Appendices for "The Glass Ceiling and The Paper Floor: Gender Differences Among Top Earners, 1981–2012"*

A Details of Decompositions

In this appendix, we provide details of the methodology underlying the decompositions presented in Table 1, Table 3, Table 7 and Table 8.

We start by establishing some notation. Let G_{it} be the gender of individual i who is included in our sample in year t, with the convention that $G_{it} = 1$ for a female and $G_{it} = 0$ for a male. Let p denote a percentile range (e.g. top 0.1 percent, second 0.9 percent or bottom 99 percent) and let D_{it}^p be an indicator variable that takes the value 1 if individual i is in the percentile range p of the earnings distribution in year t. Let σ_t^p be the fraction of top earners that are female.

$$\sigma_t^p = E_t\left[G|D^p=1\right] \tag{1}$$

Let E_t denote a moment of a time t distribution and let P_t denote a probability based on the time t distribution.

A.1 Decomposition for changing gender composition of the labor force (Table 1)

The goal is to measure how much of the observed change in σ_t^p is due to a changes in the share of females in the labor force $E_t[G]$. Using Bayes' rule we can decompose σ_t^p as

$$\sigma_t^p = \frac{P_t\left[D^p=1|G=1\right]P_t\left[G=1\right]}{P_t\left[D^p=1\right]} \tag{2}$$

$$\sigma_t^p P_t\left[D^p=1\right] = E_t\left[D^p|G=1\right]E_t\left[G\right] \tag{3}$$

$$\Delta\left(\sigma_t^p P_t\left[D^p=1\right]\right) = E_t\left[D^p|G=1\right]\left(\Delta E_t\left[G\right]\right) + \left(\Delta E_t\left[D^p|G=1\right]\right)E_{t-1}\left[G\right] \tag{4}$$

*Fatih Guvenen, University of Minnesota and NBER (guvenen@umn.edu), Greg Kaplan, Princeton University and NBER (gkaplan@princeton.edu) and Jae Song, Social Security Administration (jae.song@ssa.gov)

The term on the LHS of (4) is the change in the fraction of the workforce that are female and in percentile group p. The first term on the RHS of (4) is the component of this change that is due to changes in the share of females in the labor force. The second term on the RHS is the component that is due to changes in the fraction of females that are in percentile group p. We implement this decomposition for each pair of consecutive years using sample analogues of the moments in (4) and then summing the components over all years to get the total decomposition.

In principal $P_t\left[D^p=1\right]$ is constant for all t, since it is simply the fraction of the population in percentile group p. However, since we take different size random samples for the top percentile groups compared with the bottom 99 percent, in practice there are small year-to-year fluctuations in our sample estimates of this moment. If $P_t\left[D^p=1\right]$ were constant then the fraction of $\Delta\sigma_t^p$ that is due to changes in the gender composition of the labor force would be given by

$$\frac{E_t\left[D^p|G=1\right]\Delta E_t\left[G\right]}{P_t\left[D^p=1\right]\Delta\sigma_t^p} \tag{5}$$

With our decomposition the fraction is given by

$$\frac{E_t\left[D^p|G=1\right]\Delta E_t\left[G\right]}{P_t\left[D^p=1\right]\Delta\sigma_t^p+\sigma_{t-1}^p\Delta P_t\left[D^p=1\right]} \tag{6}$$

Since the term $\sigma_{t-1}^p\Delta P_t\left[D^p=1\right]$ is very small relative to $P_t\left[D^p=1\right]\Delta\sigma_t^p$, this sampling variation has a negligible effect on the results of the decomposition.

A.2 Decomposition for changing for age and industry composition (Table 7, Table 8)

The goal is to measure how much of the observed change in σ_t^p is due to a changes in the distribution of an observable characteristic X_{it}. We consider only characteristics that which take a discrete set of values such as age and industry. Analogously to the decomposition

above we can write

$$\begin{aligned}
\sigma_t^p P_t\left[D^p=1\right] &= E_t\left[D^p|G=1\right]E_t\left[G=1\right] \\
&= \sum_x E_t\left[D^p|G=1,X=x\right]P_t\left[X=x|G=1\right]E_t\left[G\right] \\
&= \sum_x E_t\left[D^p|G=1,X=x\right]E_t\left[G|X=x|\right]P_t\left[X=x\right] \qquad (7) \\
\Delta\left(\sigma_t^p P_t\left[D^p=1\right]\right) &= \sum_x E_t\left[D^p|G=1,X=x\right]\Delta E_t\left[G|X=x\right]P_t\left[X=x\right] \\
&\quad +\sum_x \Delta E_t\left[D^p|G=1,X=x\right]E_{t-1}\left[G|X=x\right]P_t\left[X=x\right] \\
&\quad +\sum_x E_{t-1}\left[D^p|G=1,X=x\right]E_{t-1}\left[G|X=x\right]\Delta P_t\left[X=x\right] \qquad (8)
\end{aligned}$$

The term on the LHS of (8) is the change in the fraction of the workforce that are female and in percentile group p. The first term on the RHS is the component of this change that is due to changes in the gender composition of different categories (i.e. industries or age groups).The second term on the RHS is the component that is due to changes in the fraction of females in each category that are in percentile group p. The third term on the RHS is the component that is due to changes in the fraction of the overall labor force in each category of X.

A.3 Decomposition for changes in mobility (Table 3)

The goal is to measure how much of the observed change in σ_t^p is due to changes in the transition probabilities in and out of the percentile group p. Let D_+^p be an indicator variable that takes the value 1 if an individual was in percentile group p in year $t+1$. Since gender is constant over time, $G_t = G_{t-1}$, we can decompose σ_t^p using the relationship that

$$\begin{aligned}
\sigma_t^p P_t\left[D^p=1\right] &= E_t\left[D^p|G=1\right]E_t\left[G=1\right] \\
&= \sum E_{t-1}\left[D_+^p|G=1,D^q=1\right]E_{t-1}\left[D^q|G=1\right]E_{t-1}\left[G=1\right] \\
&= \sum E_{t-1}\left[D_+^p|G=1,D^q=1\right]E_{t-1}\left[G|D^q=1\right]E_{t-1}\left[D^q\right] \qquad (9)
\end{aligned}$$

Then taking first differences yields

$$\begin{aligned}
\Delta\left(\sigma_t^p P_t\left[D^p=1\right]\right) &= \sum_q E_{t-1}\left[D_+^p|G=1,D^q=1\right]\Delta E_{t-1}\left[G|D^q=1\right]E_{t-1}\left[D^q\right] \\
&\quad +\sum_q \Delta E_{t-1}\left[D_+^p|G=1,D^q=1\right]E_{t-2}\left[G|D^q=1\right]E_{t-1}\left[D^q\right] \\
&\quad +\sum_q E_{t-2}\left[D_+^p|G=1,D^q=1\right]E_{t-2}\left[G|D^q=1\right]\Delta E_{t-1}\left[D^q\right] \qquad (10)
\end{aligned}$$

The term on the LHS of (10) is the change in the fraction of the workforce that are female and in percentile group p. The first term on the RHS is the component of the change that is due to changes in the female share of top percentiles in the previous period at the prevailing levels of persistence. The second term on the RHS is the component of this change that is due to changes in the transition probabilities into the top p-the percentile. The third term is due to sampling variation and is a negligible component of the overall change; we present the decomposition for the change net of the effects of this term.

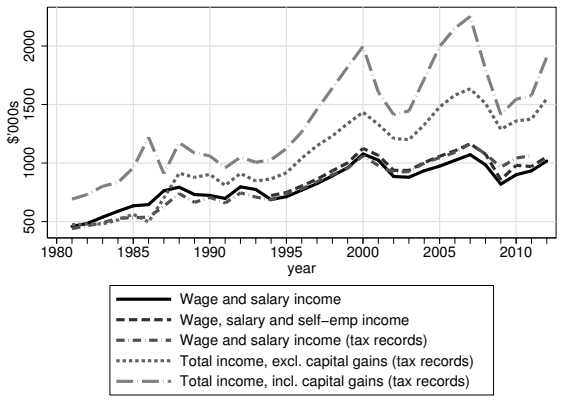
The idea behind this decomposition is that any one-time change in transition probabilities will lead to continued changes in the fraction of females in the top percentiles in subsequent years, even if there are no further changes in the transition probabilities. Hence any observed change is partly due to the effects of changes in the transition probabilities in the past as the system moves towards its new stationary distribution, and is partly due to new changes in the transition probabilities. The first term captures the former effect, the second term captures the latter effect.

B Comparison with alternative definitions of income

Figure B.1A and Figure B.1B plot the trends for the 99.9th percentile and 99th percentile, under various definitions of income, using our data and the data from aggregate tax records from Saez (2012). Note that in our data, the unit of observation is an individual, but in Saez (2012) the unit of observation is a tax unit, defined as married couples plus dependents (if any) or single adults plus dependents (if any). Figure B.2A and Figure B.2B show the trends in the number of individuals in our sample, the number of tax units, and the number filed tax returns. The difference in these growth rates, explains why the thresholds differ even when just focusing on wage and salary income, particularly in recent years. For all definitions of income, we see a significant tapering off in the growth of the top-earning thresholds during the last decade.

Figure B.1: Top earning thresholds with alternative data sources

(A) 99.9th percentile

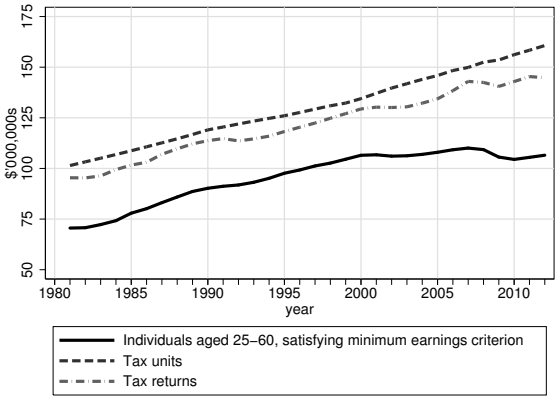


(B) 99th percentile

The following figures reconcile our findings with those in Saez (2012) that income shares for the top 1 percent and 0.1 percent have continued to trend upwards during the last decade. Figure B.3A and Figure B.3B show that below the 99.99th percentile, average income growth in the top percentiles, with or without capital gains, has remained roughly constant since 2000. Figure B.3C shows that average income for the top 0.01 percent has continued to rise during this period. Figure B.3D shows that average income for the bottom 99 percent has declined substantially more in these data than for our sample of wage and salary earners. The difference in the recent trends in top earning shares are thus due to (i) increases in capital income above the 99.99th percentile; and (ii) a larger decline in income for the bottom 99 percent that is due to the difference in the unit of observation: individuals versus tax units.

Figure B.2: Sample sizes with different units of analysis

(A) Sample sizes



(B) Normalized to 1 in 1981

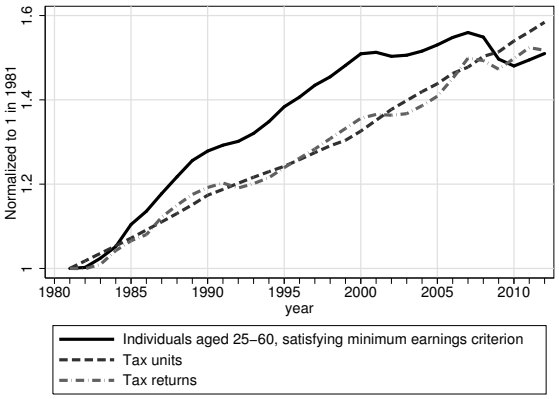
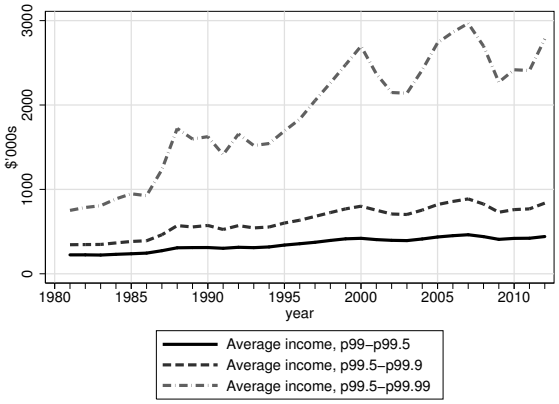
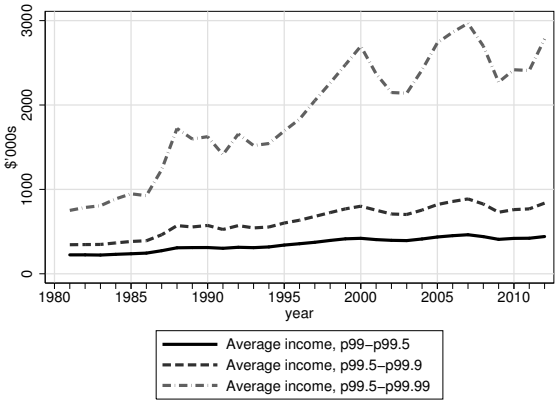


Figure B.3: Average income in top percentiles

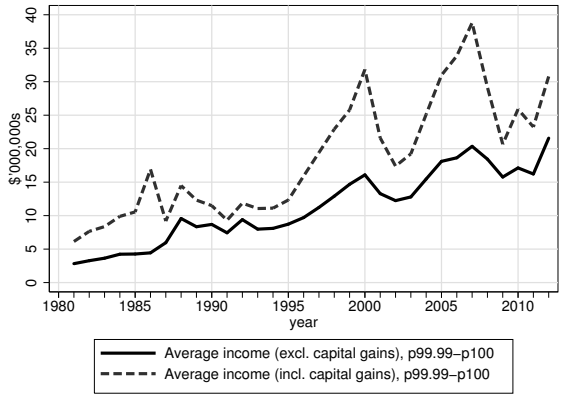
(A) Average income, excluding capital gains



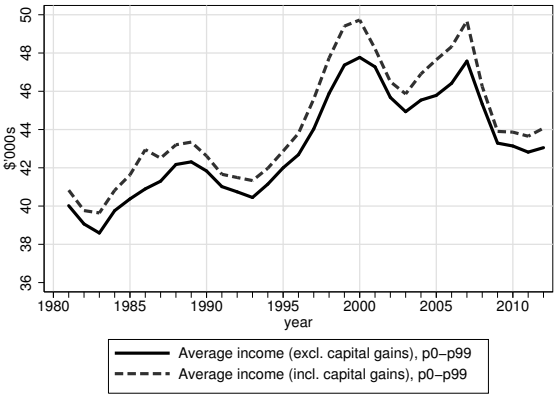
(B) Average income, including capital gains



(C) Average income of top 0.01 percent



(D) Average income of bottom 99 percent



C Lifetime earnings analysis for 30-59 year age range

This appendix reports analogous tables and figures to those in Section 6, but where the 30 year age range is taken to be the ages 30 to 59, rather than 25 to 54.

Table C.1: Lifetime earnings top earnings statistics

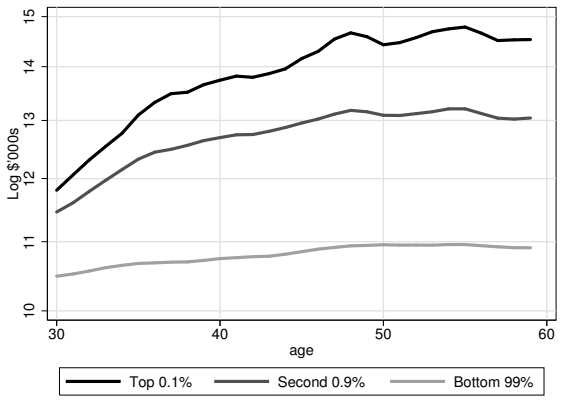
	Top 0.1%	Second 0.9%	Bottom 99%
30-year earnings thresholds:			
- 99.9th percentile ($'000s)	20,704		
- 99th percentile ($'000s)		7,043	
Mean 30-year earnings ($'000s)	38,092	10,545	1,276
Median 30-year earnings ($'000s)	29,467	9,443	1,043
Mean no. working years	27.9	28.3	25.6
Mean fraction of working years in age-specific:			
- top 0.1 percent	35%	5%	0%
- next 0.9pct	40%	42%	0%
- bottom 99 percent	25%	53%	100%

Table C.2: Gender differences among lifetime top earners

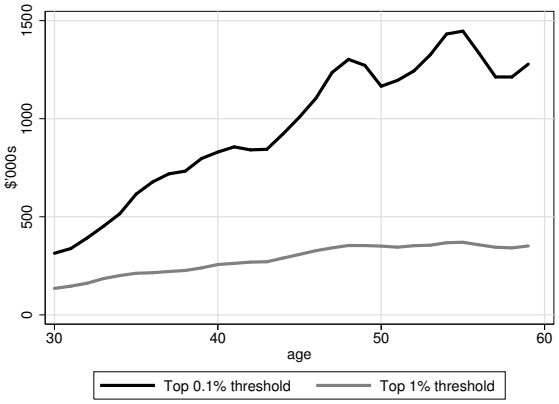
	Top 0.1%	Second 0.9%	Bottom 99%
Panel A: Overall top earners			
Female worker share	9%	11%	49%
Female earnings share	9%	10%	38%
Log mean gender gap	–0.01	0.06	0.46
Log p50 gender gap	–0.05	0.05	0.48
No. working years gender gap	0.40	0.20	0.90
Panel B: Gender-specific top earners			
Male threshold ($'000)	27,512	9,320	
Female threshold ($'000)	9,487	3,828	
Log mean gender gap	1.18	0.97	0.52
Log p50 gender gap	1.16	0.96	0.49
No. working years gender gap	–0.19	–0.01	0.94

Figure C.1: Age profiles by 30-year top earning groups

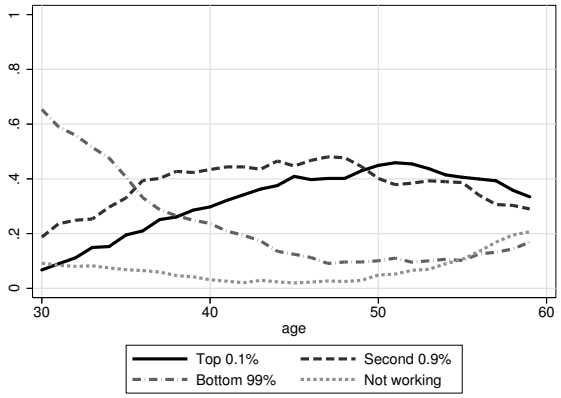
(A) Mean earnings by age



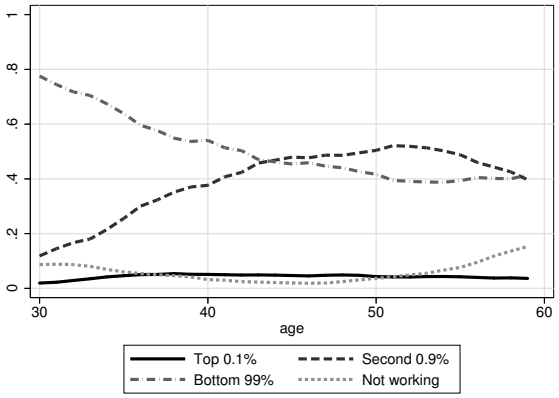
(B) Age-specific top-earning thresholds



(C) Location of lifetime top 0.1 percent in age-specific distributions



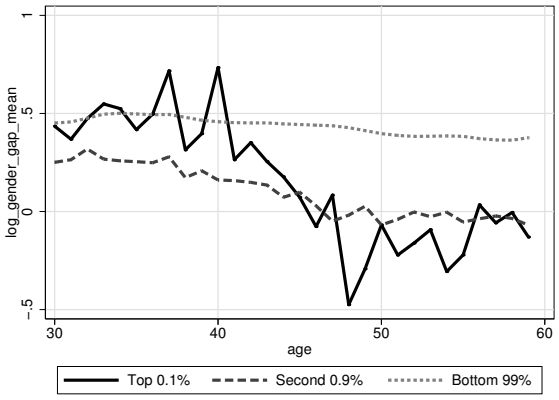
(D) Location of lifetime top 1 percent in age-specific distributions



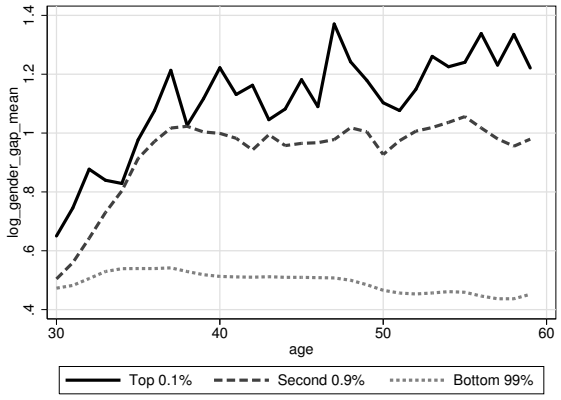
Notes: Figures refer to individuals from the 1951, 1952, and 1953 birth cohorts. Age-specific top-earning thresholds and groups are computed using only these three cohorts.

Figure C.2: Gender gap among 30-year top earners by age

(A) Overall lifetime top earners



(B) Gender-specific lifetime top earners



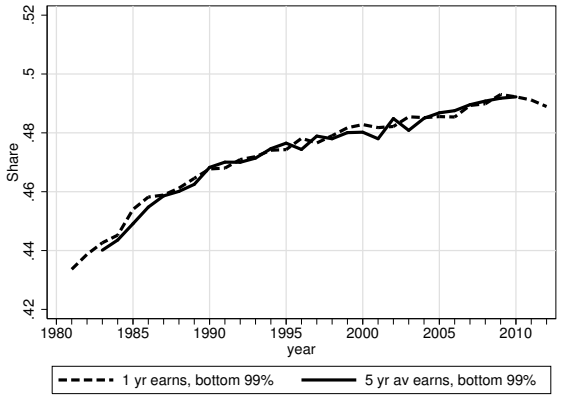
Notes: Figures refer to individuals from the 1951, 1952, and 1953 birth cohorts. Age-specific top-earning thresholds and groups are computed using only these three cohorts. Figures show mean gender gap in each part of the earnings distribution.

D Trends in the gender composition of the bottom 99 percent

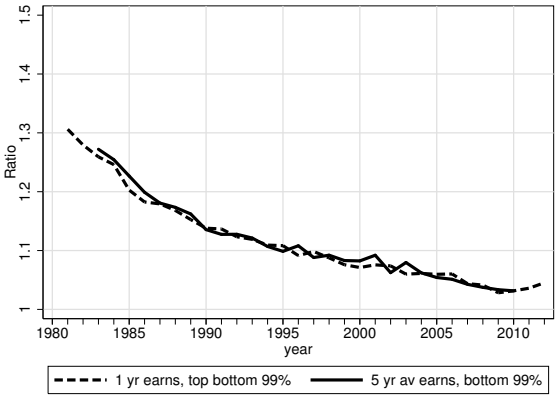
Figure D.1 plots the time trend for the female population share and the male-female population ratio, for the bottom 99 percent of the earnings distribution.

Figure D.1: Gender composition of overall top earners, bottom 99%

(A) Female population share



(B) Male-female population ratio

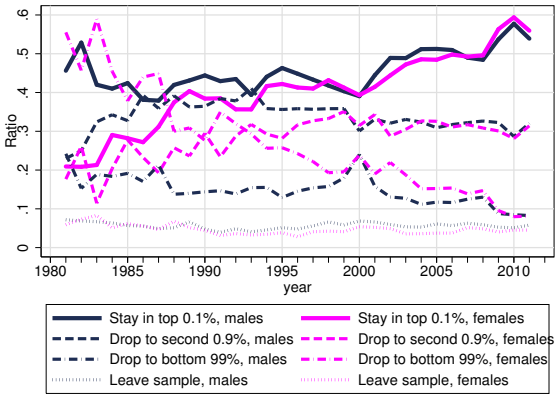


E Mobility within gender-specific distributions

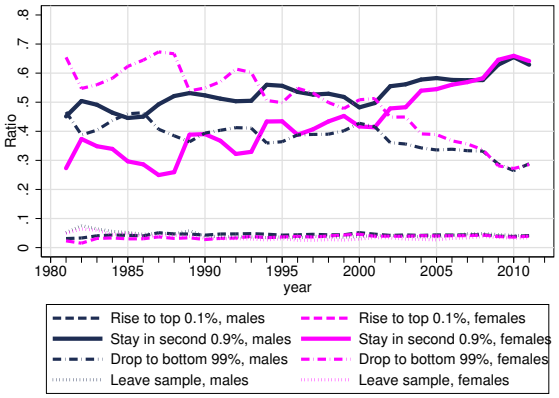
This appendix reports figures that are analogous to those in Section 5, but in which individuals are defined as top earners based on their position in their gender-specific earnings distribution, rather than the overall earnings distribution.

Figure E.1: Transition probabilities in and out of top percentiles of earnings distribution, by gender

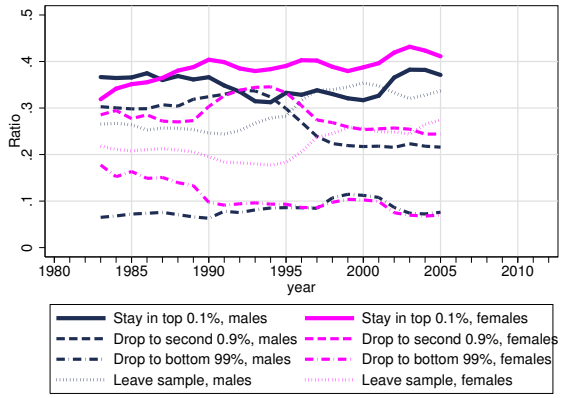
(A) One-year transition probabilities for annual earnings, top 0.1 percent



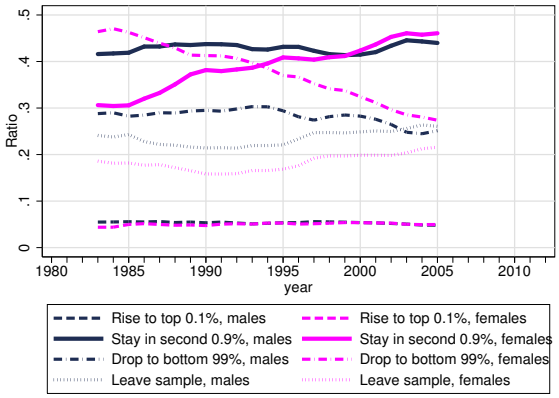
(B) One-year transition probabilities for annual earnings, second 0.9 percent



(C) Five-year transition probabilities for five–year earnings, top 0.1 percent



(D) Five-year transition probabilities for five-year earnings, second 0.9 percent



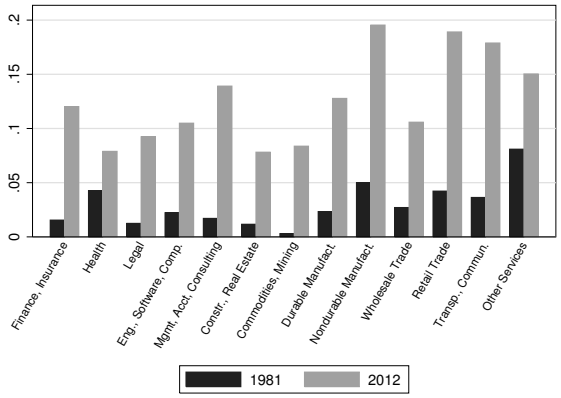
Notes: These figures show the probability that a top earner based on average earnings over the period $t-2, ..., t+2$ is a top earner based on average earnings over the period $t+3, ..., t+7$, separately for male top earners (blue) and female top earners (pink). Individuals are classified as top earners based on gender-specific earnings distributions.

F Industry analysis further figures

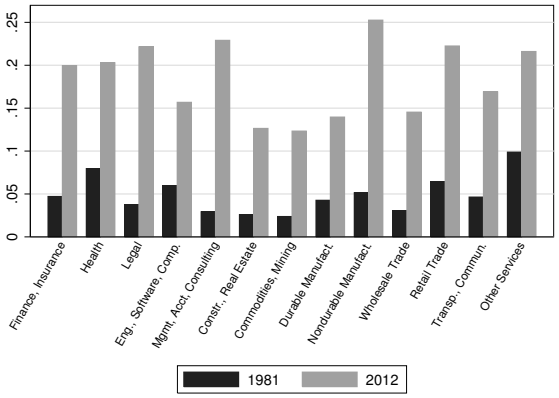
This appendix contains figures that are analogous to those in Secion 7, but which are constructed using annual earnings rather than five-year average earnings.

Figure F.1: Top earners by industry and gender, annual earnings

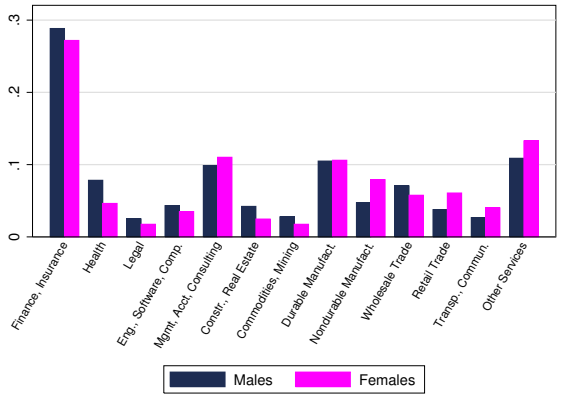
(A) Share of females by industry within top 0.1 percent



(B) Share of females by industry within top 0.9 percent



(C) Industry shares by gender within top 0.1 percent, 2008–12



(D) Industry shares by gender within second 0.9 percent, 2008–12

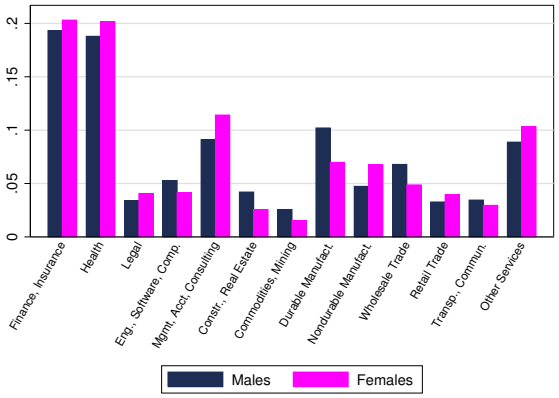


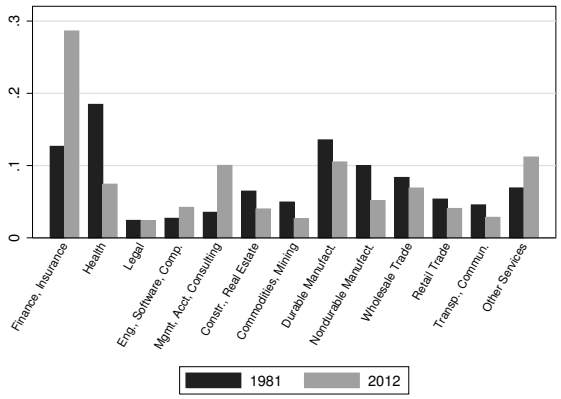
Table F.1: Selected US Companies and Associated (Primary) SIC Codes

Company Name:	Primary SIC Code	Descriptions
Google	7370	Computer Programming, Data Processing, And Computer Services
Apple,Dell	3571	Electronic computers
HP	3570	Computer and office equipment
Microsoft	7372	Prepackaged software
IBM	7371	Computer programing services
Intel	3674	Semiconductors and related services
Oracle	7372	Prepackaged software
Cisco	5045	Wholesale-Computers and Peripheral equipment and Software
Qualcomm	3663	Radio and TV broadcasting and communication equipment
Boeing	3721	Aircraft and parts
Amazon.com	5961	Retail-Catalog and Mail Order Houses
3M	3291	Abrasive products
Walmart	5331	Retail-Variety stores
Exon, Chevron, BP	2911	Petroleum refining
Total SA	1211	Crude petroleum and natural gas
Ford, GM, Tesla	3711	Motor vehicles and passenger car bodies
Berkshire-Hathaway, State Farm	6331	Fire, Marine and Casualty Insurance
General Electric:	3600	Electronic and other electrical equipment except computers
Cargill Inc	5153	Grain and field beans; Domestic Transportation of Freight
Bank of America, JP Morgan	6021	Banks
Goldman Sachs	6022	Investment bank
Morgan Stanley	6199	Investment bank
Mettle	6311	Life insurance

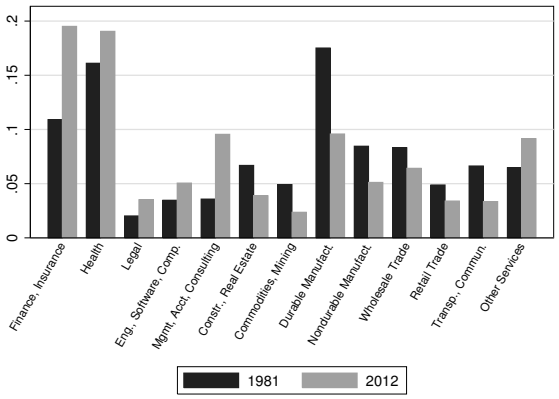
Notes: Some companies listed here have further SIC codes associated with them. For example, Microsoft: 7371, 7372, 7379 (Prepackaged software, primary), and 3944 (electronic games) and 3861 (photographic equipment). And similarly, Cargill Inc: 5153 (Grain & Field Beans); 4424 (Deep Sea Domestic Transportation of Freight); 6221 (Commodity Contracts Brokers & Dealers); 2041 (Flour & Other Grain Mill Products.)

Figure F.2: Industry composition of top earners, annual earnings

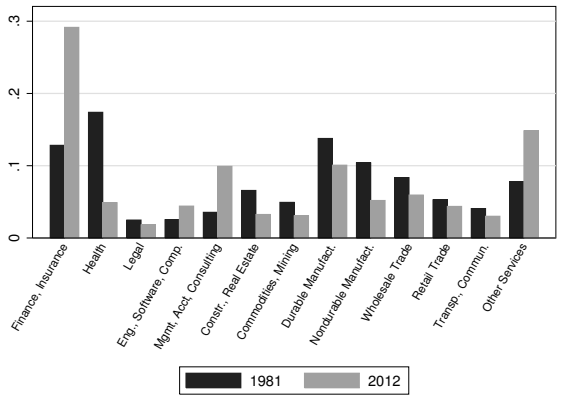
(A) Population shares, top 0.1 percent



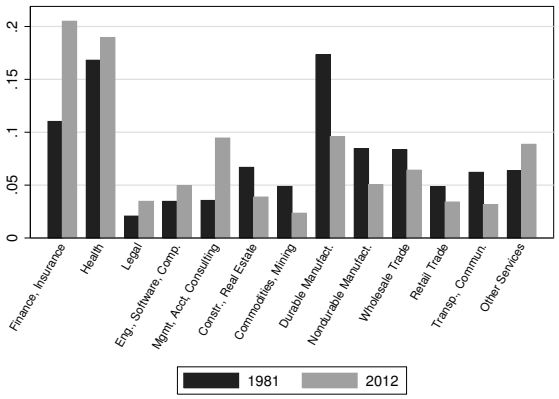
(B) Population shares, second 0.9 percent



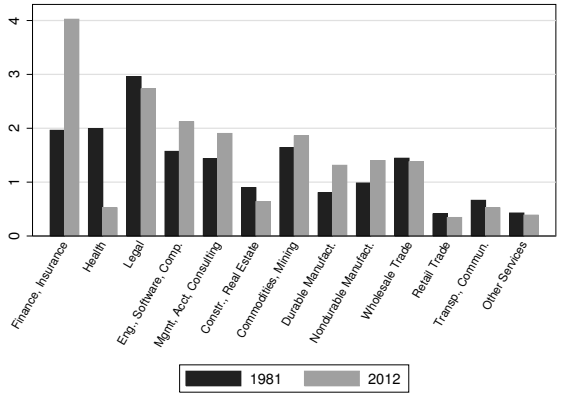
(C) Earnings shares, top 0.1 percent



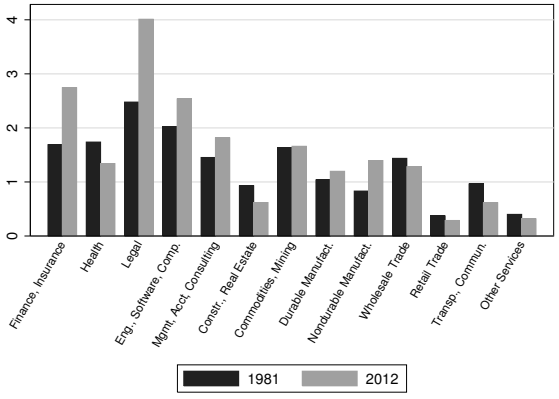
(D) Earnings shares, second 0.9 percent



(E) Population shares, top 0.1 percent relative to bottom 99 percent



(F) Population shares, second 0.9 percent relative to bottom 99 percent

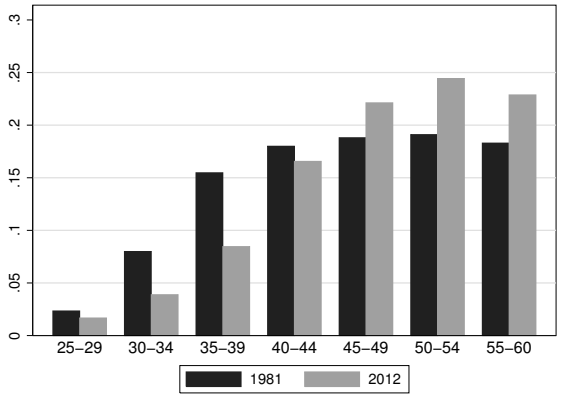


G Age analysis further figures

This appendix contains figures that are analogous to those in Section 8, but which are constructed using annual earnings rather than five-year average earnings, and additional figures that are references in Section 8.

Figure G.1: Age distribution of workers, annual earnings

(A) Age distribution of individuals in top 0.1 per-cent



(B) Age distribution of individuals in second 0.9 percent

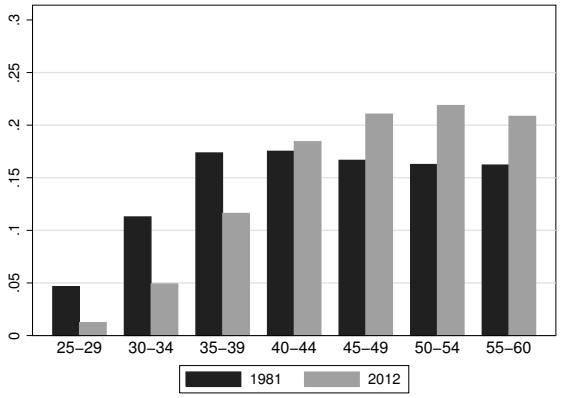
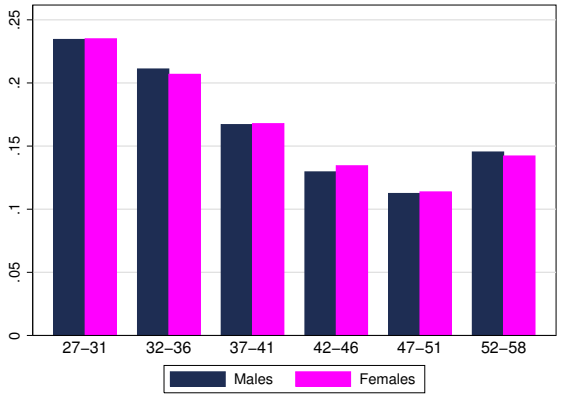


Figure G.2: Age distribution of workers by gender, overall distribution, five-year average earnings

(A) 1981-85



(B) 2008-12

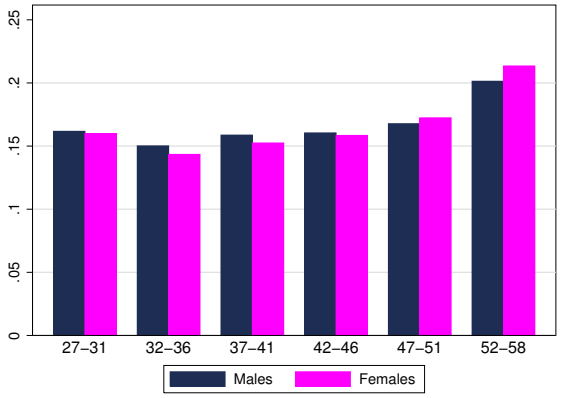
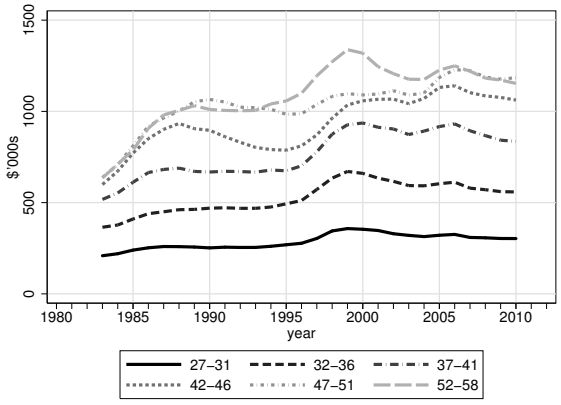
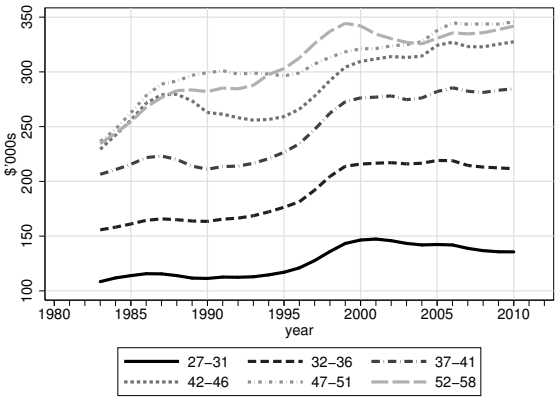


Figure G.3: Top-earning thresholds within age groups, five-year average earnings

(A) Thresholds for top 0.1 percent, by age group



(B) Thresholds for top 1 percent, by age group

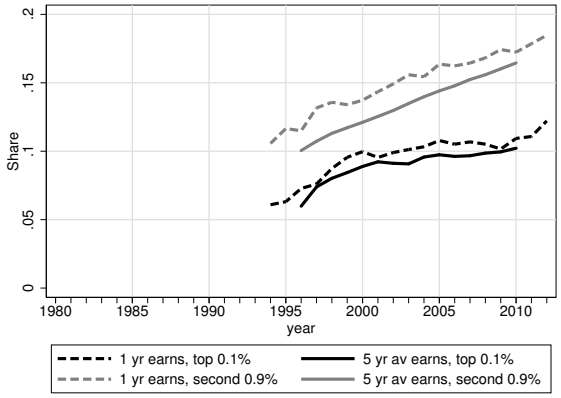


H Including self-employment income

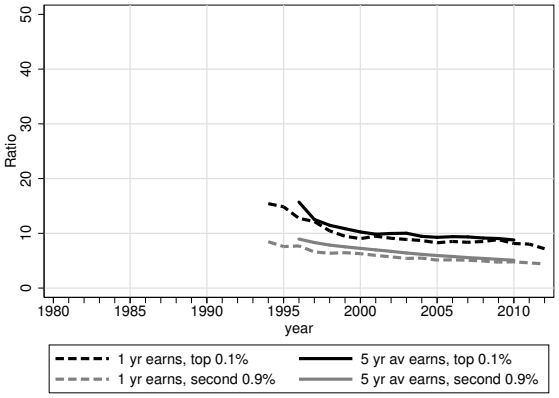
This appendix contains deleted figures from the main text, constructed using a definition of income that includes both wage and salary earnings, and earnings from self-employment income.

Figure H.1: Gender composition of top earners

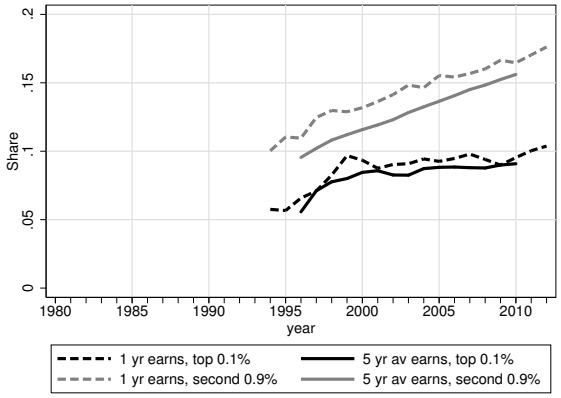
(A) Share of females among top earners



(B) Ratio of males to females among to earners



(C) Share of top earnings accruing to females



(D) Share of females among top earners, relative to share of females among all workers

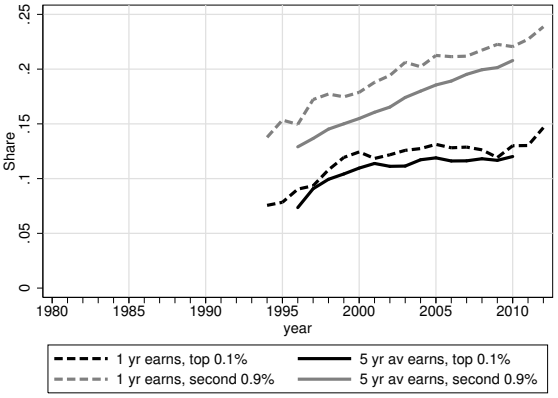
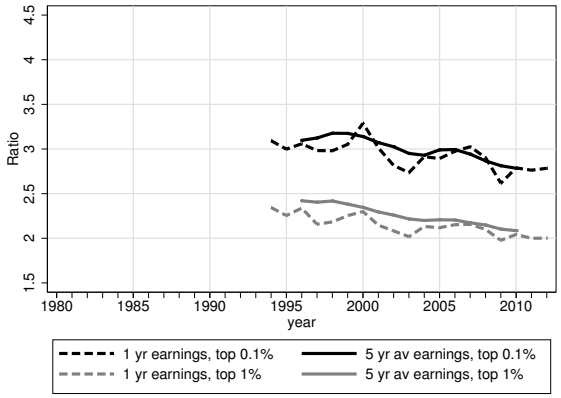
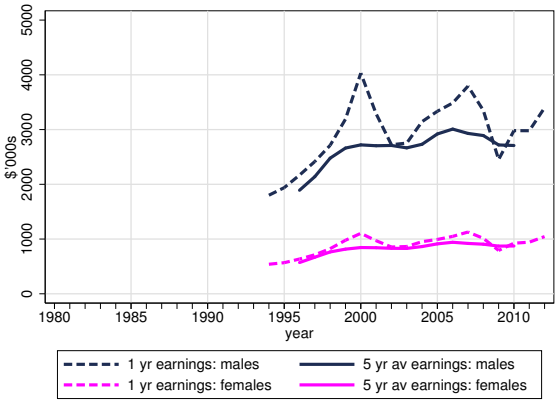


Figure H.2: Male top earners versus female top earners

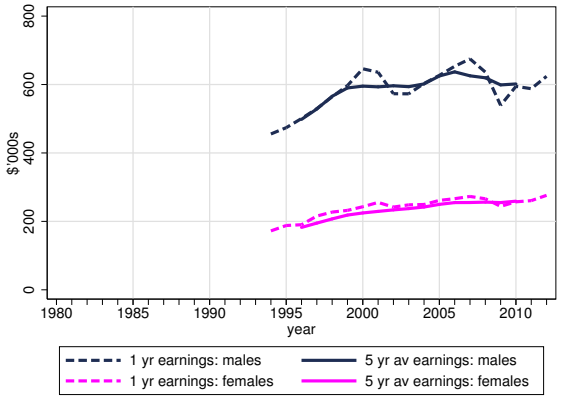
(A) Ratio of male to female top earning thresholds



(B) Average earnings among top 0.1 percent of males and top 0.1 percent of females



(C) Average earnings among second 0.9 percent of males and second 0.9 percent of females



(D) Share of top 0.1 percent earnings in top 1 percent earnings for males and females

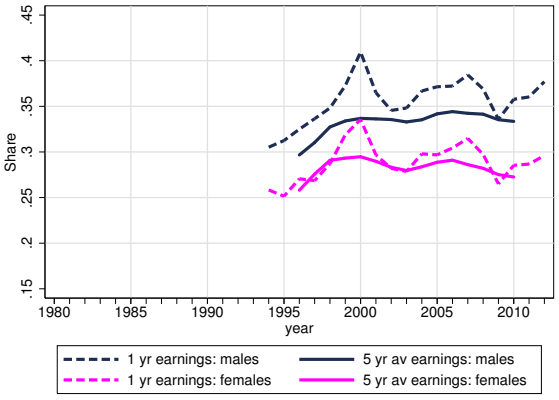
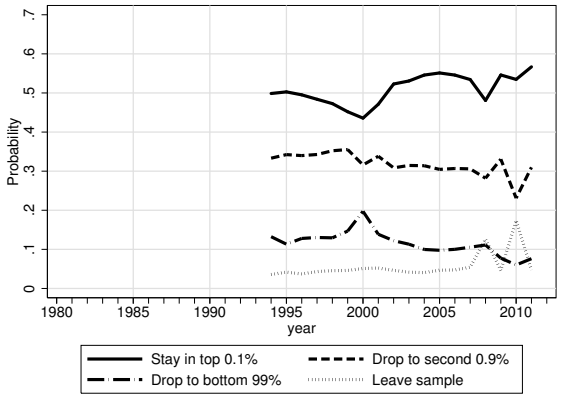
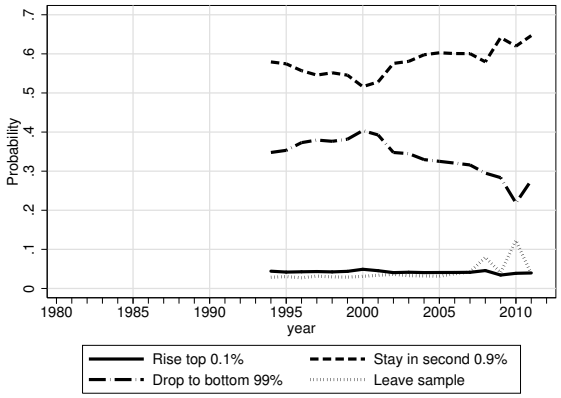


Figure H.3: Transition probabilities in and out of top percentiles of earnings distribution.

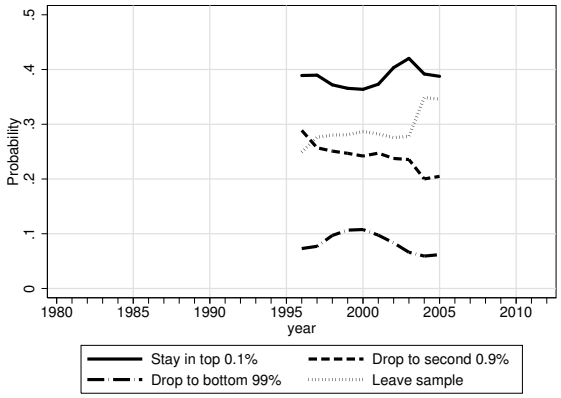
(A) 1-year transition prob. for annual earnings, top 0.1 percent



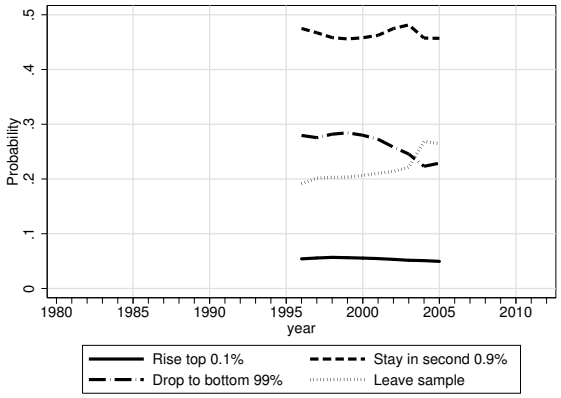
(B) 1-year transition prob. for annual earnings, second 0.9 percent



(C) 5-year transition prob. for 5-year earnings, top 0.1 percent



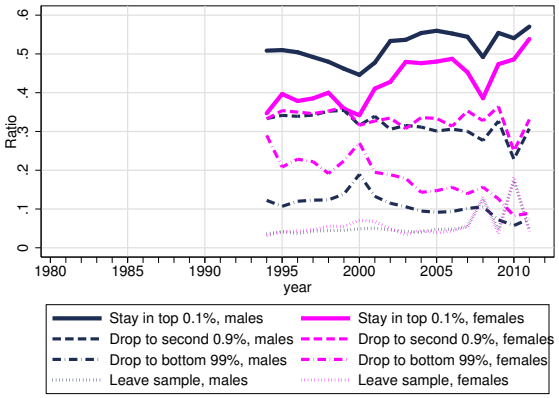
(D) 5-year transition prob. for 5-year earnings, second 0.9 percent



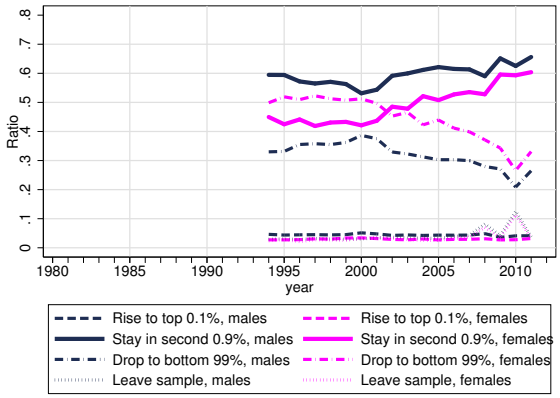
Notes: These figures show the probability that a top earner based on average earnings over the period $t-2, ..., t+2$ is a top earner based on average earnings over the period $t+3, ..., t+7$.

Figure H.4: Transition probabilities in and out of top percentiles of earnings distribution, by gender

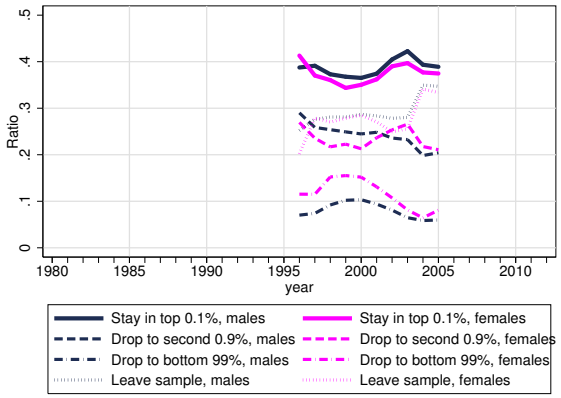
(A) 1 year transition probabilities for annual earnings, top 0.1 percent



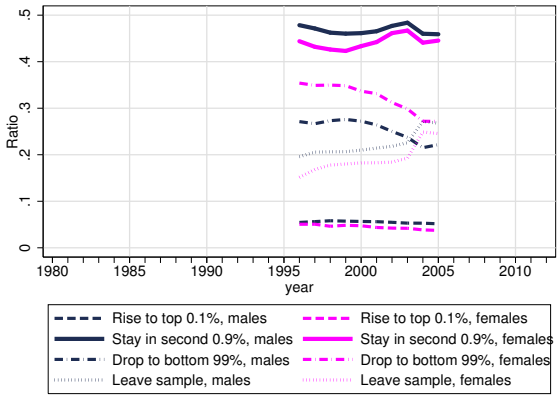
(B) 1 year transition probabilities for annual earnings, second 0.9 percent



(C) 5 year transition probabilities for 5-year earnings, top 0.1 percent



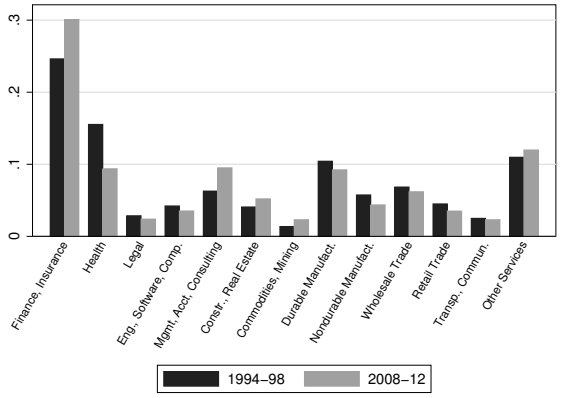
(D) 5 year transition probabilities for 5-year earnings, second 0.9 percent



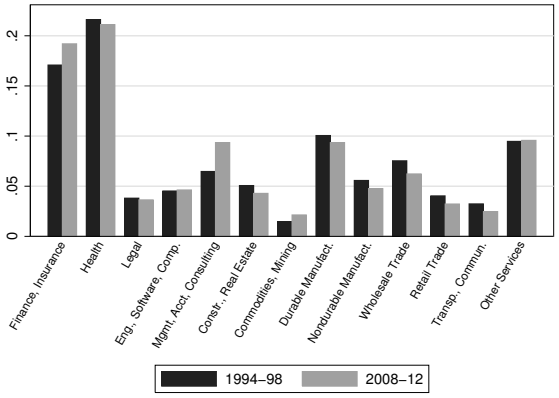
Notes: These figures show the probability that a top earner based on average earnings over the period $t-2, ..., t+2$ is a top earner based on average earnings over the period $t+3, ..., t+7$, separately for male top earners (blue) and female top earners (pink).

Figure H.5: Industry composition of top earners, 5-year average earnings

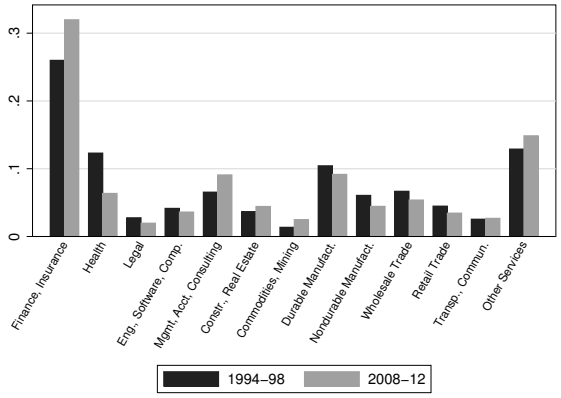
(A) Population shares, top 0.1 percent



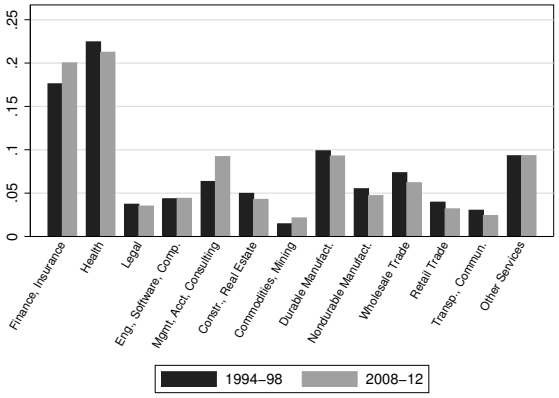
(B) Population shares, second 0.9 percent



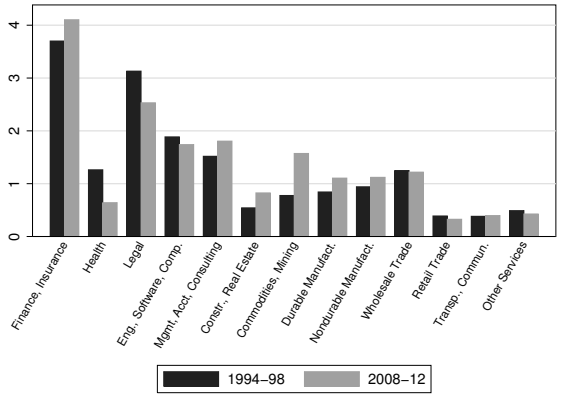
(C) Earnings shares, top 0.1 percent



(D) Earnings shares, second 0.9 percent



(E) Population shares, top 0.1 percent relative to bottom 99 percent



(F) Population shares, second 0.9 percent relative to bottom 99 percent

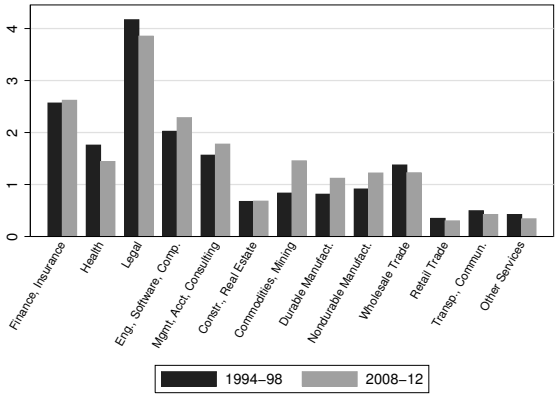
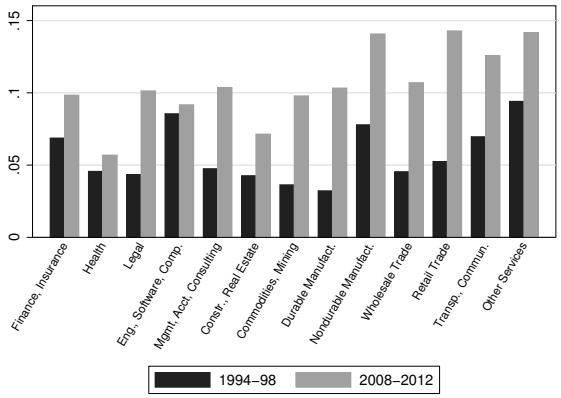
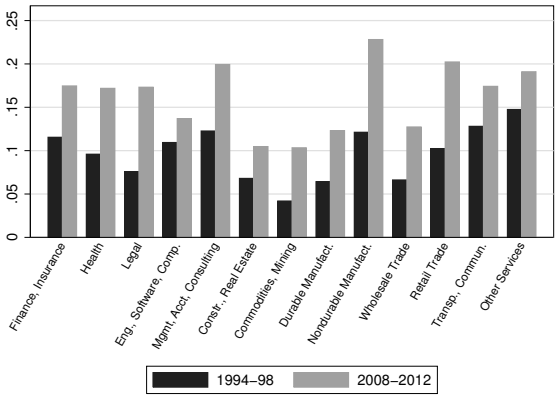


Figure H.6: Top earners by industry and gender, 5-year average earnings

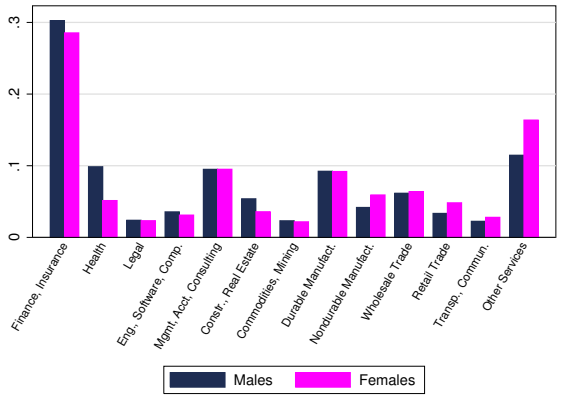
(A) Share of females by industry within top 0.1 percent



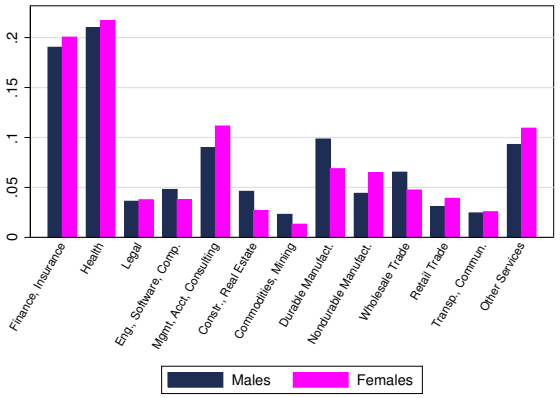
(B) Share of females by industry within top 0.9pct



(C) Industry shares by gender within top 0.1 percent, 2008–12



(D) Industry shares by gender within second 0.9 percent, 2008–12



References

Saez, E. (2012). Striking it richer: The evolution of top incomes in the United States. Working paper, University of California at Berkeley.