the baby-boom cohorts born after 1945 in the sample of all workers. Those results are striking in light of our previous results showing a worsening of inequality in annual cross-sections and for multi-year averages. Third, however, the Gini series for men only show that inequality has increased substantially across baby-boom cohorts. This sharp contrast between series for all workers versus men only reinforces our previous findings that gender effects play an important role in shaping the trends in overall inequality.

Figure 10 reports the rank correlation between (1) early career earnings and middle career earnings, (2) middle career earnings and late career earnings, (3) early career earnings and late career earnings. It also displays the same series for men only in lighter grey. Two important results should be noted. First, as we found in the unconditional mobility series, mobility over a career is fairly modest with a correlation over 0.5 from early to late career. Second, the pattern of mobility over the period displays modest increases in mobility over the period we analyze. Those changes are most visible in decreases in rank correlation from early to late career. However, those results of increasing mobility disappear in the sample for men only suggesting again that they are driven by gender effects, to which we now turn.

5.3 The Role of Gender Gaps in Long-Term Inequality and Mobility

As we saw, there are striking differences in the long-term inequality and mobility series for all workers vs. for men only: Long-term inequality has increased much less in the sample of all workers than in the sample of men only. Long-term mobility has increased over the last 4 decades in the sample of all workers but not in the sample of men only. Such differences can be explained by the reduction in the gender gap that has taken place over the period.

Figure 11 plots the fraction of women in our core sample and in various upper earnings

groups: the fourth quintile group (P60-80), the ninth decile group (P80-90), the top decile group (P90-100), and the top percentile group (P99-100). As adult women aged 25 to 60 are about half of the adult population aged 25 to 60, with no gender differences, those fractions should be approximately 0.5. Those representation indices with no adjustment capture the total realized gaps including labor supply decisions.²⁷ We use those representation indices instead of the traditional ratio of mean (or median) female earnings to male earnings because such representation indices remain meaningful in the presence of differential changes in labor force participation or changes in the wage structure, and we do not have covariates to control for such changes as is done in survey data (see e.g. Blau et al., 2006). Two elements on Figure 11 are worth noting.

First, the fraction of women in the core sample of commerce and industry workers has increased from around 23% in 1937 to about 44% in 2004. World War II generated a temporary surge in women labor force participation, two thirds of which was reversed immediately after the war.²⁸ Women labor force participation has been steadily and continuously increasing since the mid 1950s and has been stable at around 43-44% since 1990.

Second, Figure 11 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.²⁹ For example, the fraction of women in P60-80 starts to increase in 1966 from around 8% and reaches about 34% in the early 1990s and has remained about stable since then. The fraction of women in the top percentile (P99-100) does not really start to increase significantly before 1980. It grows from around 2% in 1980 to almost 14% in 2004 and is still quickly increasing.

Figure 12 displays the upward mobility (defined as the probability of moving from the bottom two quintile groups to the top quintile group) after 20 years for all workers, men only, and women

²⁷As 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).

²⁸This is consistent with the analysis of Goldin (1991) who uses a unique micro survey data covering women workforce history from 1940 to 1951.

²⁹There 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.

only. The figure shows a striking heterogeneity across groups. First, men have significantly higher levels of upward mobility than overall workers while women 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 increase in upward mobility for women compensate for the stagnation or slight decline in mobility for men so that the overall upward mobility for all workers is slightly increasing.³⁰ Figure 12 also suggests that the gains in annual earnings made by women documented earlier were in part due to women already in the labor force making earnings gains rather than gains entirely due to the entry of new cohorts of women with higher earnings.

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 entirely the possibility that mobility patterns across those two groups were different decades ago.³¹

³⁰It 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.

³¹The same critique can be made for movements between employment and self-employment.

6 Conclusion

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 short-term 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. In particular, we find that increases in annual earnings inequality are driven almost entirely by increases in permanent earnings inequality with much more modest changes in the variability of transitory earnings.

However, our key finding is that while the overall measures of mobility are fairly stable, they hide heterogeneity by gender groups. Inequality and mobility 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 compensate for the increase in inequality for males.

References

- 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.
- Blau, Francine D.**, “Trends in the Well-being of American Women, 1970-1995,” *Journal of Economic Literature*, March 1998, 36 (1), 112–165.
- , **Marianne Ferber, and Anne Winkler**, *The Economics of Women, Men and Work*, 4th ed., Prentice-Hall, 2006.
- Carroll, Robert, David Joulfaiian, 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.
- Ferrie, Joseph**, “Moving Through Time: Mobility in America Since 1850,” Manuscript under contract, Cambridge University Press 2008.
- 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.
- , “The Quiet Revolution That Transformed Women’s Employment, Education, and Family,” *American Economic Review Papers and Proceedings*, May 2006, 96 (2), 1–21.

- **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.
- Gottschalk, Peter**, “Inequality, Income Growth, and Mobility: The Basic Facts,” *Journal of Economic Perspectives*, Spring 1997, *11* (2), 21–40.
- **and Robert Moffitt**, “The growth of earnings instability in the U.S. labor market,” *Brookings Papers on Economic Activity*, 1994, (2), 217–54.
- 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.
- Hungerford, Thomas L**, “U.S. Income Mobility in the Seventies and Eighties,” *Review of Income and Wealth*, December 1993, *39* (4), 403–417.
- 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.
- 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, Emmanuel Saez, 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.
- 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.

Panis, Constantijn, Roald Euler, 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.

Shorrocks, Anthony F., "Income Inequality and Income Mobility," *Journal of Economic Theory*, December 1978, 19 (2), 376–93.

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.

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.

Solow, Robert M., "On the Dynamics of the Income Distribution." PhD dissertation, Harvard University 1951.

Figures footnotes

Figure 1: The figure displays the Gini coefficients from 1937 to 2004 for earnings of (1) individuals in the core sample, (2) men in the core sample, (3) women in the core sample. The core sample in year t is defined as all employees with Commerce and Industry earnings above a minimum threshold (\$2,575 in 2004 and indexed using average wage for earlier years) and aged 25 to 60 (by January 1st of 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. Self-employment earnings are fully excluded. See appendix XX for complete details.

Figure 2: Sample is the core sample (commerce and industry employees aged 25 to 60, see Figure 1 footnote). The figure displays the log of the 50th to 20th percentile earnings ratio (upper part of the figure) and the log of the 80th to 50th percentile earnings ratio (lower part of the figure) for the distribution of all workers and for the distribution of men only (in lighter grey).

Figure 3: The Figure displays the Gini coefficient for annual earnings and for earnings averaged over 5 years from 1939 to 2002. In year t , the sample for both series is defined as all individuals aged 25 to 60 in year t , with commerce and industry earnings above the minimum threshold in all 5 years $t-2, t-1, t, t+1, t+2$. Earnings are averaged over the 5 year span using the average earnings index. The Gini coefficient for annual earnings displayed for year t is the average of the Gini coefficient for annual earnings in years $t-2, \dots, t+2$. The same series are reported in lighter grey for the sample restricted to men only.

Figure 4: The Figure displays the Shorrocks mobility coefficient based on annual earnings Gini vs. five year average earnings Gini and the rank correlation between earnings in year t and year $t+1$. The Shorrocks mobility coefficient in year t is defined as the ratio of the 5 year earnings (from $t-2$ to $t+2$) Gini coefficient to the average of the annual earnings Gini for years $t-2, \dots, t+2$ (those two series are displayed on Figure 3). The rank correlation in year t is estimated on the sample of individuals present in the core sample (commerce and industry employees aged 25 to 60, see Figure 1 footnote) in *both* year t and year $t+1$.

The same series are reported in lighter grey for the sample restricted to men only.

Figure 5: The Figure displays the variance of (log) annual earning, the variance of (log) five year average earnings, and the variance of the difference between (log) annual earnings and (log) five year average earnings. In year t , the sample for all three series is defined as all individuals aged 25 to 60 in year t , with commerce and industry earnings above the minimum threshold in all 5 years $t - 2, t - 1, t, t + 1, t + 2$. The (log) annual earnings variance is estimated as the average (across years $t - 2, \dots, t + 2$) of the variance of (log) annual earnings. The same series are reported in lighter grey for the sample restricted to men only.

Figure 6A: The sample in year t is all individuals aged 25 to 60 in year t and commerce and industry earnings above the minimum threshold in all 5 years $t - 2, t - 1, t, t + 1, t + 2$. In year t , the Figure displays (1) the share of total year t earnings accruing to the top 1% earners in that year, (2) the share to total five year average earnings (from year $t - 2, \dots, t + 2$) accruing to the top 1% earners (in terms of average 5 year earnings).

Figure 6B: The figure displays the probability of staying in the top 1% annual earnings group after X years (where $X = 1, 3, 5$). The sample in year t is all individuals present in the core sample (commerce and industry employees aged 25 to 60, see Figure 1 footnote) in *both* year t and year $t + X$.

Figure 7: The figure displays the Gini coefficients from 1942 to 1999 for 11 year average earnings for all workers, men only, and women only. The sample in year t is defined as all employees aged 25 to 60 in year t , alive in all years $t - 5$ to $t + 5$, and with average Commerce and Industry earnings (averaged using the average wage index) from year $t - 5$ to $t + 5$ above the minimum threshold.

Figure 8: The figure displays in year t the rank correlation between 11 year average earnings centered around year t and 11 year average earnings centered around year $t + X$ where $X = 10, 15, 20$. The sample is defined as all individuals aged 25 to 60 in year t and $t + X$, with average 11 year earnings around years t and $t + X$ above the minimum threshold. 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 $t + 1$. The same series are reported in lighter grey for the sample restricted to men only (in which case, rank is estimated within the sample of men only).

Figure 9: 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. Early career is from age 25 to 36, middle career is from age 37 to 48, late career is from age 49 to 60. 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 10: The figure displays the rank correlation between earnings in two stages of the career. The sample is defined as individuals present in the career sample (see Figure 9) for both corresponding stages of the career.

Figure 11: Sample is core sample (commerce and industry employees aged 25 to 60, see Figure 1 footnote). The figure displays the fraction of women in various group. $P60 - 80$ denotes the fourth quintile group from percentile 60th to percentile 80, $P90-100$ denotes the top 10%, etc. 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 12: The figure displays in year t the probability of moving to the top quintile group ($P80-100$) for 11 year average earnings centered around year $t + 20$ conditional on having 11 year average earnings centered around year t in the bottom two quintile groups ($P0-40$). The sample is defined as all individuals aged 25 to 60 in year t and $t + 20$, with average 11 year earnings around years t and $t + 20$ above the minimum threshold. 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 $t + 1$. The

series are reported for all workers, men only, and women only. In all three cases, quintile groups are defined based on the sample of all workers.

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

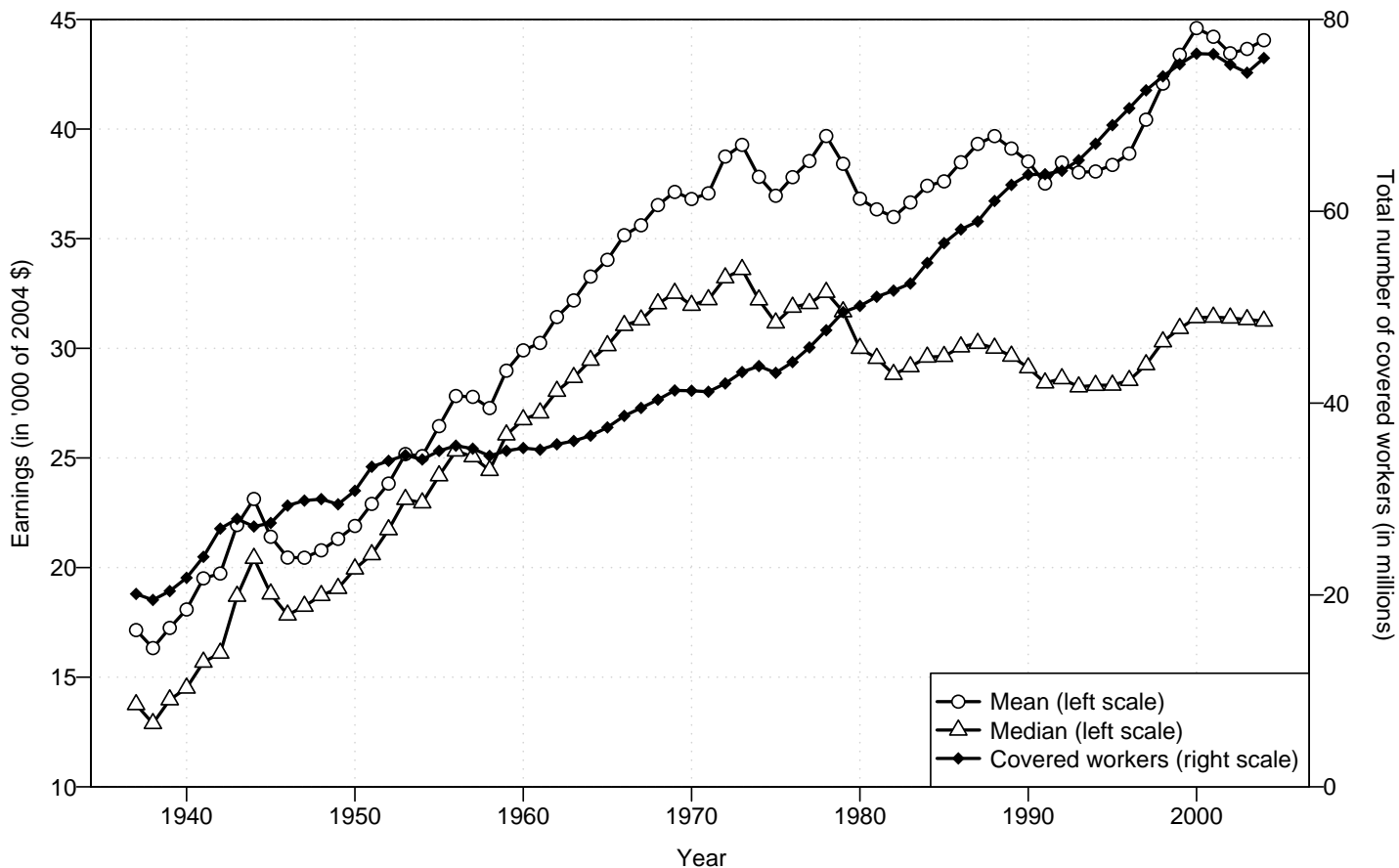


Figure 1: Gini coefficient

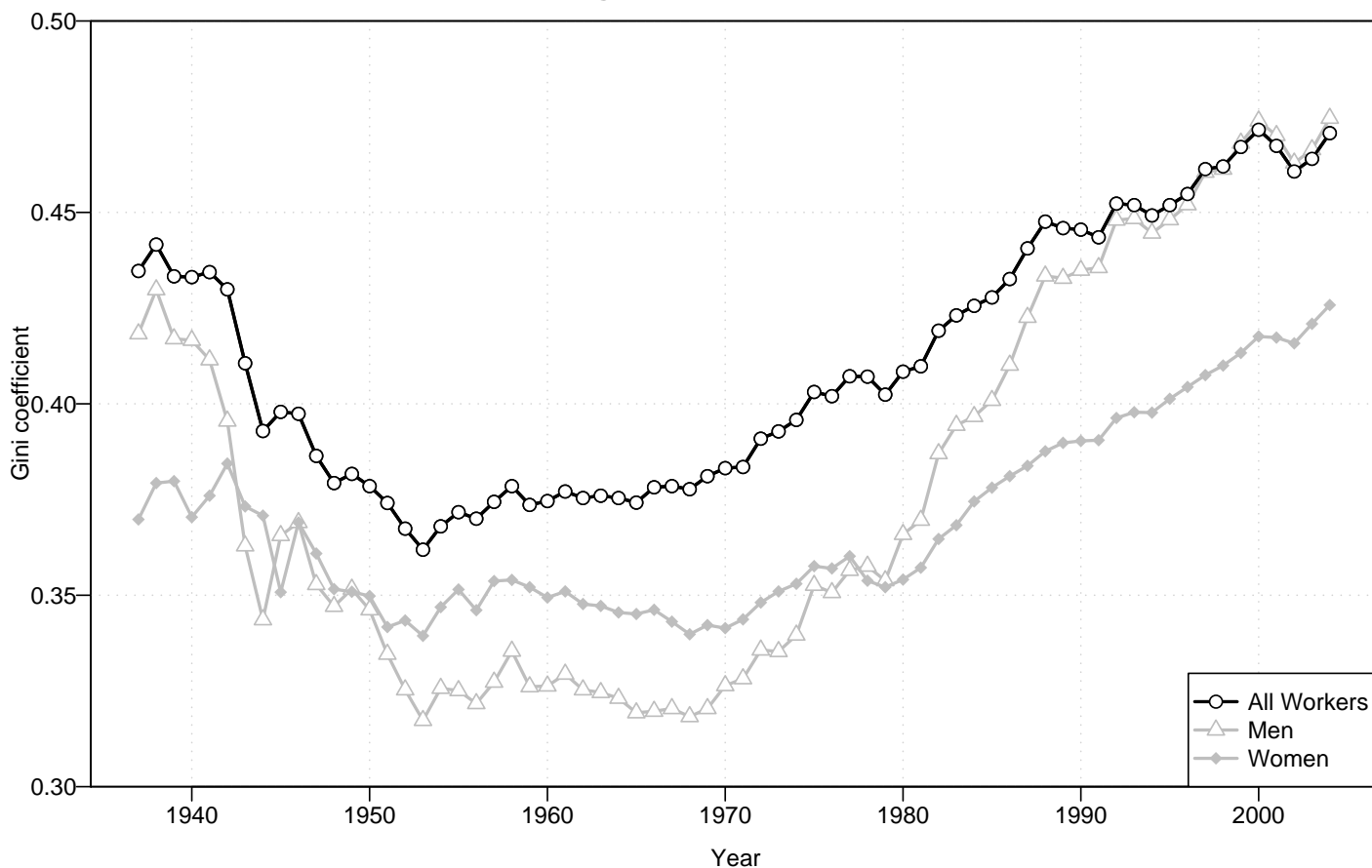


Figure 2: Percentile ratios Log(P80/P50) and Log(P50/P20)

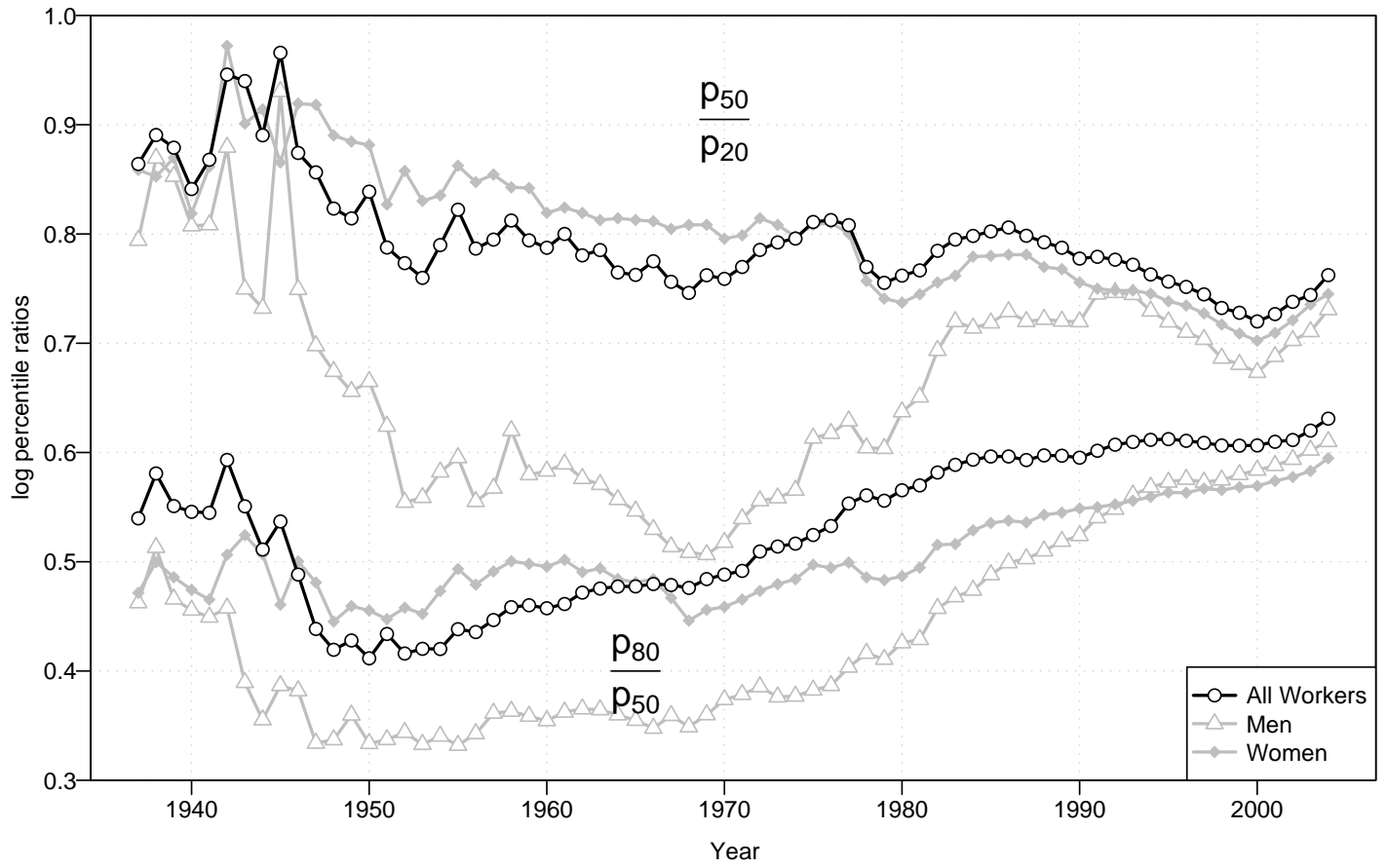


Figure 3: Gini coefficient: Annual Earnings vs 5-Year Earnings

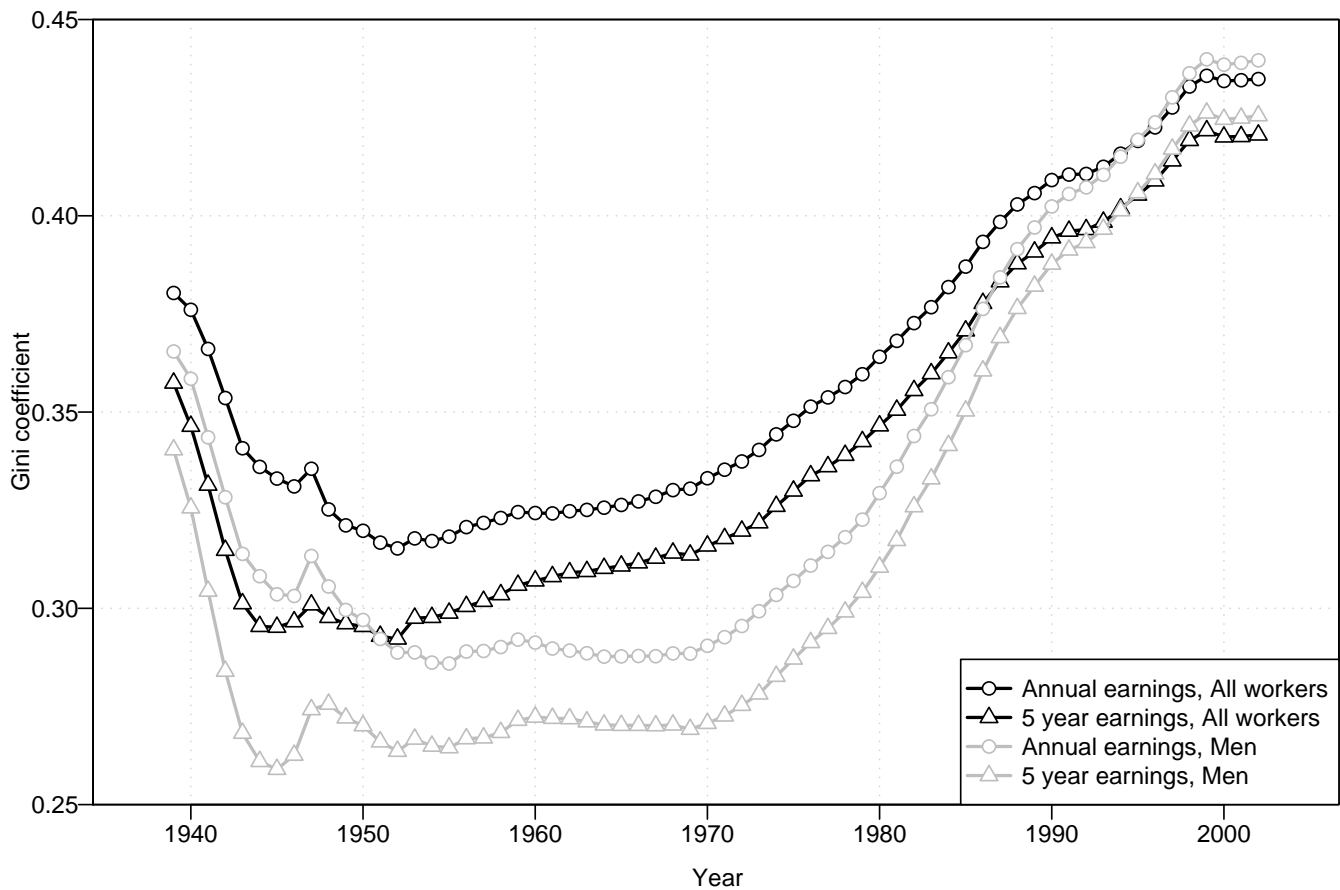


Figure 4: Mobility: Shorrocks Index and Rank Correlation



Figure 5: Variance of Annual, Permanent and Transitory (log) Earnings

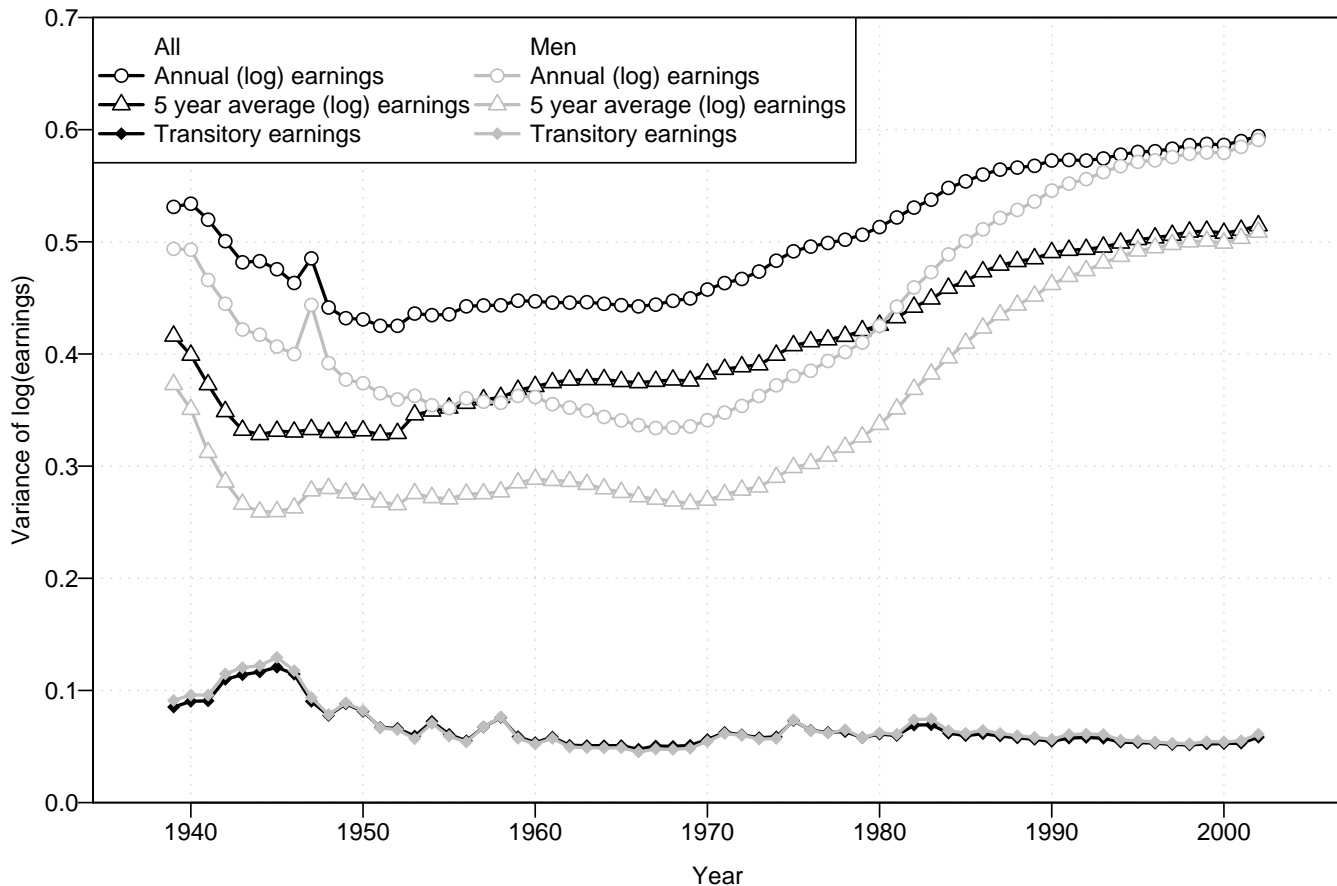


Figure 6A: Top 1% Earnings Share : Annual vs Multi-Year Earnings



Figure 6B: Probability of Staying in Top 1%



Figure 7: Gini coefficient: 11 Year Average Earnings

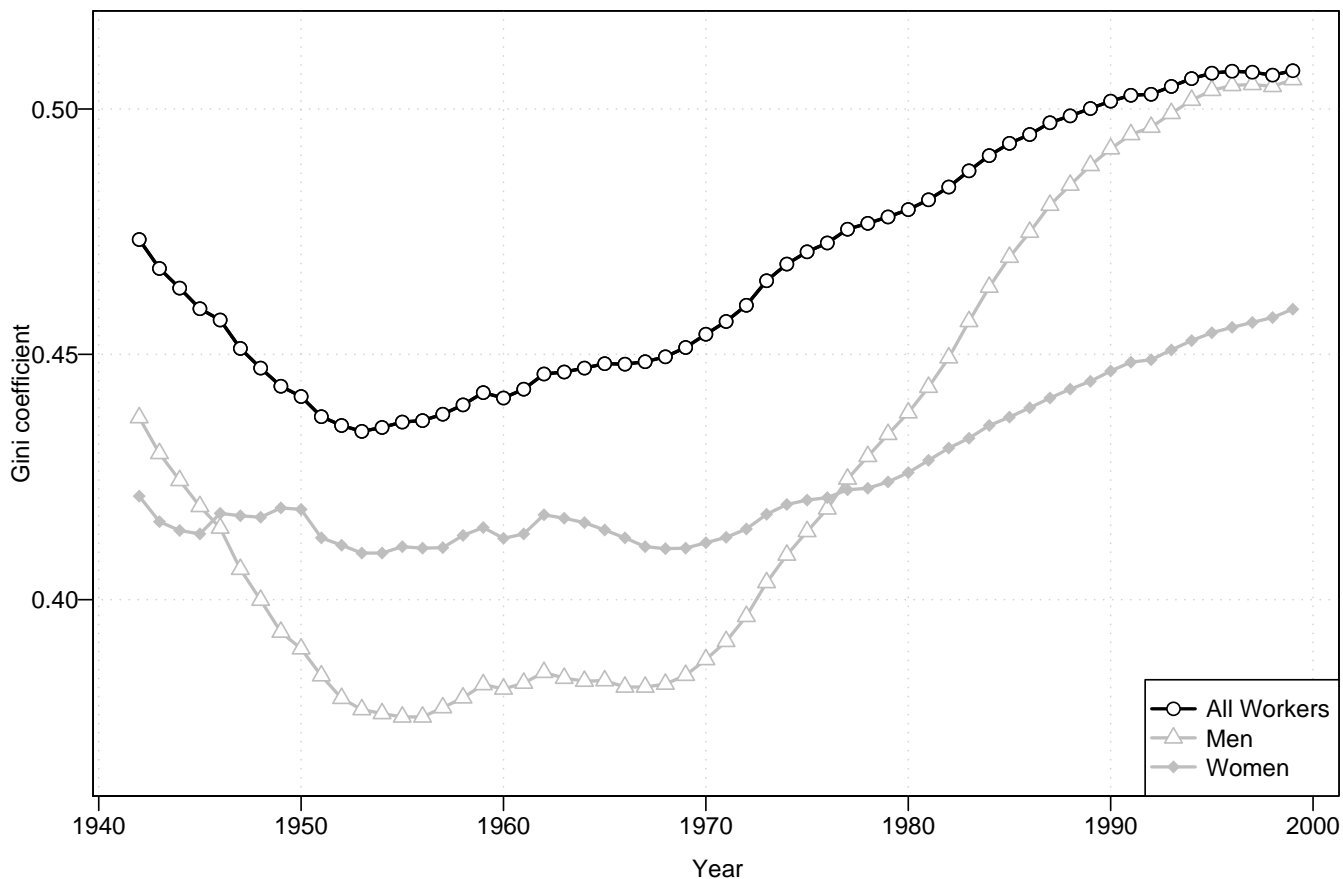


Figure 8: Long-Term Mobility: Rank Correlation between 11 Year Earnings Spans

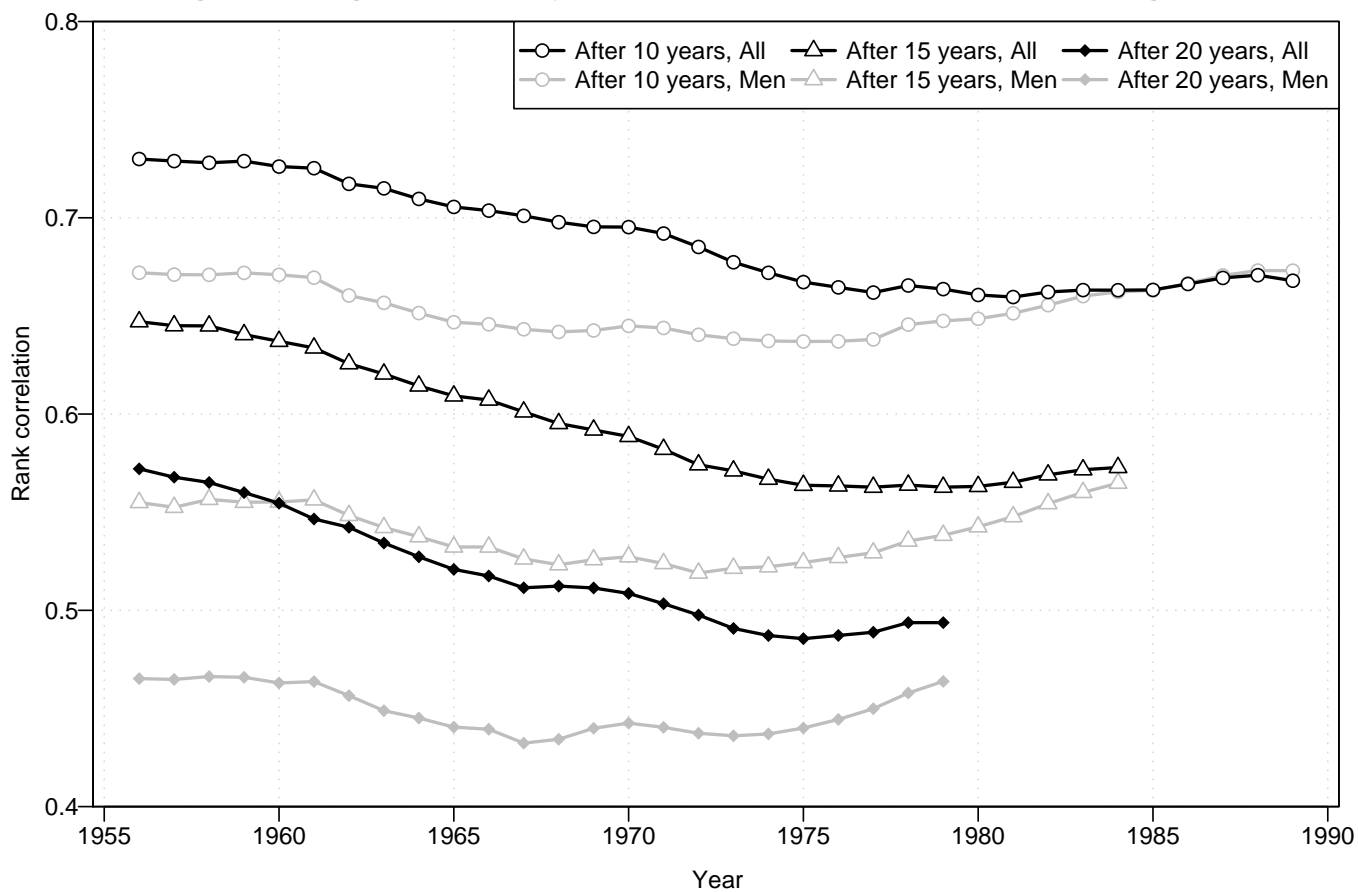


Figure 9: Cohort-Based Long-Term Gini Coefficient

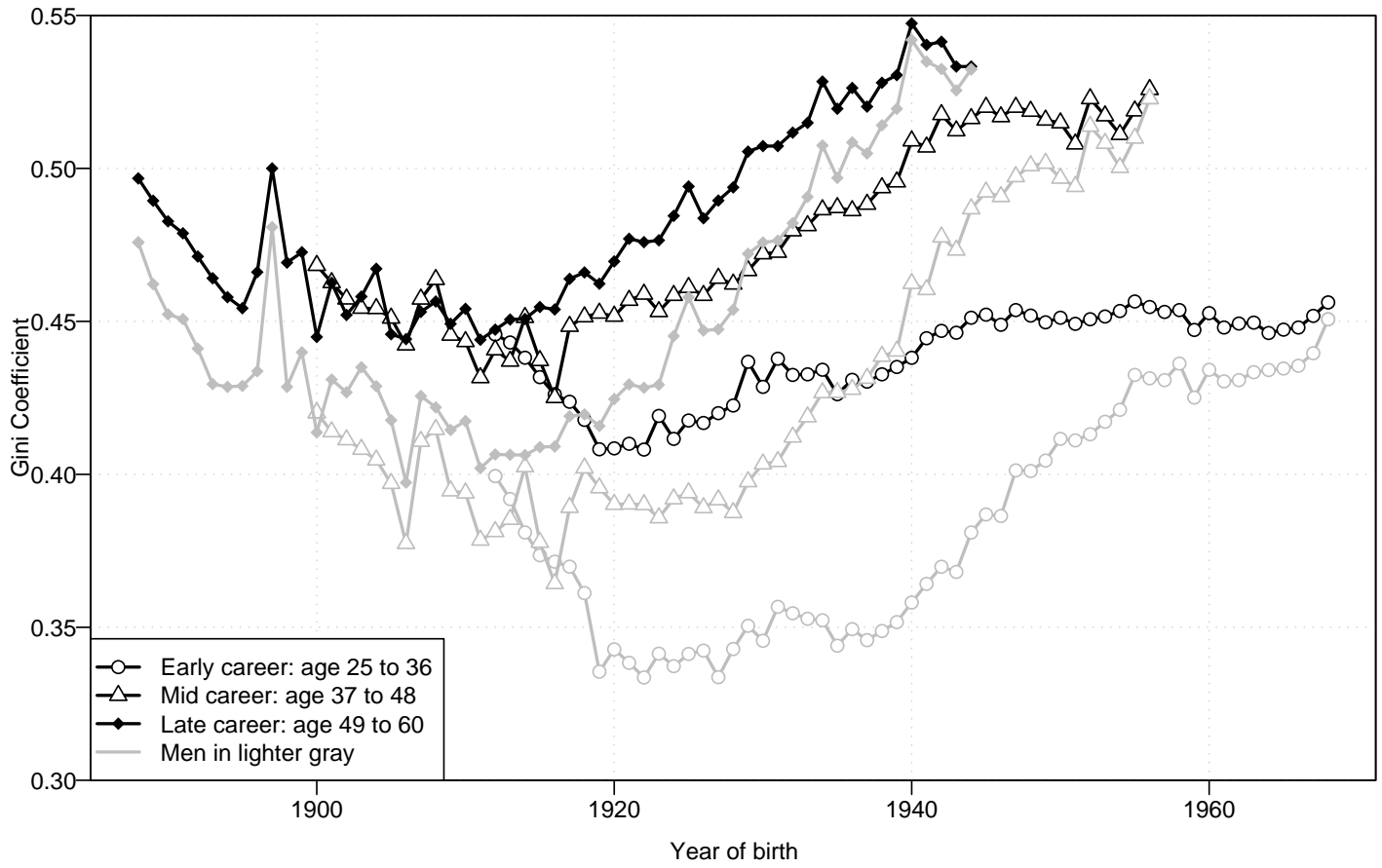


Figure 10: Cohort-Based Long-Term Mobility: Earnings Correlation across Career Stages

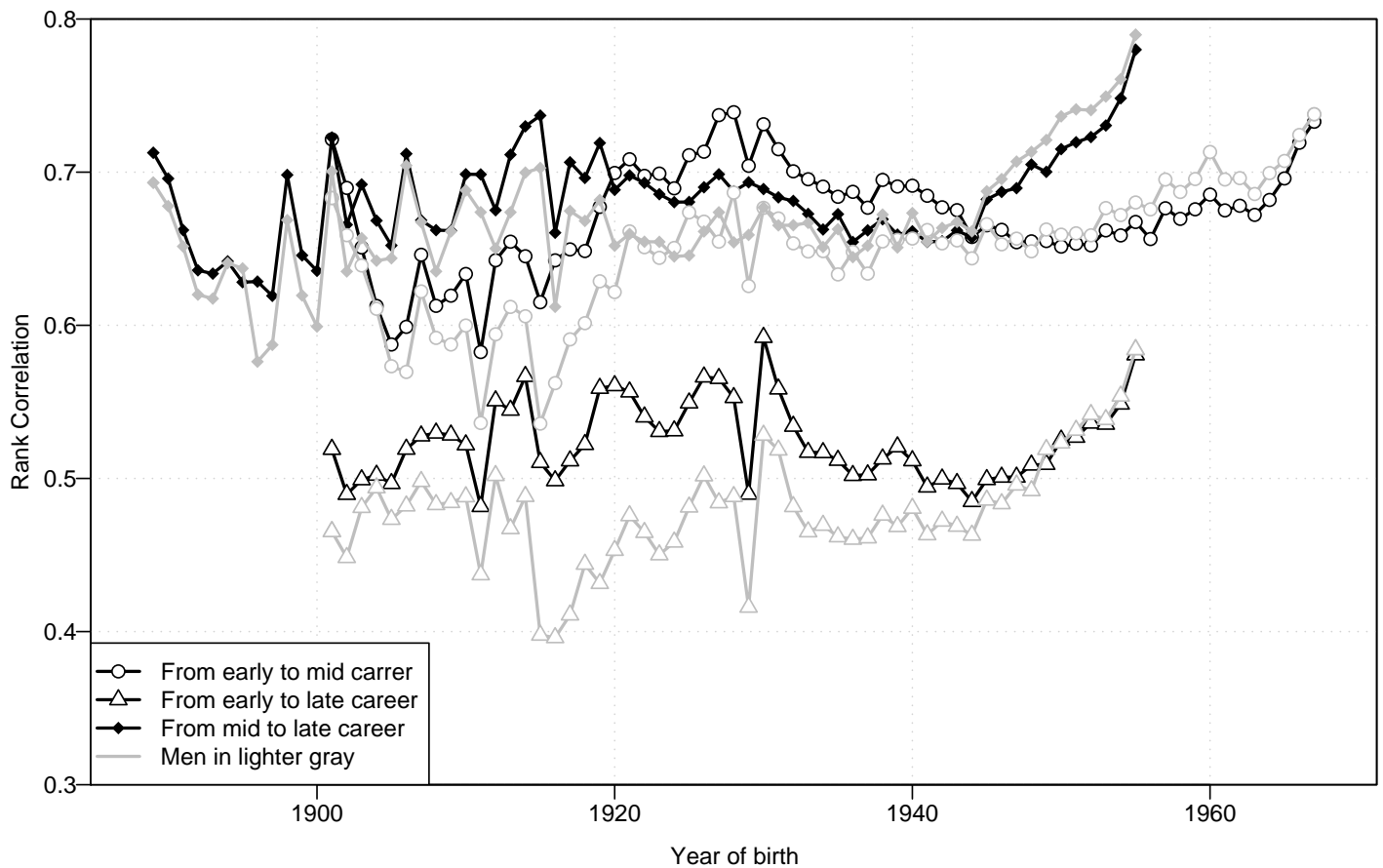


Figure 11A: Gender Gap in Upper Groups

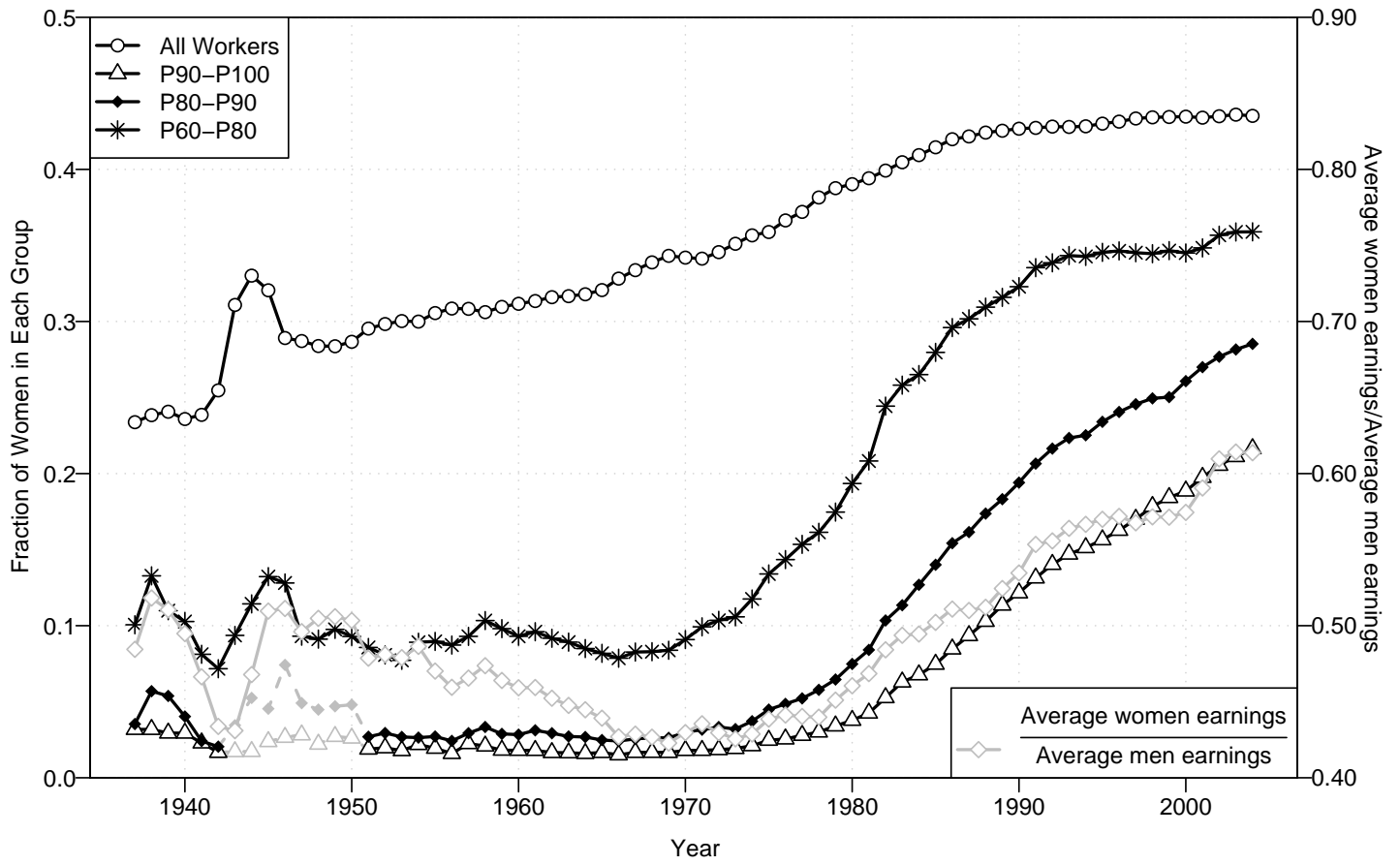


Figure 11B: Gender Gap in Top Groups

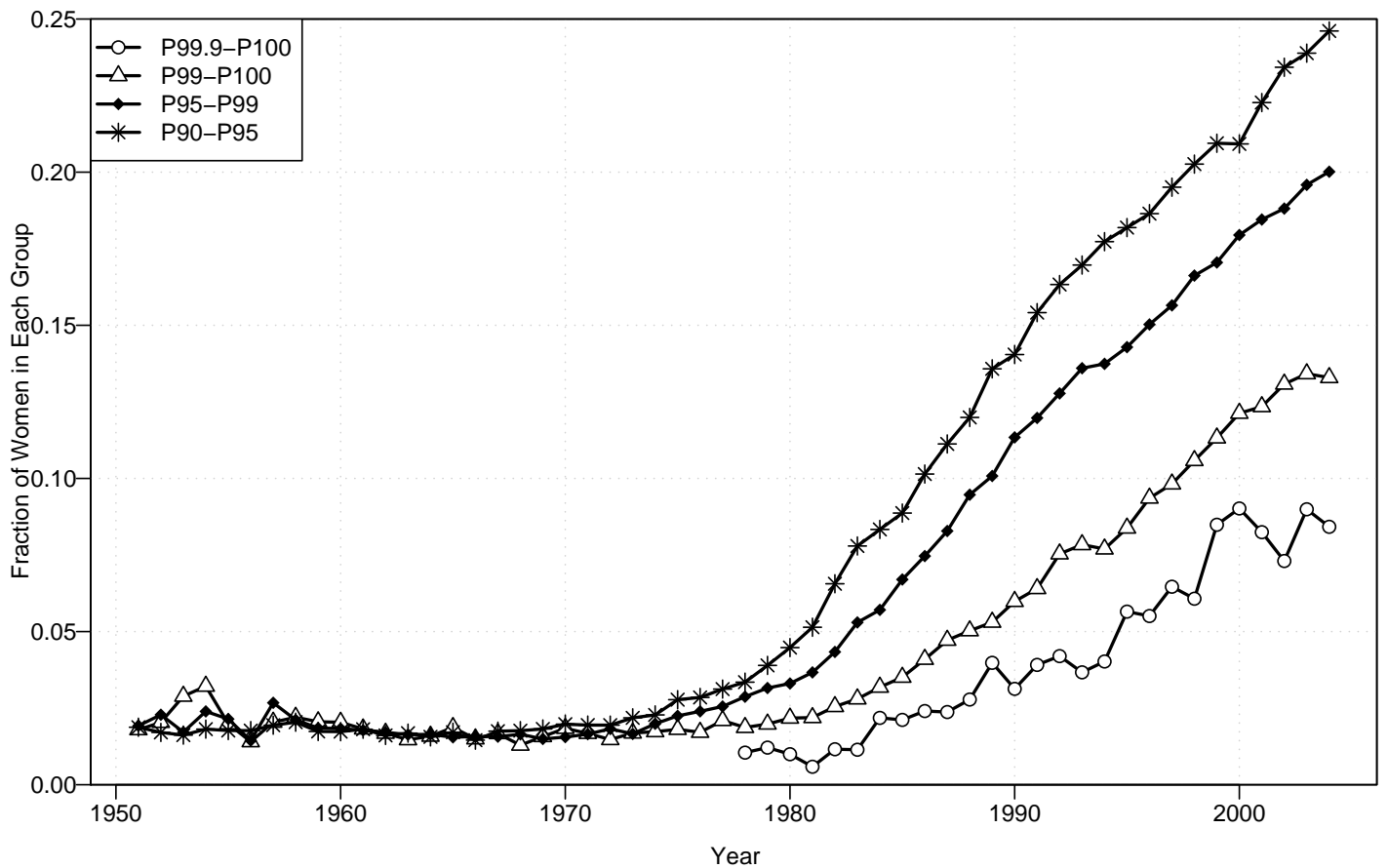


Figure 12: Upward Mobility After 20 Years (11 year averages)

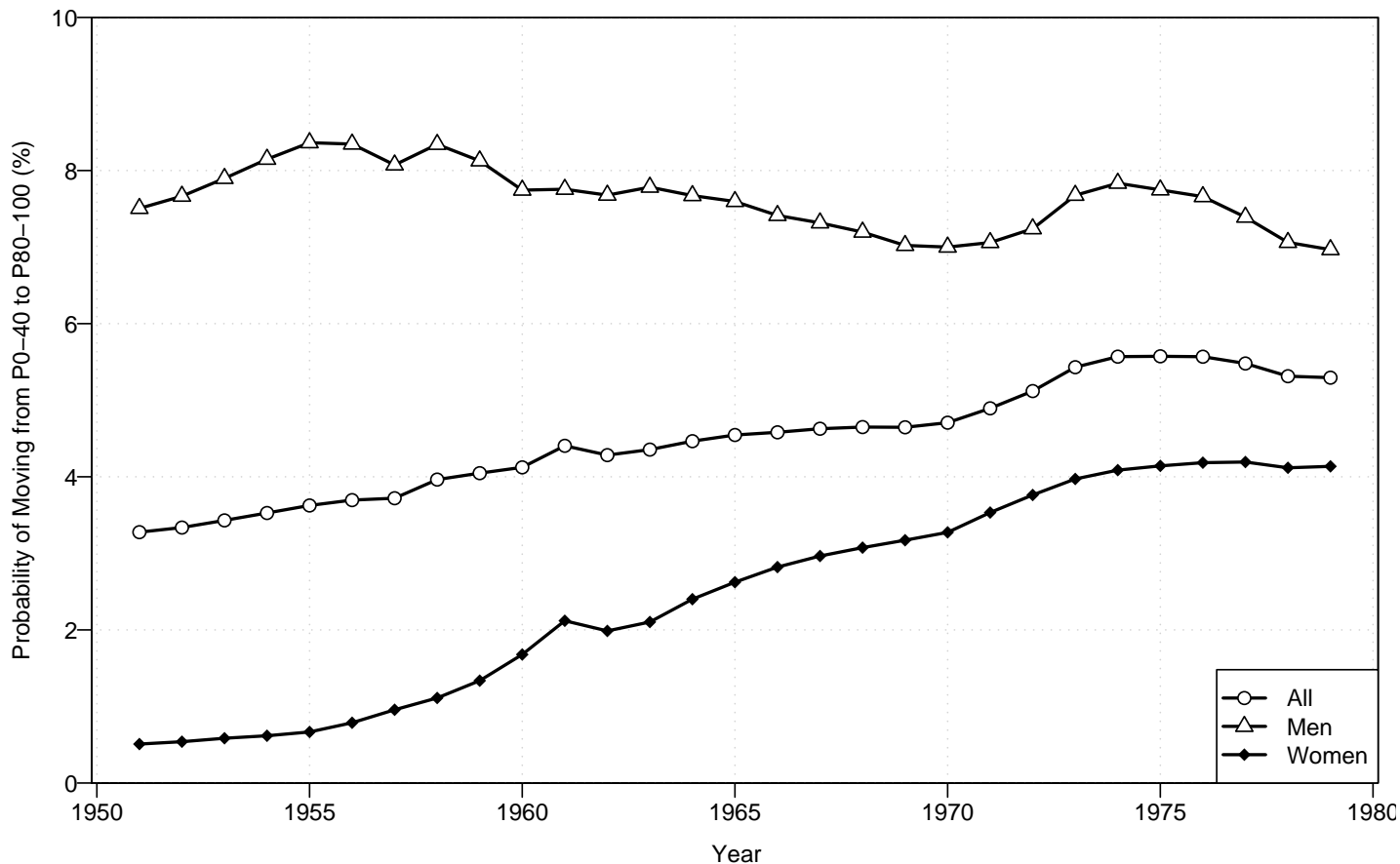


Figure 12: Probability of Upward Mobility over a Career by Gender

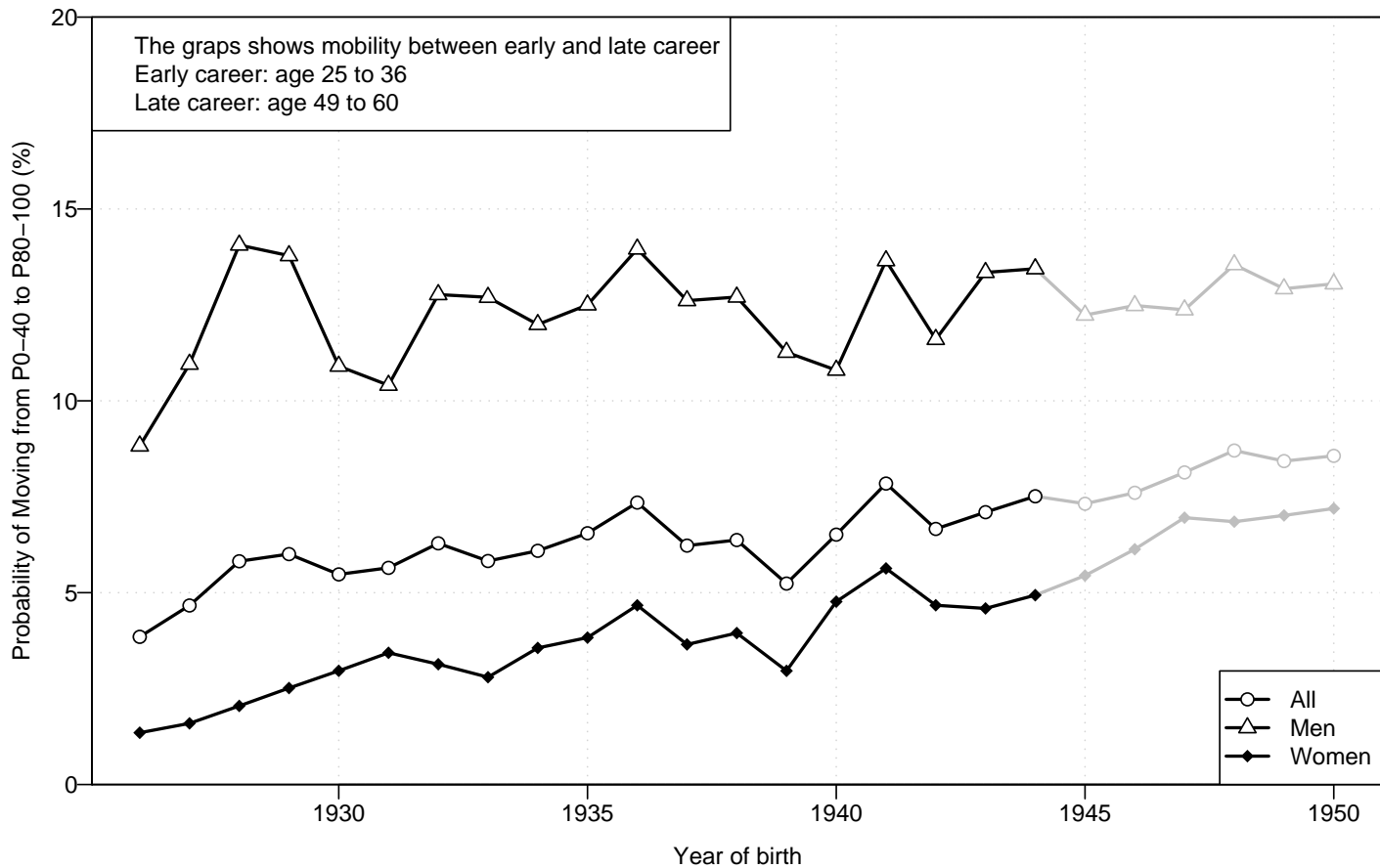


Figure A1: Gini coefficient

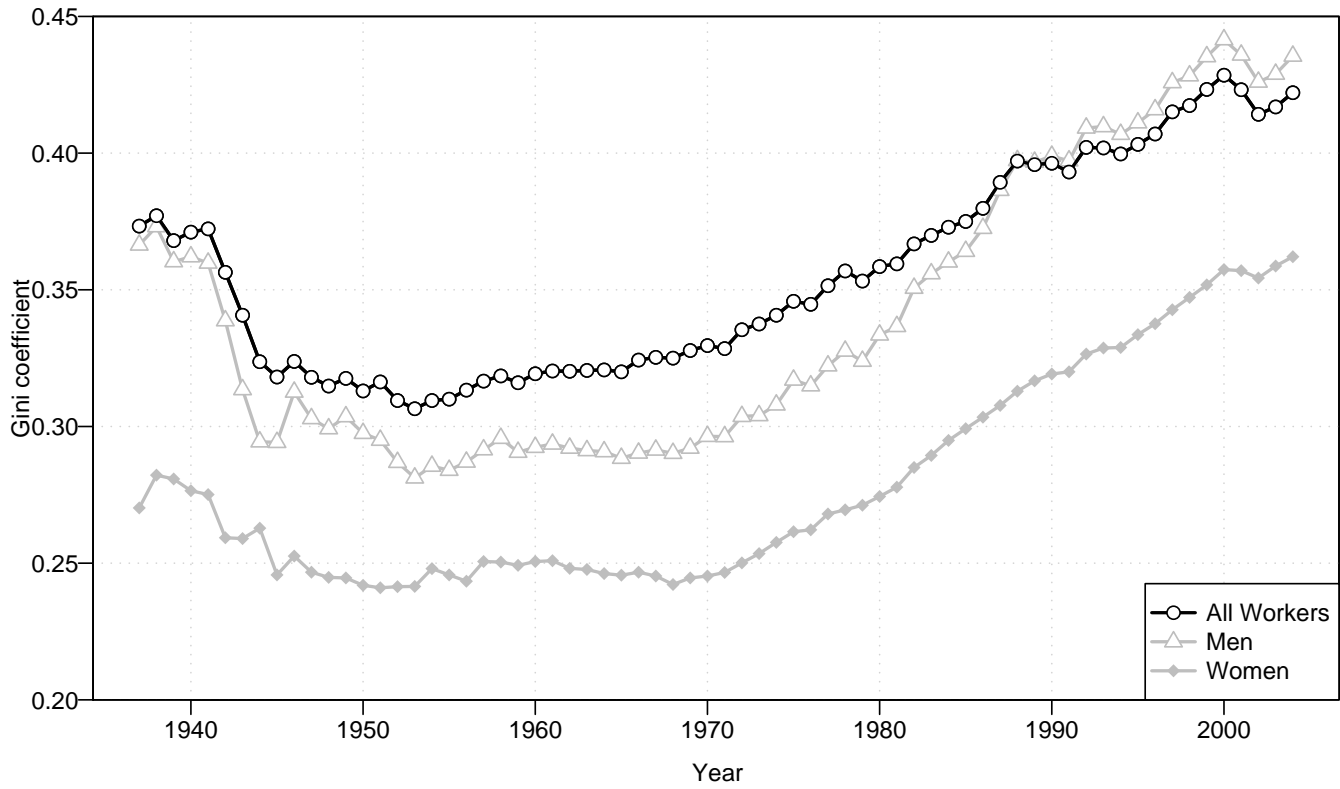


Figure A2: Percentile ratios $\text{Log}(P80/P50)$ and $\text{Log}(P50/P20)$

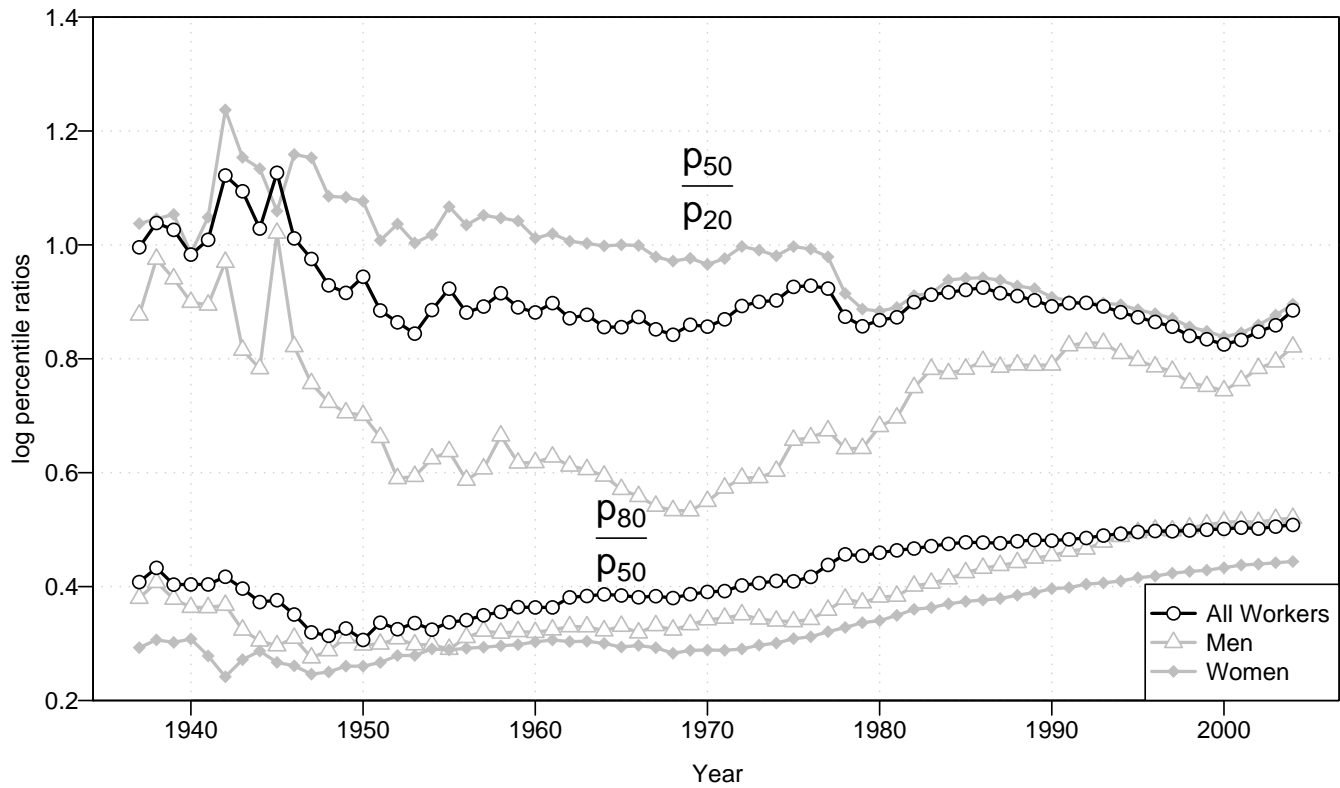


Figure A3: Gini coefficient: Annual Earnings vs 5-Year Earnings

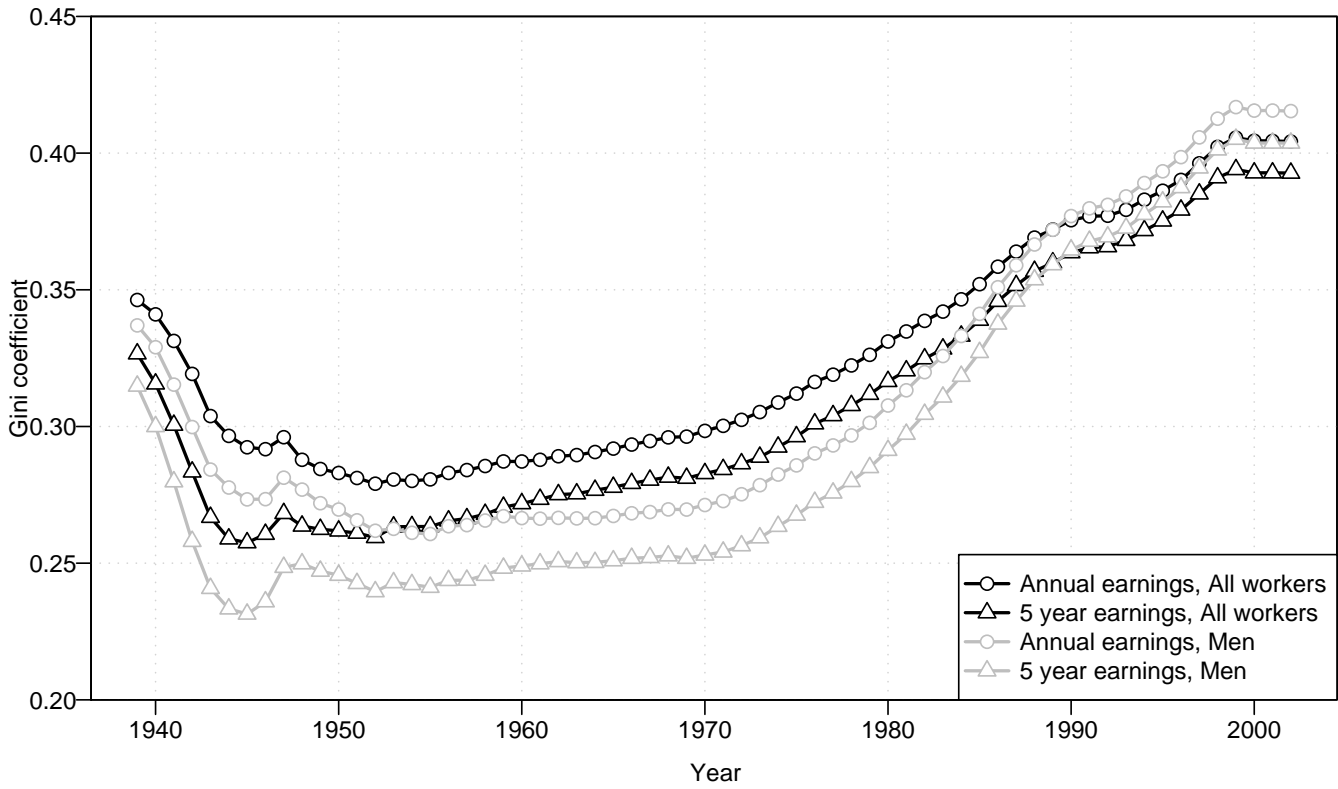


Figure A4: Gini coefficient: 11 Year Average Earnings

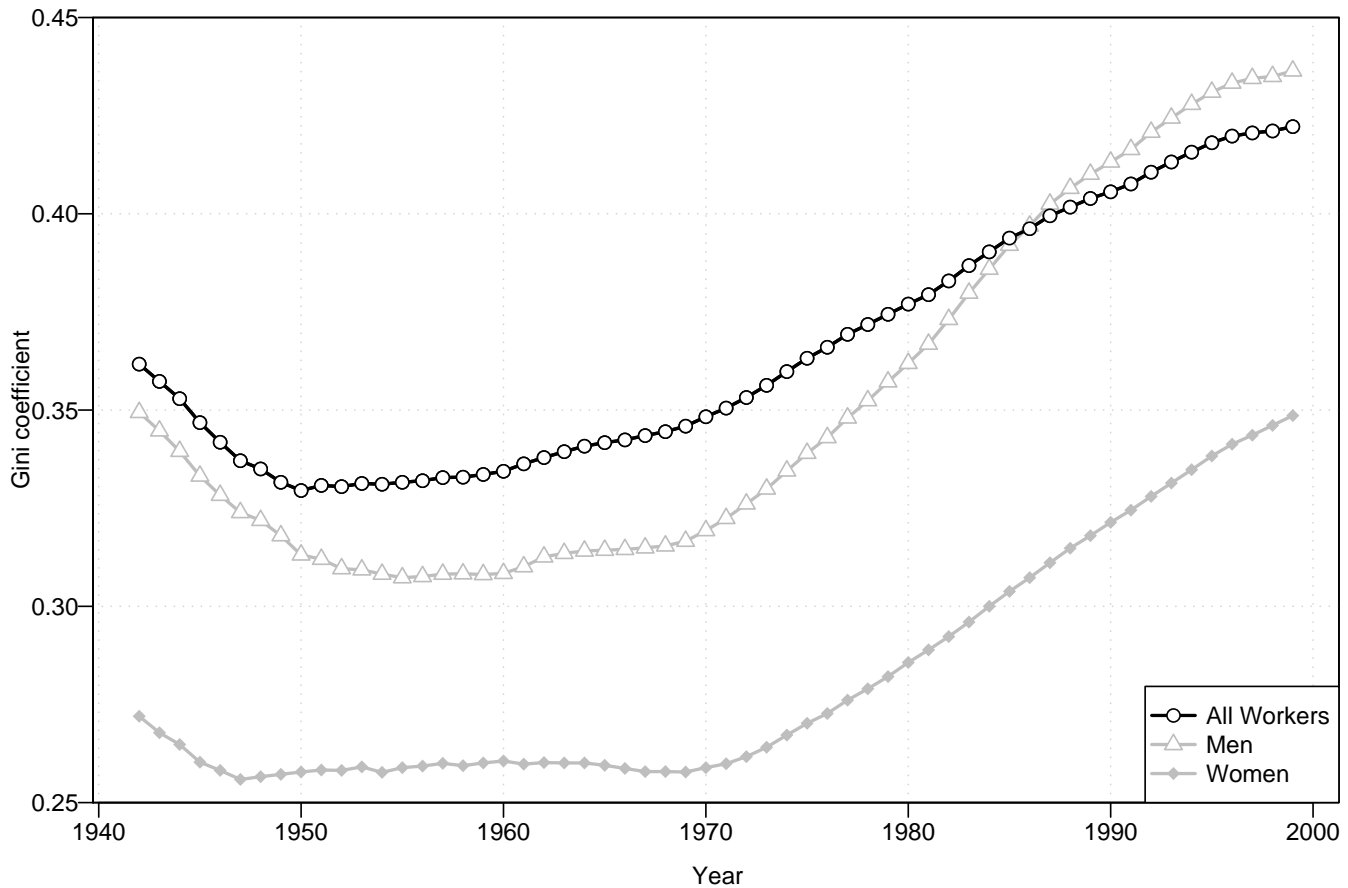


Figure A5: Long-Term Mobility: Rank Correlation between 11 Year Earnings Spans

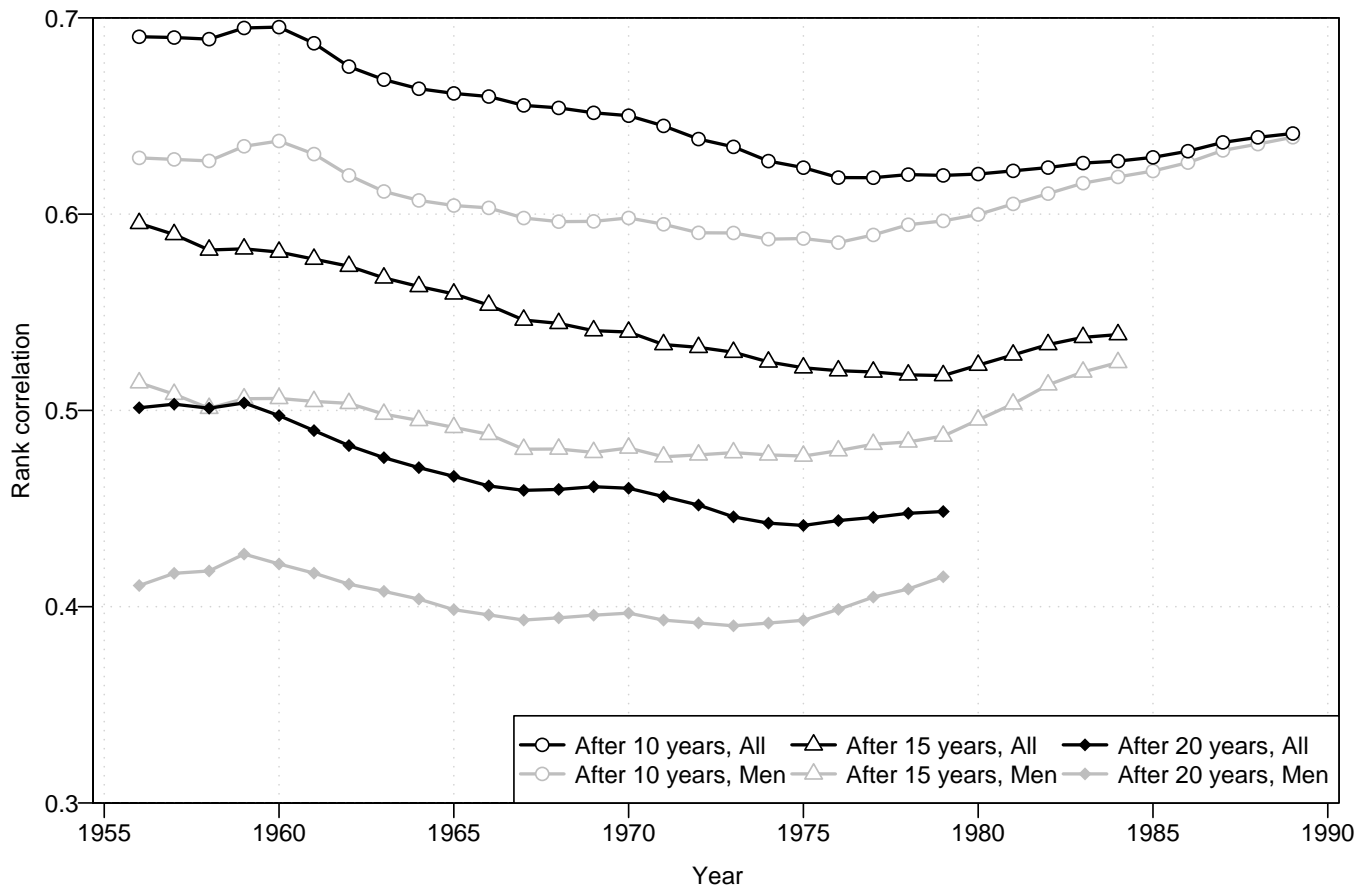


Figure B1: Gini coefficient

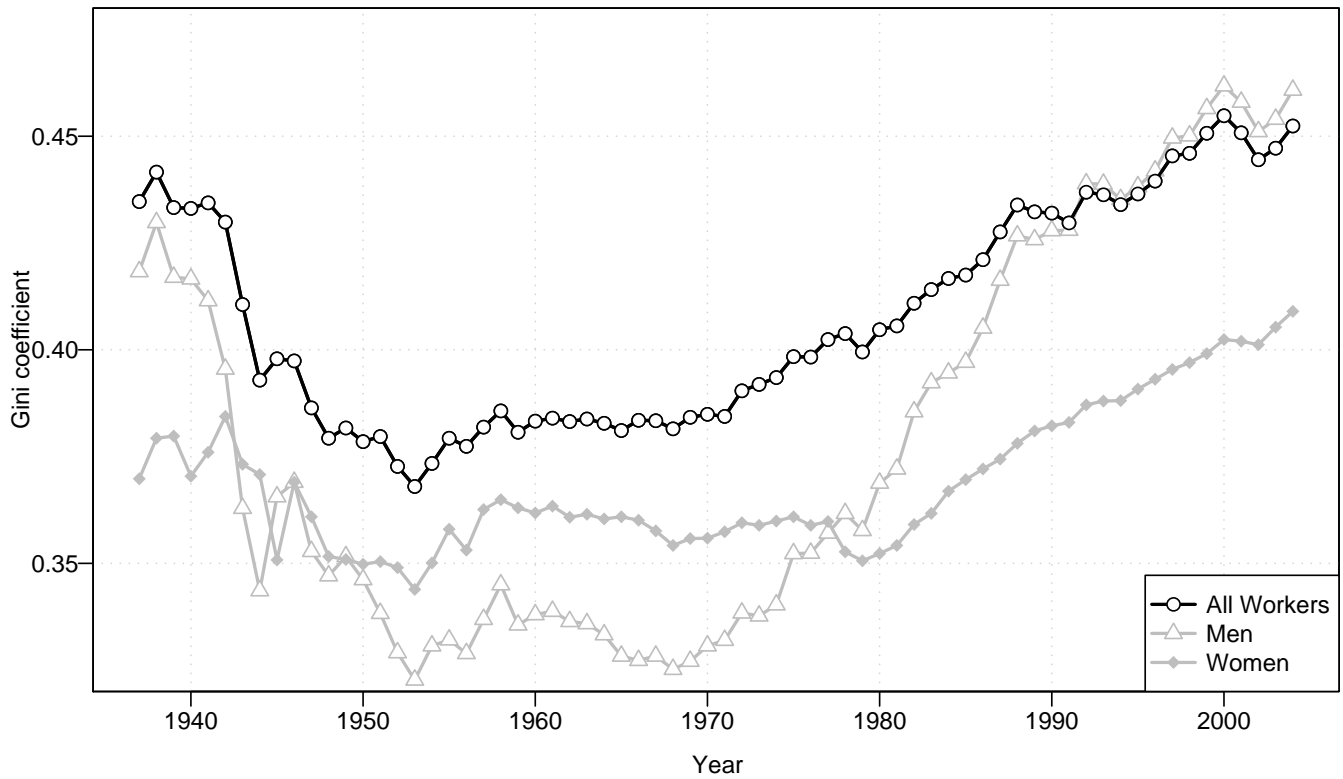


Figure B2: Percentile ratios Log(P80/P50) and Log(P50/P20)

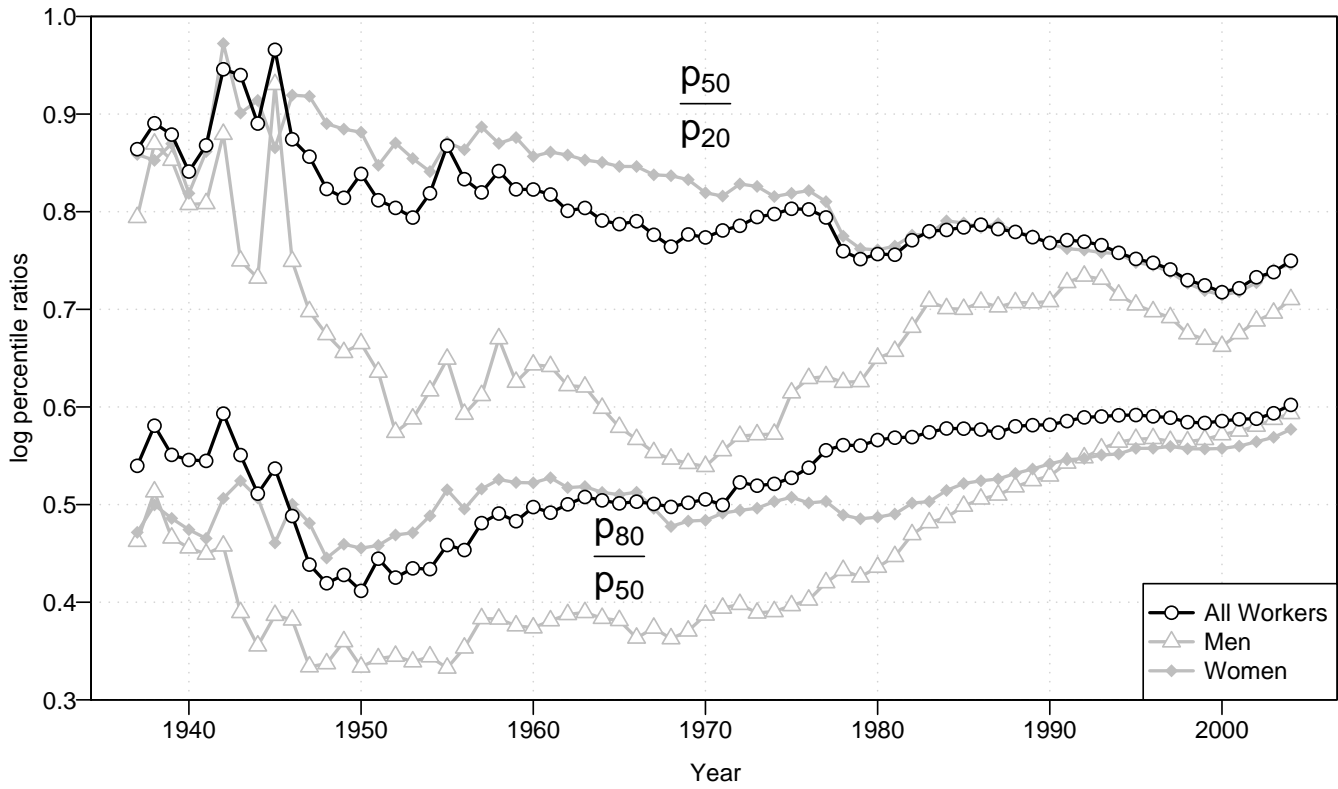


Figure B3: Gini coefficient: Annual Earnings vs 5-Year Earnings

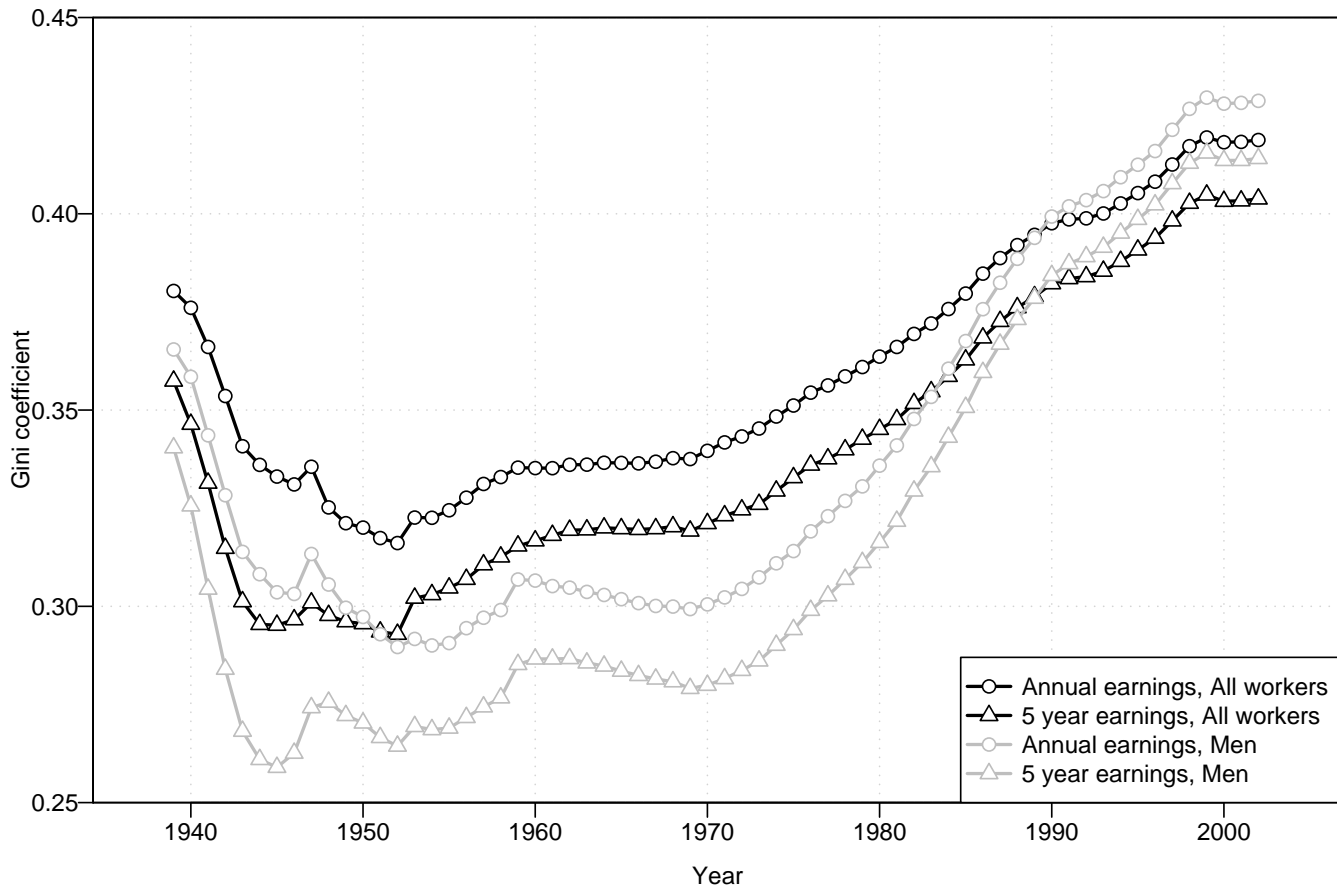


Figure B4: Mobility: Shorrocks Index and Rank Correlation



Figure B5: Variance of Annual, Permanent and Transitory (log) Earnings

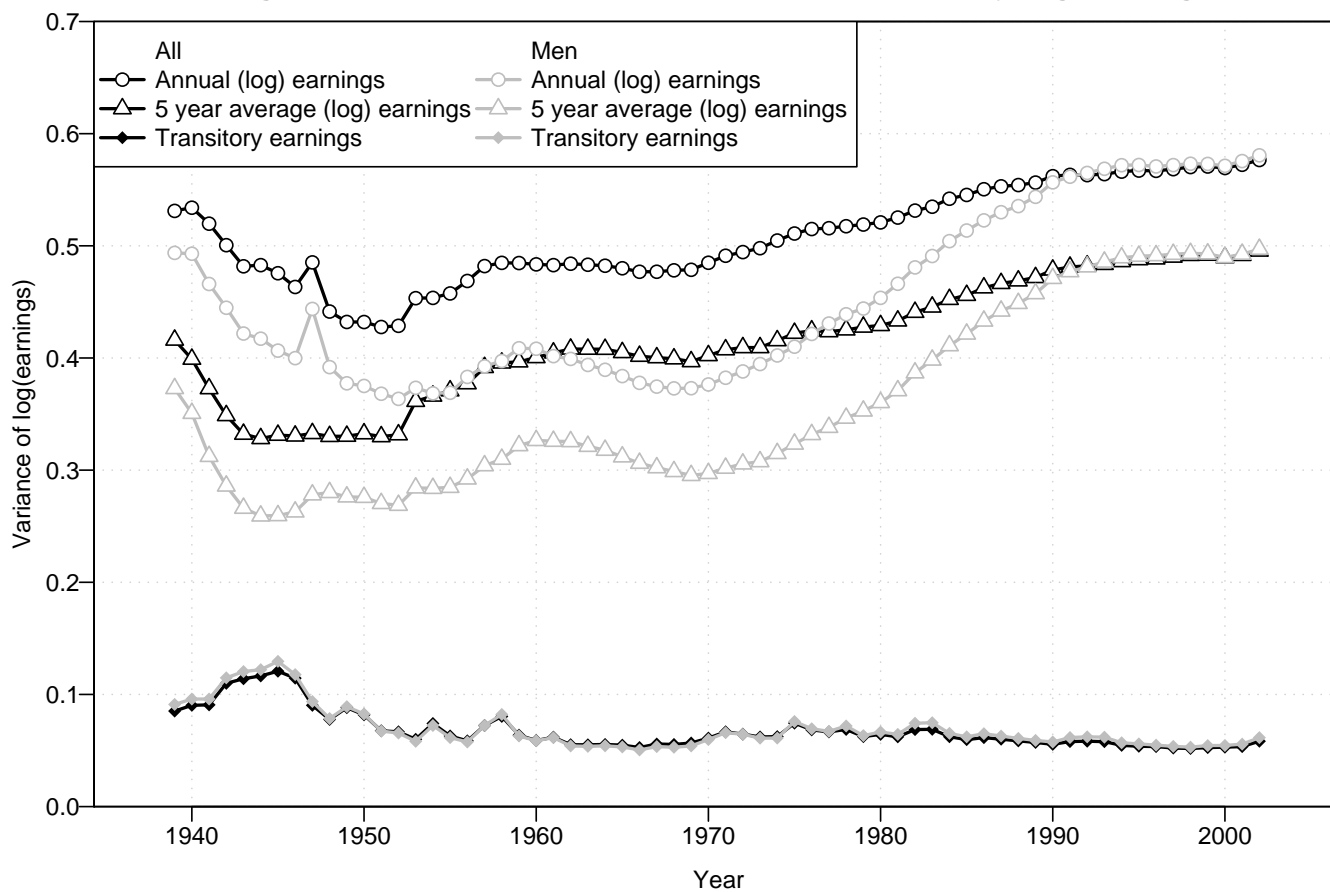


Figure B6: Gini coefficient: 11 Year Average Earnings

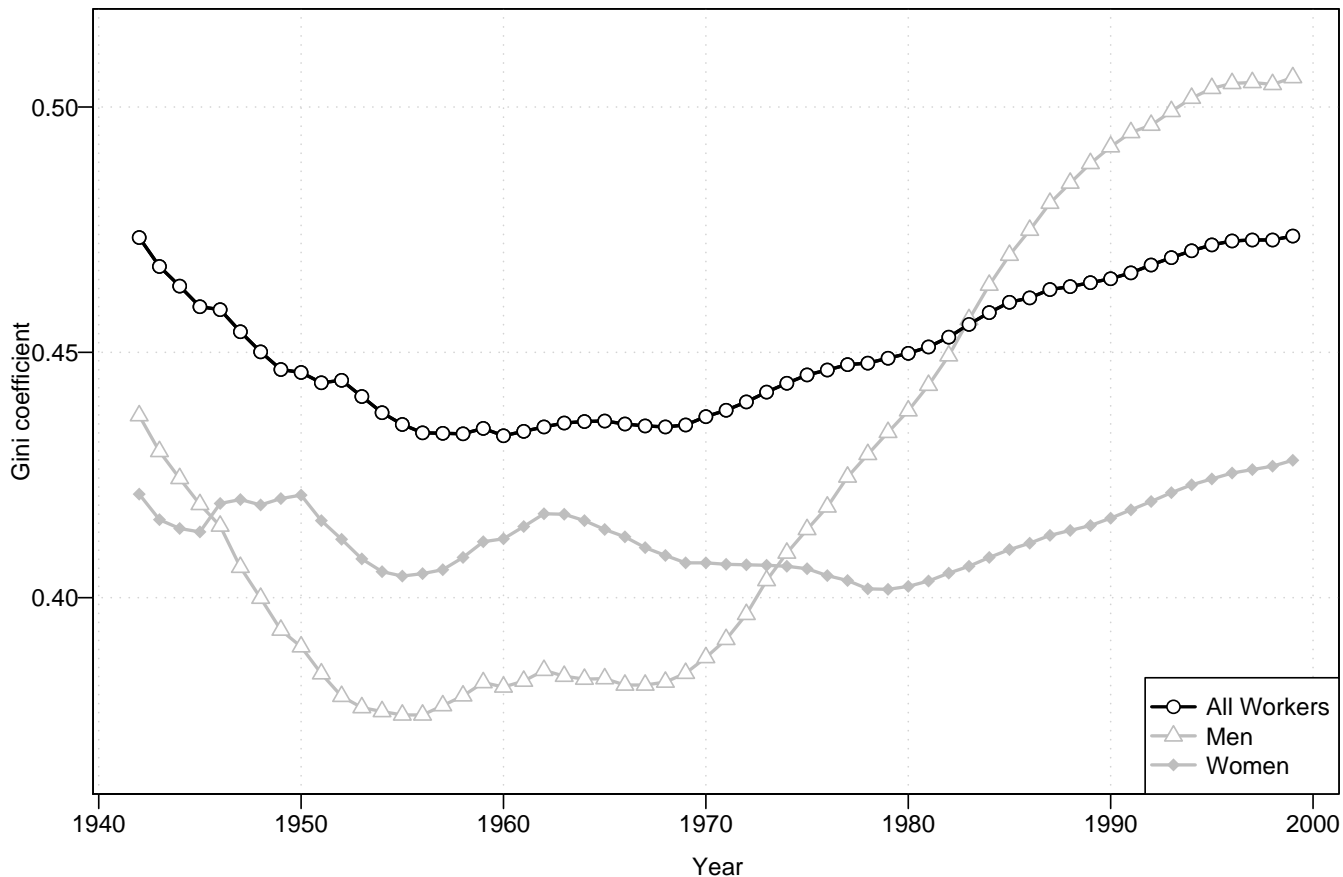


Figure B7: Long-Term Mobility: Rank Correlation between 11 Year Earnings Spans

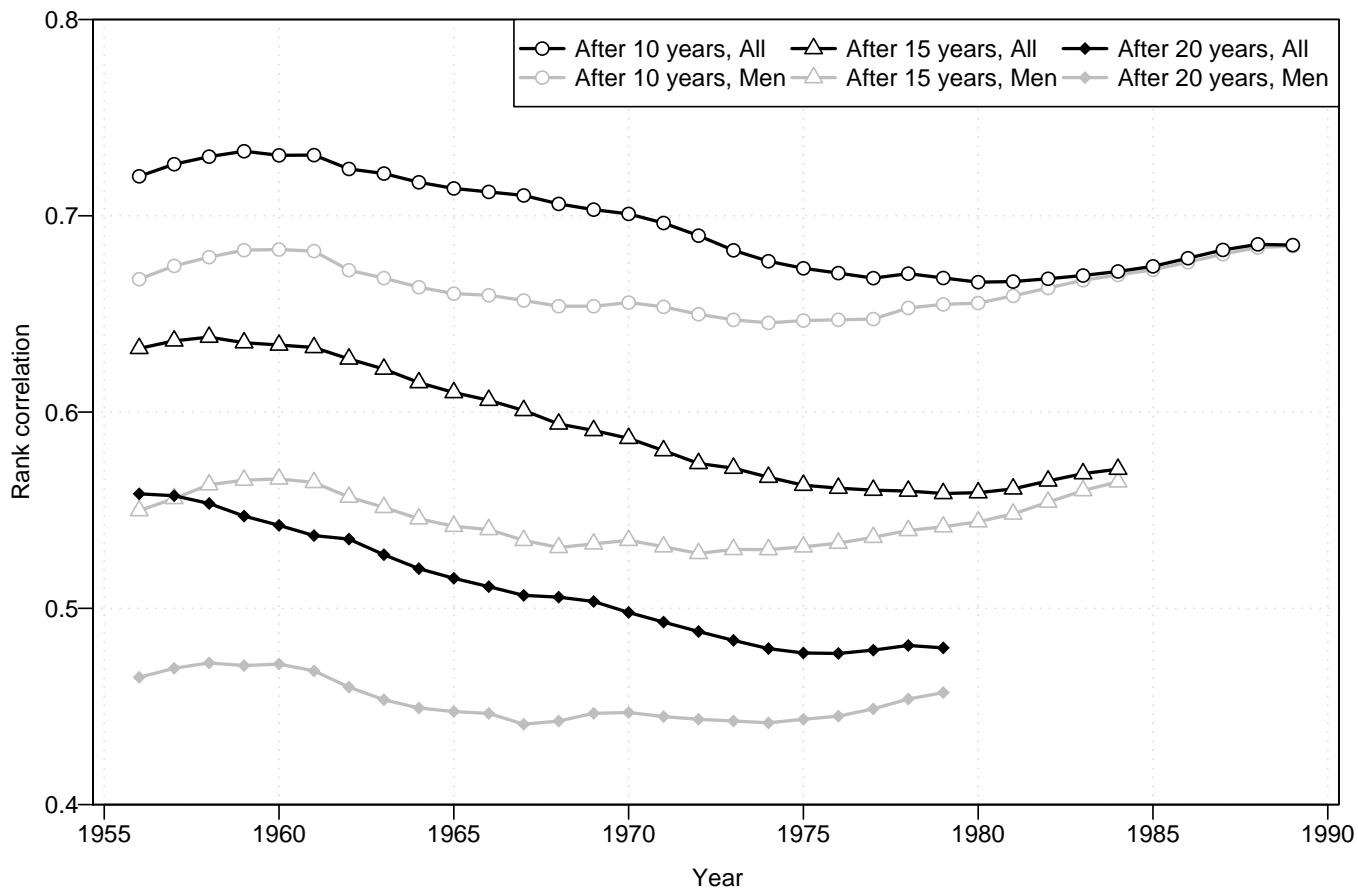


Figure C1: Gini coefficient

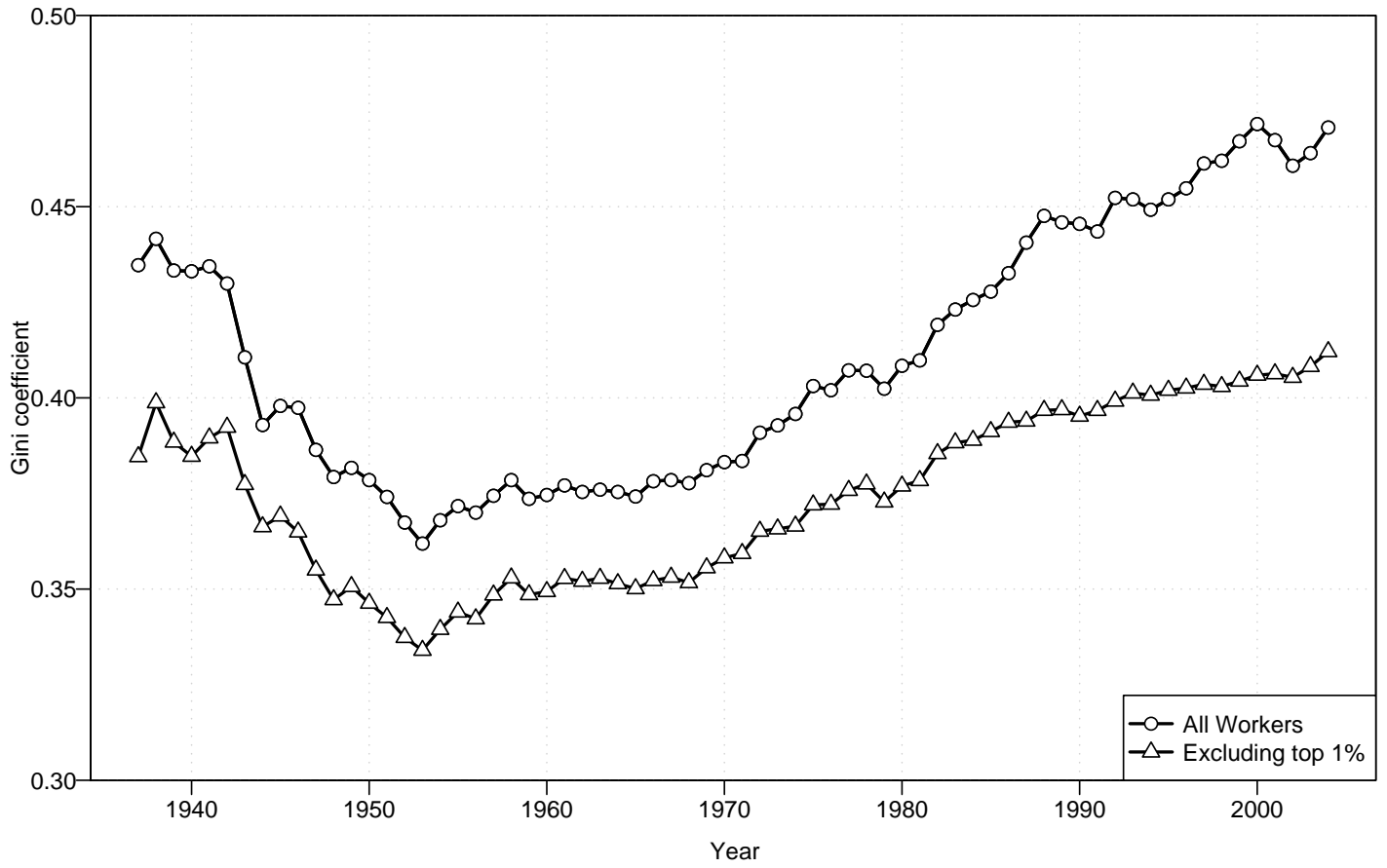


Table 1: Thresholds and Average Earnings by Group in 2004 — Commerce and Industry Sample

Percentile threshold (1)	Earnings threshold (2)	Earnings group (3)	Average earnings (4)	Number of workers (5)
P50	31,245	P0-100	44,052	75,970,900
P0	2,575	P0-20	8,606	15,193,400
P20	14,578	P20-40	20,097	15,189,700
P40	25,500	P40-60	31,393	15,198,300
P60	37,888	P60-80	46,929	15,194,900
P80	58,722	P80-90	68,883	7,597,500
P90	81,530	P90-95	90,788	3,798,500
P95	106,245	P95-99	145,179	3,038,800
P99	235,786	P99-99.5	280,130	379,900
P99.5	343,310	P99.5-99.9	489,854	303,900
P99.9	833,895	Top .1%	2,050,142	76,000

Table 2: Inequality

Year	Gini	Variance log earnings	P80/P20 (logs)	P50/P20 (logs)	P80/P50 (logs)	P0-20	P80-100	P99-100	Average Earnings	#Workers
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
All										
1939	0.433	0.826	1.43	0.88	0.55	3.64	46.82	9.55	17,248	20,404
1960	0.375	0.681	1.24	0.79	0.46	4.54	41.66	5.92	29,908	35,315
1980	0.408	0.730	1.33	0.76	0.57	4.34	44.98	7.21	36,824	50,129
2004	0.471	0.791	1.39	0.76	0.63	3.91	51.41	12.28	44,052	75,971
Men										
1939	0.417	0.800	1.32	0.85	0.47	3.82	45.52	9.58	19,551	15,493
1960	0.326	0.533	0.94	0.58	0.35	5.89	38.80	5.55	35,971	24,309
1980	0.366	0.618	1.06	0.64	0.43	5.25	42.02	6.85	46,646	30,564
2004	0.475	0.797	1.34	0.73	0.61	3.92	51.83	13.44	52,955	42,908
Women										
1939	0.380	0.635	1.36	0.87	0.49	4.49	42.25	6.11	9,979	4,911
1960	0.349	0.570	1.31	0.82	0.50	4.98	39.18	4.05	16,517	11,006
1980	0.354	0.564	1.22	0.74	0.49	5.15	40.38	4.37	21,480	19,566
2004	0.426	0.693	1.34	0.74	0.59	4.45	47.36	8.00	32,499	33,063

0.1% sample from 1937 to 1956, 1% from 1957 to 2004. Number of workers in thousands.

Table 3: Short-Term Mobility

Year	Gini MA5	Gini MA1 (average)	Rank Correlation After 1 Year	Variance of MA5 log earnings	Variance of MA1 log earnings (average)	Variance log deviation	#Workers
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All							
1939	0.357	0.380	0.859	0.416	0.531	0.085	14,785
1960	0.307	0.324	0.883	0.371	0.447	0.054	26,479
1980	0.347	0.364	0.885	0.426	0.513	0.061	35,500
2002	0.421	0.435	0.897	0.514	0.594	0.058	55,108
Men							
1939	0.340	0.365	0.853	0.373	0.494	0.091	11,700
1960	0.272	0.291	0.855	0.288	0.362	0.052	19,577
1980	0.310	0.329	0.869	0.337	0.425	0.062	23,190
2002	0.426	0.440	0.898	0.509	0.591	0.061	32,259

0.1% sample from 1937 to 1956, 1% from 1957 to 2004. Number of workers in thousands.

Table 4: Long-Term Mobility

Year	Gini MA11	Rank Correlation After 20 Years	Upward Mobility After 20 Years	#Workers
(1)	(2)	(3)	(4)	(5)
All				
1956	0.437	0.572	0.037	42,753
1978	0.477	0.494	0.053	61,828
1999	0.508			94,930
Men				
1956	0.376	0.465	0.084	27,952
1978	0.429	0.458	0.071	37,187
1999	0.506			52,761
Women				
1956	0.410	0.361	0.008	14,801
1978	0.423	0.358	0.041	24,641
1999	0.459			42,169

0.1% sample from 1937 to 1956, 1% from 1957 to 2004. Number of workers in thousands.