ONLINE APPENDIX

A Additional Data Source and Methods Tables and Figures

In Section 2 we discussed the construction of our eligible-workers frame. Here we provide further details on which workers are excluded from the all-workers frame to arrive at the eligible-workers frame and how this impacts the earnings coverage of LEHD when compared to NIPA.

A.1 All-Workers Frame

The all-workers frame contains earnings for all jobs reported on the UI data for each date regime in the relevant years from 1990-2013, as noted in Figure 1 and summarized in Table A.1 below.

Using the person level earnings, e_{it}, an estimate of annual earnings for the all-workers frame in year t is calculated as follows:

$$E_t^{AW} = \sum_{i \in AW_t} e_{it},$$

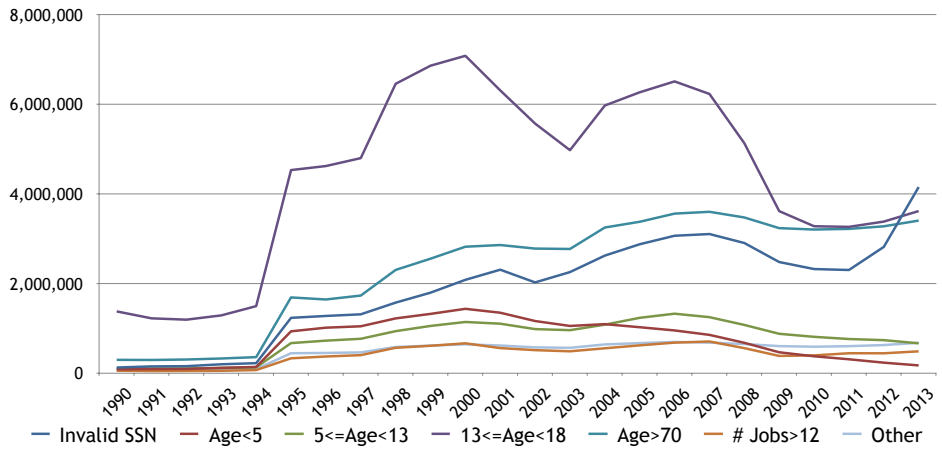
where AW_t is the set of workers in the all-workers frame in year t.

Table A.1: LEHD Regimes

Count	State	First YYYYQ	Last YYYYQ	Entry Order	Pct 2012Q1 QCEW Emp
	Regime 1 - 1990Q1 to 2013Q4 - 19.35% of 2012Q1 QCEW Employment				
1	Maryland	1985Q2	2014Q3	1	1.83%
2	Alaska	1990Q1	2014Q3	2	0.22%
3	Colorado	1990Q1	2014Q3	3	1.70%
4	Idaho	1990Q1	2014Q3	4	0.45%
5	Illinois	1990Q1	2014Q3	5	4.38%
6	Indiana	1990Q1	2014Q3	6	2.19%
7	Kansas	1990Q1	2013Q4	7	0.98%
8	Louisiana	1990Q1	2014Q2	8	1.41%
9	Missouri	1990Q1	2014Q3	9	1.99%
10	Washington	1990Q1	2014Q3	10	2.12%
11	Wisconsin	1990Q1	2014Q3	11	2.08%
	Regime 2 - 1995Q1 to 2013Q4 - 48.28% of 2012Q1 QCEW Employment				
12	North Carolina	1991Q1	2014Q3	1	2.92%
13	Oregon	1991Q1	2014Q3	2	1.23%
14	Pennsylvania	1991Q1	2014Q3	3	4.44%
15	California	1991Q3	2014Q3	4	11.37%
16	Arizona	1992Q1	2014Q3	5	1.85%
17	Wyoming	1992Q1	2014Q3	6	0.19%
18	Florida	1992Q4	2014Q2	7	5.78%
19	Montana	1993Q1	2014Q3	8	0.31%
20	Georgia	1994Q1	2014Q3	9	2.90%
21	South Dakota	1994Q1	2014Q2	10	0.30%
22	Minnesota	1994Q3	2014Q3	11	2.05%
23	New York	1995Q1	2014Q3	12	6.49%
24	Rhode Island	1995Q1	2014Q3	13	0.35%
25	Texas	1995Q1	2014Q3	14	8.10%
	Regime 3 - 1998Q1 to 2013Q4 - 17.66% of 2012Q1 QCEW Employment				
26	New Mexico	1995Q3	2014Q3	1	0.55%
27	Hawaii	1995Q4	2014Q3	2	0.44%
28	Connecticut	1996Q1	2014Q3	3	1.26%
29	Maine	1996Q1	2014Q3	4	0.43%
30	New Jersey	1996Q1	2014Q3	5	2.87%
31	Kentucky	1996Q4	2014Q3	6	1.32%
32	West Virginia	1997Q1	2014Q3	7	0.52%
33	Michigan	1998Q1	2014Q3	8	3.04%
34	Nevada	1998Q1	2014Q3	9	0.89%
35	North Dakota	1998Q1	2014Q3	10	0.31%
36	South Carolina	1998Q1	2014Q3	11	1.35%
37	Tennessee	1998Q1	2014Q3	12	2.03%
38	Virginia	1998Q1	2014Q2	13	2.65%
	Regime 4 - 2004Q1 to 2013Q4 - 14.71% of 2012Q1 QCEW Employment				
39	Delaware	1998Q3	2014Q3	1	0.31%
40	Iowa	1998Q4	2014Q3	2	1.12%
41	Nebraska	1999Q1	2014Q3	3	0.69%
42	Utah	1999Q1	2014Q3	4	0.91%
43	Ohio	2000Q1	2014Q3	5	3.93%
44	Oklahoma	2000Q1	2014Q3	6	1.11%
45	Vermont	2000Q1	2014Q3	7	0.22%
46	Alabama	2001Q1	2014Q3	8	1.34%
47	Massachusetts	2002Q1	2014Q2	9	2.55%
48	District of Columbia	2002Q2	2014Q3	10	0.43%
49	Arkansas	2002Q3	2014Q3	11	0.86%
50	New Hampshire	2003Q1	2014Q2	12	0.47%
51	Mississippi	2003Q3	2014Q3	13	0.77%

Notes: This table presents information on the states that make up each date regime. Each panel gives the first and last quarter of data available, the entry order, and the employment coverage (percent of 2012Q1 private QCEW employment) of each state in the regime. OPM data for federal workers is not shown in this table, but is available beginning in 2000Q1.

Figure A.1: Immigrant Candidates – Excluded Earnings Records



Notes: This figure presents the count of earnings records excluded from the eligible-workers frame each year, disaggregated by the different eligibility requirements the record failed to meet: (i) *Invalid SSN* are records that are only on the UI; (ii) *Age<5* are records where the SNN is valid, but the age of the worker is less than 5; (iii) *5≤Age<13* are records where the worker is between 5 and 13 years old; (iv) *13≤Age<18* are records where the worker is between 13 and 18 years old; (v) *Age>70* are records where the worker is more than 70 years old; (vi) *#Jobs>12* are records where the worker has more than 12 jobs a year; and (vii) *Other* are records that fail to meet the other eligibility requirements, such as the year is greater than or equal to the SSN year of issue and less than year of death (when available).

Table A.2: Immigrant Candidates – Excluded Earnings Records

Year	Total	Invalid SSN	Age<5	5≥Age<13	13≤Age<18	Age>70	#Jobs>12	Other
1990	2,173,054	131,768	92,173	115,966	1,383,852	302,791	61,336	85,168
1991	2,029,041	156,980	96,503	110,535	1,228,937	300,232	53,311	82,543
1992	2,024,225	161,800	99,380	111,528	1,199,329	310,526	55,873	85,789
1993	2,227,908	204,299	123,925	122,587	1,294,809	333,303	59,024	89,961
1994	2,546,460	228,963	145,038	136,015	1,500,927	363,506	74,634	97,377
1995	9,875,811	1,240,177	939,315	676,532	4,536,074	1,695,371	337,545	450,797
1996	10,144,571	1,282,244	1,020,460	731,340	4,625,974	1,649,645	377,807	457,101
1997	10,560,373	1,318,787	1,051,685	773,013	4,802,606	1,737,019	408,080	469,183
1998	13,680,138	1,579,419	1,227,565	942,868	6,460,058	2,308,455	571,745	590,028
1999	14,850,424	1,801,636	1,328,052	1,059,582	6,864,218	2,559,284	617,195	620,457
2000	15,909,402	2,087,866	1,441,233	1,147,779	7,084,996	2,826,633	671,695	649,200
2001	15,142,444	2,313,768	1,354,067	1,109,587	6,313,180	2,864,144	565,342	622,356
2002	13,646,946	2,030,273	1,168,828	988,866	5,573,020	2,784,977	519,677	581,305
2003	13,105,529	2,260,426	1,059,202	965,151	4,979,593	2,776,405	493,455	571,297
2004	15,254,789	2,628,435	1,099,414	1,087,743	5,976,072	3,254,876	561,150	647,099
2005	16,109,360	2,881,580	1,030,810	1,240,576	6,271,025	3,383,095	626,426	675,848
2006	16,830,576	3,071,079	959,130	1,332,606	6,513,877	3,564,841	686,925	702,118
2007	16,464,027	3,109,359	860,258	1,254,957	6,233,964	3,605,470	712,999	687,020
2008	14,509,746	2,909,378	683,388	1,081,938	5,135,680	3,478,821	564,086	656,455
2009	11,701,711	2,484,829	471,798	884,181	3,620,311	3,240,941	390,427	609,224
2010	11,019,697	2,328,456	382,395	816,592	3,283,378	3,210,027	402,839	596,010
2011	10,942,606	2,307,310	315,743	767,636	3,269,325	3,224,106	450,244	608,242
2012	11,556,277	2,822,199	240,123	742,658	3,386,957	3,282,004	449,498	632,838
2013	13,216,695	4,157,518	178,979	671,775	3,622,084	3,409,276	492,710	684,353

Notes: The first column presents of the total number of earnings records excluded from the eligible-workers frame each year. The remaining columns disaggregate this count by the different eligibility requirements the record failed to meet: (i) *Invalid SSN* are records that are only on the UI; (ii) *Age<5* are records where the SNN is valid, but the age of the worker is less than 5; (iii) *5≤Age<13* are records where the worker is between 5 and 13 years old; (iv) *13≤Age<18* are records where the worker is between 13 and 18 years old; (v) *Age>70* are records where the worker is more than 70 years old; (vi) *#Jobs>12* are records where the worker has more than 12 jobs a year; and (vii) *Other* are records that fail to meet the other eligibility requirements, such as the year is greater than or equal to the SSN year of issue and less than year of death (when available). The frame is complete from 2004 forward.

A.2 Comparison to NIPA

The BEA NIPA estimates are based primarily on the BLS Quarterly Census of Employment and Wages (QCEW), an alternative source of employment and earnings with similar coverage as the UI based job level data used in this paper. A firm typically files a QCEW firm-level report in conjunction with the UI job-level data each quarter. The QCEW report is sent to BLS where it undergoes edits and imputations before the final statistics are released.[53] These data are then used by BEA as the primary input when estimating the wage and salary component of the NIPA tables.[54]

Table A.3 presents a comparison of our estimates of annual earnings with the BEA NIPA data. Figure A.2 plots this comparison. Our estimates of total annual earnings using the all-workers frame vary from 16.5% of NIPA wage and salary estimates in 1990, the beginning of LEHD date regime 1; to 60.1% in 1995, the beginning of date regime 2; to 76.4% in 1998, the beginning of date regime 3; to 90.6% in 2004, the beginning of date regime 4. Once LEHD data are complete in 2004, the two series track almost exactly. By 2013 the all-workers estimate is about 91.7% of

[53] See http://www.bls.gov/opub/hom/pdf/homch5.pdf for more information

[54] The BLS QCEW estimates account for about 95% of the BEA wage and salary component of the NIPA tables. See http://www.bea.gov/faq/index.cfm?faq_id=104 for more information.

the NIPA wage and salary estimates. The eligible-workers estimates follow a similar pattern as the all-workers estimates, with about two percentage points lower coverage relative to the all-workers frame after 2004.

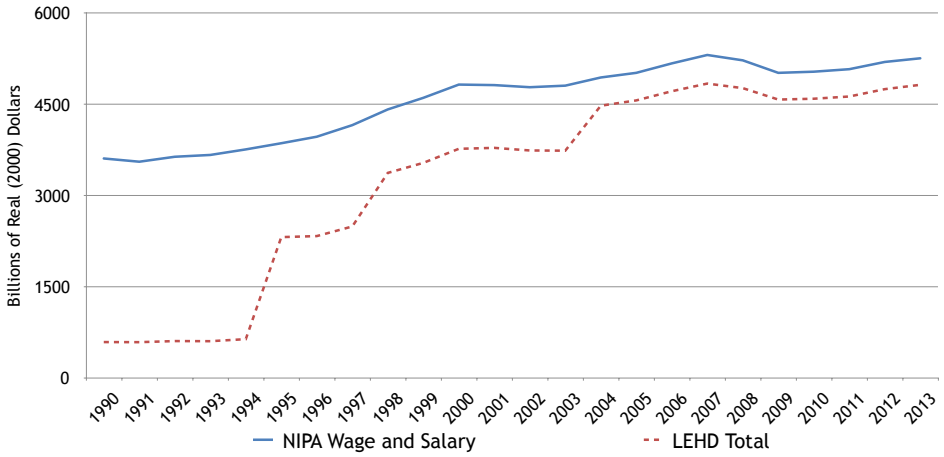
The coverage of both the all-workers and eligible-workers frames is very low relative to the NIPA estimates in the early 1990s but increases dramatically in 1995 once the historical data for the more populous states (CA, FL, NY, and TX) have entered the LEHD infrastructure files. When the frame is complete (date regime 4), there is an apparent coverage gap of about eight to nine percentage points for the all-workers frame and ten to 11 percentage points for the eligible-workers frame. About half of this gap is due to differences between the statutory-employer population for UI wage records and the NIPA definition of wage and salary income. When comparing frames with similar coverage definitions (UI wage records vs. QCEW), our results suggest that the gap between the two frames is about four to five percentage points for the all-workers frame and five to six percentage points for the eligible-workers frame.

Table A.3: Earnings Measures–National Income and Product Accounts versus LEHD Data

Year	NIPA Wage and Salary	LEHD Total	Eligible Workers	Immigrant Candidates
1990	3,611.6	594.7	587.4	7.3
1991	3,558.4	593.2	585.7	7.5
1992	3,639.8	611.2	603.8	7.4
1993	3,669.6	609.3	601.6	7.7
1994	3,760.7	642.6	633.9	8.7
1995	3,862.1	2,319.3	2,279.7	39.6
1996	3,969.2	2,336.3	2,294.5	41.8
1997	4,159.4	2,494.2	2,448.4	45.8
1998	4,417.6	3,374.6	3,312.9	61.7
1999	4,607.8	3,539.3	3,469.8	69.5
2000	4,825.9	3,770.5	3,694.7	75.8
2001	4,817.3	3,785.9	3,707.7	78.2
2002	4,782.5	3,743.2	3,666.4	76.8
2003	4,808.3	3,739.8	3,663.8	76.0
2004	4,942.6	4,478.7	4,387.3	91.4
2005	5,018.8	4,565.8	4,469.3	96.5
2006	5,174.0	4,716.5	4,613.0	103.5
2007	5,312.4	4,842.3	4,736.2	106.1
2008	5,224.3	4,767.6	4,667.3	100.3
2009	5,018.6	4,579.8	4,489.4	90.4
2010	5,037.6	4,593.4	4,503.7	89.7
2011	5,078.9	4,630.1	4,539.8	90.3
2012	5,197.7	4,750.8	4,652.5	98.3
2013	5,257.9	4,822.0	4,706.0	116.0

Notes: This table compares total earnings as measured in the BEA NIPA to earnings computed from LEHD. *LEHD Total* presents total annual earnings for the all-workers frame. This total is decomposed into earnings attributed to workers included in the eligible-workers frame (*Eligible Workers*) and to workers who are not included (*Immigrant Candidates*). Units are in billions of real (2000) dollars, converted using CPI-U. The frame is complete from 2004 forward.

Figure A.2: NIPA



Notes: This figure compares total earnings as measured in BEA NIPA (blue line) to earnings computed from LEHD using all workers (red line).

A.3 Estimation of the Earnings/Inactivity Distribution

Table A.4: Labor Force Activity of Workers in Each Earnings Bin

Quarters Worked	**Longest Job**	**Workers** *Counts*	*Percent*	**Jobs** (*Avg*)	**Earnings** (*Avg*)
Bottom 20% of Earnings Distribution					
1	1	8,543,957	30.6%	1.066	$1,366
2	1	1,883,159	6.7%	1.996	$2,187
2	2	5,806,138	20.8%	1.213	$2,824
3	1	520,324	1.9%	2.594	$3,029
3	2	2,467,851	8.8%	2.297	$3,480
3	3	2,591,936	9.3%	1.263	$3,726
4	1	58,758	0.2%	4.542	$3,480
4	2	949,367	3.4%	3.429	$4,274
4	3	932,150	3.3%	2.602	$4,544
4	4	187,115	0.7%	1.716	$4,161
4	5	1,078,088	3.9%	1.440	$4,178
4	6	2,893,038	10.4%	1.251	$4,227
Middle 60% of Earnings Distribution					
1	1	853,497	1.0%	1.051	$13,637
2	1	489,513	0.6%	1.643	$14,924
2	2	2,697,567	3.2%	1.176	$14,375
3	1	680,994	0.8%	1.475	$19,879
3	2	2,409,536	2.9%	2.119	$15,891
3	3	4,976,450	5.9%	1.233	$17,446
4	1	52,620	0.1%	3.726	$17,579
4	2	2,746,891	3.3%	3.287	$17,604
4	3	7,105,740	8.5%	2.592	$20,563
4	4	841,481	1.0%	2.109	$19,230
4	5	8,869,511	10.6%	1.602	$22,405
4	6	52,012,001	62.1%	1.212	$26,107
Top 20% of Earnings Distribution					
1	1	75,101	0.3%	1.038	$146,574
2	1	34,381	0.1%	1.361	$138,531
2	2	112,925	0.4%	1.096	$102,246
3	1	94,047	0.3%	1.178	$92,110
3	2	171,999	0.6%	1.605	$95,079
3	3	434,213	1.6%	1.128	$89,432
4	1	7,589	0.0%	2.608	$90,693
4	2	312,325	1.1%	2.752	$84,965
4	3	1,383,555	5.0%	2.323	$87,727
4	4	139,347	0.5%	1.993	$90,280
4	5	2,493,150	8.9%	1.500	$92,054
4	6	22,653,328	81.2%	1.181	$88,447

Notes: Each row in the table represents a specific combination of quarters worked and number of quarters in the longest job. A five quarter longest job is active in **either** the fourth quarter of the previous year or the first quarter of the subsequent year, while a six quarter longest job is active in **both**. The number of quarters in the longest job takes on values from one to six. The counts are averages per year.

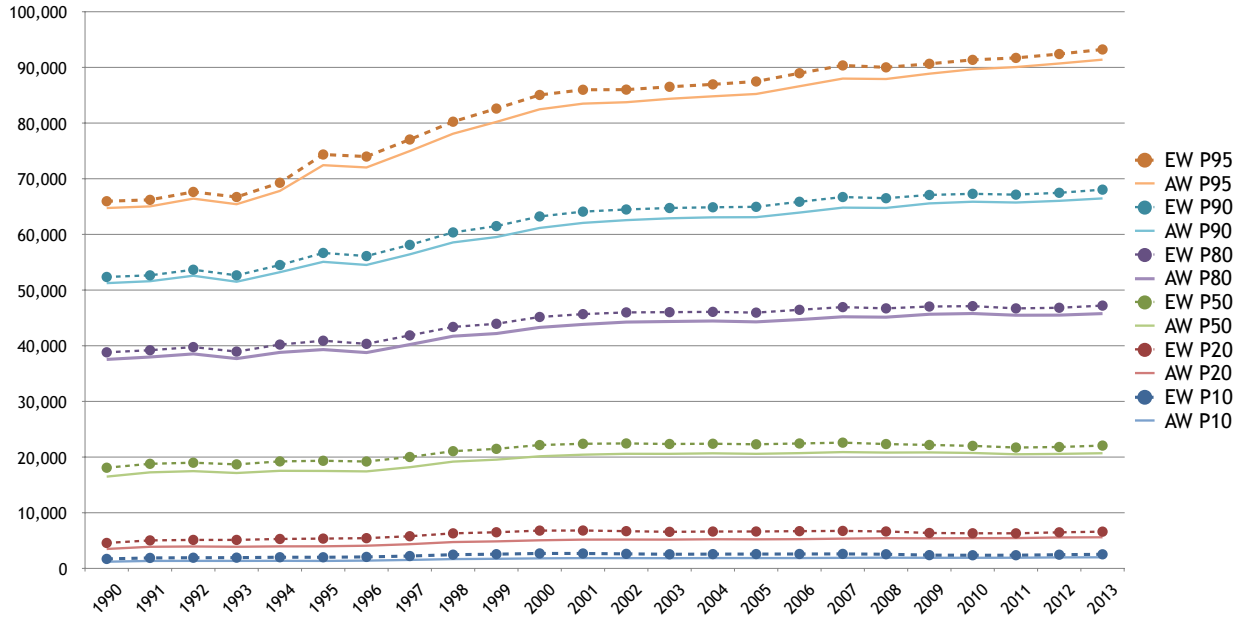
B Inequality Trends in the LEHD All-Workers Frame (1990-2013)

In Section 3 we discussed the trends in earnings inequality observed in the eligible-workers frame. Here, we detail the inequality trends in the all-workers frame, and analyze how they differ from the trends observed in the eligible-workers frame.

With a better understanding of how the exclusion of specific workers affects the distribution of earnings, we then turn our attention to earnings inequality. We analyze how various measurements of the gap between the top and bottom of the earnings distribution have changed over time and how the trends change as we move from the all-workers to the eligible-workers frame.

Figure B.1 plots selected percentiles for the two worker frames: the solid lines are the percentiles computed from the all-workers frame, while the dotted lines are the percentiles computed from the eligible-workers frame. Comparing the solid and dotted lines in Figure B.1, it is clear that the main consequence of shifting the frame from all workers to eligible workers is an increase in the percentile values, particularly at the bottom of the earnings distribution.

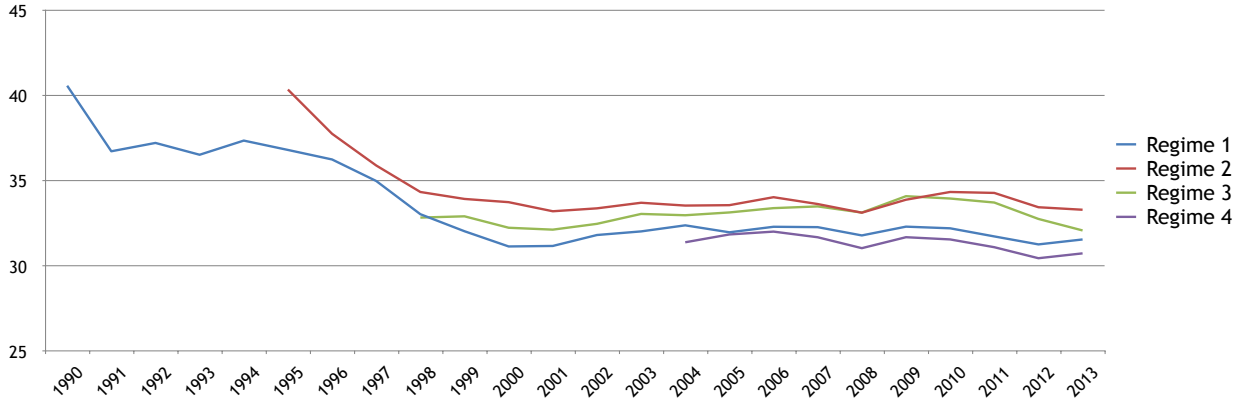
Figure B.1: Percentiles of the Earnings Distribution by Worker Frame



Notes: This figure plots the 10^{th}, 20^{th}, 50^{th}, 80^{th}, 90^{th}, and 95^{th} percentiles of the earnings distribution by worker frame and year. The solid lines are the percentiles of the earnings distribution of all workers (AW) by year. The dotted lines are the percentiles of the earnings distribution of eligible workers (EW) by year.

Figure B.2 plots the ratio of the 90^{th} percentile to the 10^{th} percentile for each date regime using the all-workers frame. The figure confirms that there are some differences in the levels of these curves but the trend analysis is largely unchanged.

Figure B.2: Ratio of the 90th to the 10th Percentile of the Earnings Distribution



Notes: This figure plots the ratio of the 90th to the 10th percentile for *all* workers by date regime.

To see this more clearly, Table B.1 presents the average percentile values from 1995-2013 for both the all-workers and the eligible-workers frames, and the last row computes their ratio (eligible workers to all workers). First, notice that the ratio is always above one, meaning that each percentile computed from the eligible-workers frame is greater than the equivalent percentile computed from the all-workers frame. Removing the immigrant candidates from the all-workers frame to construct the eligible-workers frame eliminates an unknown number of individuals who make very low earnings and, thus, tend to be at the bottom of the all-workers earnings distribution. For example, in 2006, immigrant candidates held about 8% of all jobs, but only contributed about 2% to total earnings. Furthermore, average earnings for immigrant candidates were about $6,150 in 2006 as compared with $32,865 for eligible workers. Thus, the removal of these low-earnings workers from the all-workers frame makes the ratio of EW to AW percentiles in Table B.1 higher towards the bottom of the earnings distribution. Specifically, notice that the 1^{st} percentile in the eligible-workers earnings distribution is, on average, about 32% greater than the 1^{st} percentile in the all-workers earnings distribution; the 5^{th} percentile is about 41% greater, the 10^{th} percentile is about 36% greater, and the 20^{th} percentile is about 26% greater. From the median onwards, while the absolute differences in the percentile values are large, the relative differences are not as stark, with the percentiles in the eligible-workers earnings distribution being about 2% to 8% greater than the corresponding percentile in the all-workers earnings distribution. Finally, notice that regardless of the worker frame used, there is a large number of workers with very low earnings in LEHD, with the average 10^{th} percentile at only $1,858 in the all-workers frame and $2,527 in the eligible-workers frame.

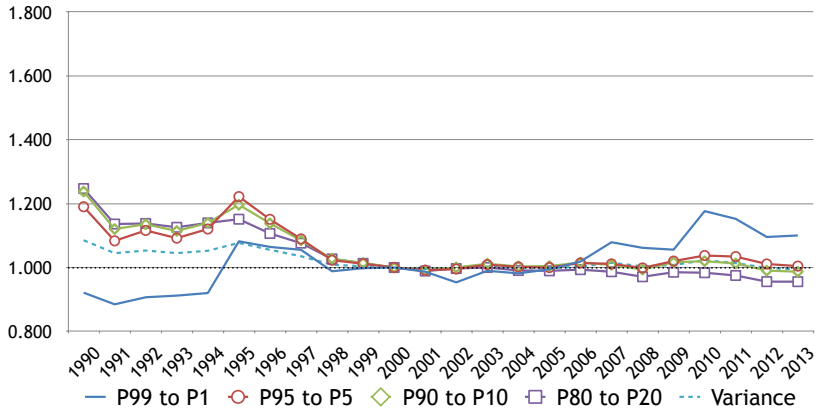
Table B.1: Average Percentiles of the Earnings Distribution by Worker Frame (1995-2013)

	Percentiles								
Frame	*1st*	*5th*	*10th*	*20th*	*50th*	*80th*	*90th*	*95th*	*99th*
All Workers	100	713	1,858	5,141	20,093	43,741	62,277	84,012	173,847
Eligible Workers	132	1,005	2,527	6,463	21,762	45,343	64,021	86,108	178,304
Ratio of EW to AW	1.3195	1.4088	1.3605	1.2572	1.0831	1.0366	1.0280	1.0249	1.0256

Notes: The first row presents the average percentile values from the earnings distribution of *all* workers in all states from 1995-2013. The second row presents the average percentile values from the earnings distribution of *eligible* workers in all states from 1995-2013. The last row computes the ratio of each percentile from the eligible-workers frame to all-workers frame.

Starting with the all-workers frame in Figure B.3, notice that all the measures show a decline in earnings inequality from 1995 to 2000. This can also be seen in Table B.2. The first row presents the average of each ratio from 1995-1999. Notice that they are all above one, meaning that earnings inequality was greater in the late 1990s than in 2000. Then, after 2000, except for the 99/1 ratio, which has a slight upward trend, all other measures of earnings inequality remain relatively stable. The second row of Table B.2 presents the average of each ratio (relative to 2000) from 2001-2013. Notice that aside from the 99/1 ratio, which on average increased by about 5% after 2000, the other measures have remained around their 2000 levels. Thus, aside from differences at the very top or the very bottom of the earnings distribution, earnings inequality among all workers has apparently seen little or no change over the last 10 years.

Figure B.3: Selected Inequality Measures 1990-2013, Relative to 2000 (All Workers)



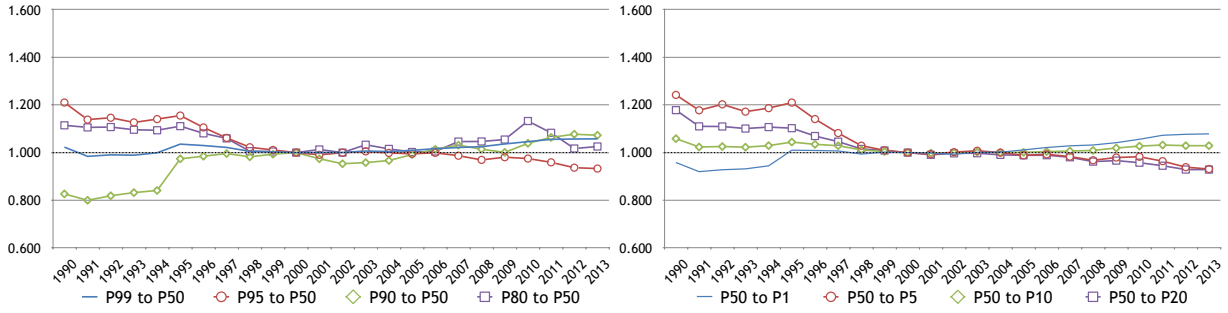
Notes: Subplot (a) presents measures of earnings inequality for *all* workers in all states relative to 2000 from 1990-2013. Subplot (b) presents measures of earnings inequality for *eligible* workers in all states relative to 2000 from 1990-2013. The measures of earnings inequality considered are (i) *P99 to P1*: the ratio of the 99th to the 1st percentile; (ii) *P95 to P5*: the ratio of the 95th to the 5th percentile; (iii) *P90 to P10*: the ratio of the 90th to the 10th percentile; (iv) *P80 to P20* the ratio of the 80th to the 20th percentile; and (v) *Variance*: the variance of log annual earnings.

Table B.2: Inequality Measures Relative to 2000 by Worker Frame

	Inequality Measures				
	99th/1st	*95th/5th*	*90th/10th*	*80th/20th*	*Variance*
			All Workers		
Pre-2000	1.038	1.099	1.092	1.075	1.036
Post-2000	1.050	1.010	1.004	0.983	1.004
Pre-GR	1.001	1.003	1.005	0.992	1.001
GR	1.059	1.009	1.005	0.978	1.006
Post-GR	1.131	1.022	1.002	0.968	1.007
			Eligible Workers		
Pre-2000	1.085	1.119	1.103	1.080	1.047
Post-2000	1.154	1.136	1.114	1.064	1.054
Pre-GR	1.063	1.075	1.067	1.039	1.031
GR	1.209	1.181	1.151	1.084	1.073
Post-GR	1.286	1.222	1.178	1.099	1.086

Notes: The first panel presents measures of earnings inequality for *all* workers in all states relative to 2000, while the second panel presents the same measures for *eligible* workers. The measures of earnings inequality considered are (i) *99th/1st*: the ratio of the 99th to the 1st percentile; (ii) *95th/5th*: the ratio of the 95th to the 5th percentile; (iii) *90th/10th*: the ratio of the 90th to the 10th percentile; (iv) *80th/20th* the ratio of the 80th to the 20th percentile; and (v) *Variance*: the variance of log annual earnings. The values in the table are averages before and after 2000: (i) *pre-2000*: 1995-1999; and (ii) *post-2000*: 2001-2013. The *post-2000* years are further subdivided into three periods: (i) *pre-GR*: 2001-2007; (ii) *GR*: 2008-2009; and (iii) *post-GR* 2010-2013.

Figure B.4: Selected Inequality Measures for the Top and Bottom of the Earnings Distribution 1990-2013, Relative to 2000 (All Workers)

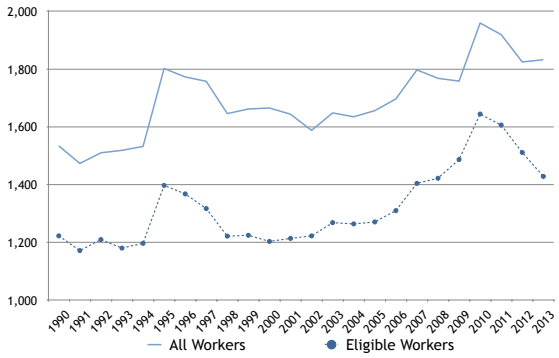


(a) All Workers - Top

(b) All Workers - Bottom

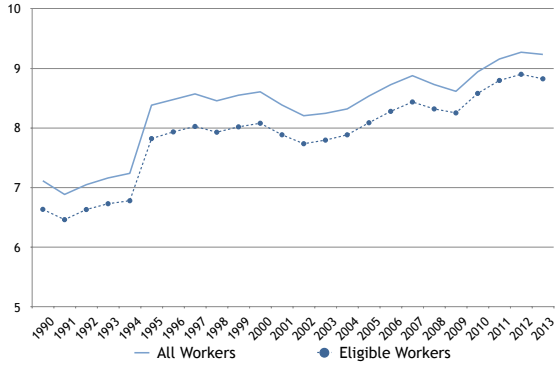
Notes: Subplots (a) and (b) decompose the 99/1 ratio, the 95/5 ratio, the 90/10 ratio, and the 80/20 for eligible workers in all states relative to 2000 from 1990-2013 relative to the median. Subplot (a) plots the following ratios for the top half of the earnings distribution: (i) *P99 to P50*: the ratio of the 99^{th} to the 50^{th} percentile; (ii) *P95 to P50*: the ratio of the 95^{th} to the 50^{th} percentile; (iii) *P90 to P50*: the ratio of the 90^{th} to the 50^{th} percentile; and (iv) *P80 to P50* the ratio of the 80^{th} to the 50^{th} percentile. Subplot (b) plots the following ratios for the bottom half of the earnings distribution: (i) *P50 to P1*: the ratio of the 50^{th} to the 1^{st} percentile; (ii) *P50 to P5*: the ratio of the 50^{th} to the 5^{th} percentile; (iii) *P50 to P10*: the ratio of the 50^{th} to the 10^{th} percentile; and (iv) *P50 to P20* the ratio of the 50^{th} to the 20^{th} percentile. The estimates are based on the all-workers frame from the LEHD infrastructure files.

Figure B.5: Percentile Ratios of the Earnings Distribution by Worker Frame

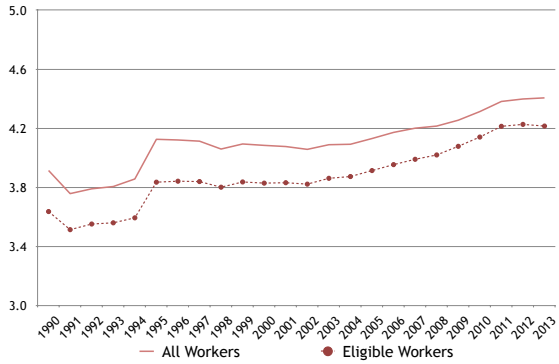


(a) Ratio of 99th to 1st

(b) Ratio of 95th to 5th

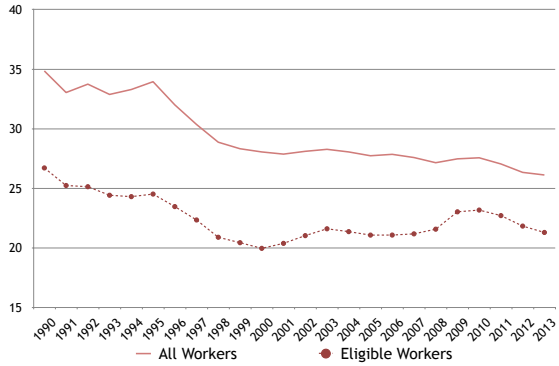


(c) Ratio of 99th to 50th



(d) Ratio of 95th to 50th

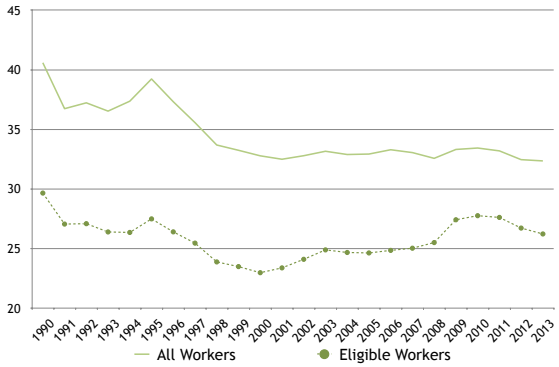
(e) Ratio of 50th to 1st



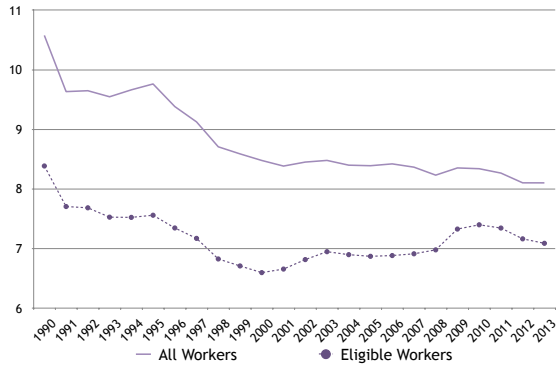
(f) Ratio of 50th to 5th

Notes: This figure plots ratios of top and bottom percentiles for *all* workers (solid lines) and for *eligible* workers (dotted lines). Subplot (a) plots the ratio of the 99th to the 1st percentile by worker frame. This ratio is decomposed into the ratio of the 99th to the 50th percentile in subplot (c) and the ratio of the 50th to the 1st percentile in subplot (e). Subplot (b) plots the ratio of the 95th to the 15th percentile by worker frame. This ratio is decomposed into the ratio of the 95th to the 50th percentile in subplot (d) and the ratio of the 50th to the 5th percentile in subplot (f).

Figure B.6: Percentile Ratios of the Earnings Distribution by Worker Frame



(a) Ratio of 90th to 10th



(b) Ratio of 80th to 20th

(c) Ratio of 90th to 50th

(d) Ratio of 80th to 50th

(e) Ratio of 50th to 10th

(f) Ratio of 50th to 20th

Notes: This figure plots ratios of top and bottom percentiles for *all* workers (solid lines) and for *eligible* workers (dotted lines). Subplot (a) plots the ratio of the 90th to the 10th percentile by worker frame. This ratio is decomposed into the ratio of the 90th to the 50th percentile in subplot (c) and the ratio of the 50th to the 10th percentile in subplot (e). Subplot (b) plots the ratio of the 80th to the 20th percentile by worker frame. This ratio is decomposed into the ratio of the 80th to the 50th percentile in subplot (d) and the ratio of the 50th to the 20th percentile in subplot (f).

C Comparison with Household Surveys

In Section 3, we discussed the trends in earnings inequality based on our analysis of the eligible-workers frame, which we constructed using the LEHD infrastructure data and supplementary information from the Census Bureau's enhanced version of SSA's Numident file. Section 3.1 discussed the highlights of the comparison of our data to the Current Population Survey and the American Community Survey. To put our inequality measures in the context of a broader literature, we compare results based on the administrative data frame discussed in main text with similar measures constructed using household survey data.[55]

C.1 Household Survey Data

To create our household survey analysis file, we use the following records from the Current Population Survey-Annual Social and Economic Supplement (March) and the American Community Survey:

- Current Population Survey Annual Social and Economic Supplement (CPS-ASEC): all persons from survey years 1990-2004
- American Community Survey (ACS): all persons from survey years 2000 to 2013

In the CPS-ASEC, the respondent is surveyed in March and reports earnings for the previous calendar year. We date the earnings accordingly. However, in the ACS the respondent reports earnings for the past 12 months and the survey is in the field continuously throughout the year. Our approach in this case is to date the earnings with the calendar year containing the majority of the months covered by the response, with ties going to the more recent year. As in the LEHD data, nominal earnings are deflated to real 2000 dollars using the CPI-U. In all cases, we used the internal (confidential) versions of the CPS-ASEC and ACS. None of the household survey data are topcoded. We did not replace the Census Bureau's edit and imputation routines with our own. We used the allocated values in the files.

Similar to the workers in LEHD, we consider two samples of individuals from the household surveys. The first includes all individuals. The second isolates workers whose employment should be covered under Unemployment Insurance (including federal employees) and who should, therefore, appear in the LEHD administrative data. We designate a survey respondent as a "covered worker" if he or she meets the following conditions:

- Person interviewed is not living in group quarters
- Individual is employed at a private firm, the local/state/federal government, or is self-employed in an incorporated firm
- Labor earnings are positive
- Individual is between 18 and 70 years old, inclusive.

The last two restrictions combined are included to match the earnings and age restrictions used to identify active eligible workers in the LEHD data.

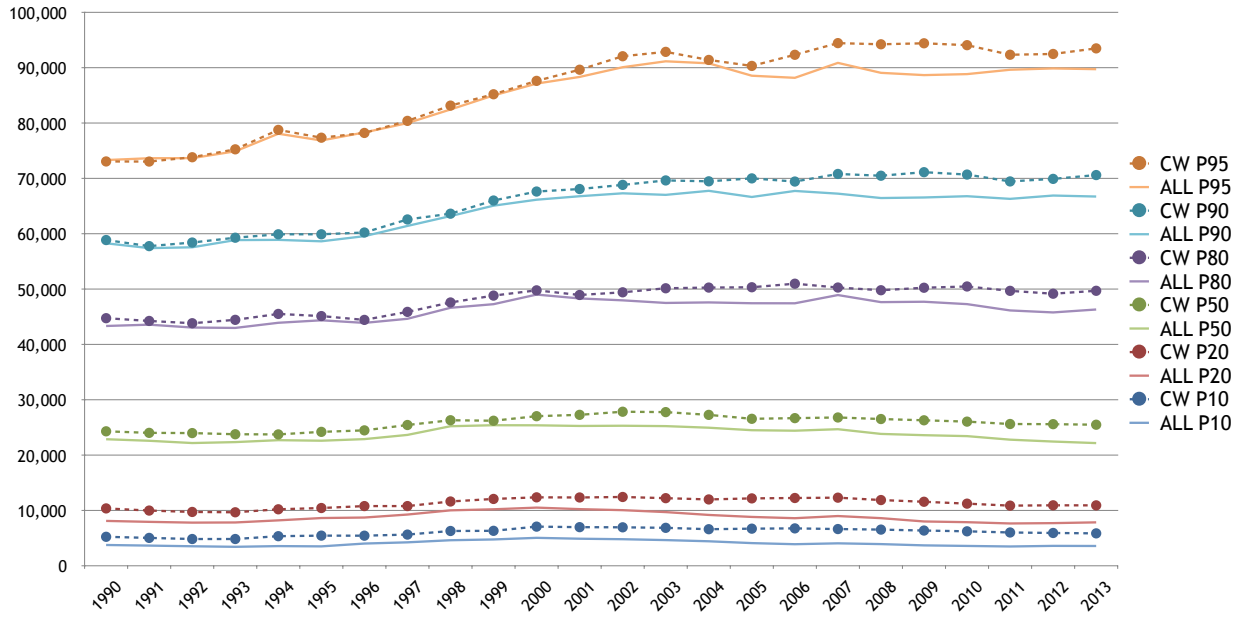
Finally, in most of the results to follow, we do not report results separately for CPS-ASEC and ACS individuals. Instead, in the overlapping years (2000-2003), we interpolate estimates computed from the CPS-ASEC and the ACS to create a single time series using the method in Abowd and Vilhuber (2011).

[55] See Spletzer (2014) for a very similar comparison.

C.2 Comparison of Aggregate Summaries

We start by analyzing how the earnings distribution in household surveys compares to the one computed from administrative records. In the household survey data, the estimated percentile values tend to be greater for covered workers than for all workers. Figure C.1 presents the percentiles of the earnings distribution for all and covered workers in the CPS/ACS. Comparing these values to the ones estimated from the LEHD data, shown in Figure B.1, notice that for percentiles above the median, the values from the eligible-workers frame are fairly close to the ones from the household surveys. Below the median, however, the differences are greater, with the percentiles estimated from the household surveys being much greater than the percentiles estimated from LEHD. For example, notice that earnings associated with the 10^{th} percentile in the CPS/ACS data are close to the 20^{th} percentile in LEHD.

Figure C.1: Percentiles of the Earnings Distribution for All and Covered Workers by Year



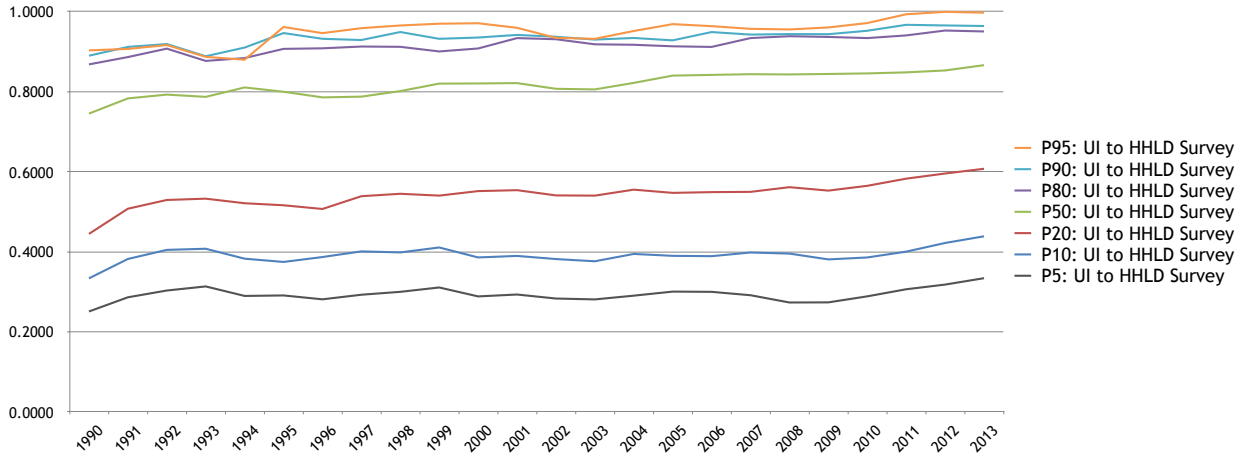
Notes: This figure plots the 10^{th}, 20^{th}, 50^{th}, 80^{th}, 90^{th}, and 95^{th} percentiles of the earnings distribution for all workers (ALL) and for covered workers (CW) in the CPS-ASEC (1990-2003) and the ACS (2000-2013) surveys by year. Covered workers include respondents whose employment relation should be covered as a statutory employee under state UI or as a federal employee, and therefore appear in the LEHD data.

To see these differences in percentiles more clearly, Figure C.2 plots the ratio of the percentiles of the earnings distribution measured using the LEHD eligible-workers frame to the same percentiles measured from the covered workers in the combined CPS/ACS data. First, notice that all the ratios in Figure C.2 are below one, meaning that the percentiles estimated from the household surveys are always greater than the corresponding percentiles estimated from the administrative data. However, the magnitude of this difference varies greatly across the percentiles of the earnings distribution. Specifically, notice that the relative differences in the 95^{th}, 90^{th}, and 80^{th} percentiles are very small compared to the relative differences in the 5^{th}, 10^{th}, and 20^{th} percentiles.

In the main text, Table 4 presents averages of the percentiles from 1995 to 2013 for CPS/ACS and LEHD. Notice that on average the earnings associated with the 80^{th}, 90^{th}, and 95^{th} percentiles are about \$3,500 less in the LEHD data than in CPS/ACS data. Further, as can be seen in

Figure C.2, this gap is decreasing over time, such that in 2013 the difference in the 95^{th} percentile is only \$264. In the bottom half of the earnings distribution, however, a CPS/ACS covered worker earns about \$4,000 more than an LEHD worker at the same point in the earnings distribution. While this absolute difference may not seem that large, relatively, a CPS/ACS worker at the 10^{th} percentile is making 2.54 times more than his LEHD counterpart, and 3.40 times more for a CPS/ACS worker in the 5^{th} percentile. This means that the survey data include more low-earning jobs that are not statutory employment relationships, or are not reported as such to state UI systems. Lastly, notice that the percentiles in LEHD increase faster than their CPS/ACS counterparts since all the ratios exhibit an upward trend, especially after the Great Recession.

Figure C.2: Ratio of UI Earnings to Household Survey Reported Earnings by Percentile



Notes: This figure presents the ratio of earnings for eligible workers in LEHD to the earnings reported by covered workers in CPS/ACS for the 5^{th}, 10^{th}, 20^{th}, 50^{th}, 80^{th}, 90^{th}, and 95^{th} percentile.

To see whether differences in the earnings distribution between workers in CPS/ACS and eligible workers in the LEHD data translate into differences in trends in inequality, we compute various measures of earnings inequality in CPS/ACS and compare them to their LEHD counterparts. In particular, we compute the 95/5, 90/10, and 80/20 ratios, and the variance of log annual earnings. We plot their time series in Figures 5a and 5b for all workers and covered workers, respectively. Both the all-workers and the covered-workers samples show a decline in earnings inequality during the late 1990s that reverses after 2000. However, in the all-workers sample, the magnitude of this increase in inequality in the post-2000 period strongly depends on the measure considered. For example, from 2000 to 2013, the 95/5 ratio increased by 66% from 36.30 to 60.26, while the 90/10 ratio increased by 42% from 12.95 to 18.35. On the other hand, the 80/20 ratio increased by 26% from 4.64 to 5.86, while the variance of log earnings increased by only 5% from 1.23 to 1.26. Thus, while the measures are all trending upwards after 2000 in the all-workers sample, it is unclear whether this increase has been large or small. In the covered-workers sample, earning inequality has also been increasing after 2000, however the magnitude of this increase is relatively consistent across the different measures of earnings inequality. The 95/5 ratio increased by 32% from 2000 to 2013, while the 90/10 ratio increased by 26%. The 80/20 ratio and the variance in log earnings increased less over this period, by about 13% and 14%, respectively. On the other hand, notice that the decline in inequality in the 1990s is very similar across the various measures and samples.

These trends in earnings inequality are very similar to the ones observed among eligible

workers in the LEHD data. Specifically, comparing the time series of earnings inequality for covered workers in CPS/ACS (Figure 5b) to the one for eligible workers in LEHD (Figure 3), notice that the general patterns are very similar. Both of these figures show a decline in inequality during the 1990s and a steady increase in inequality after 2000. The magnitude of this increase is also similar between the covered workers in CPS/ACS and the eligible workers in the LEHD data. Compare the second panel of Table C.1 to the second panel of Table B.2. The second row in both tables shows the average of the 95/5 ratio, the 90/10, ratio, the 80/20 ratio, and the variance of log earnings (relative to 2000) after 2000. Both the covered workers in CPS/ACS and the eligible workers in LEHD saw an increase in the 95/5 ratio and the 90/10 ratio above 10%, and an increase in the 80/20 ratio and the variance of log earnings around 5-6%. Furthermore, most of this increase occurred during or after the Great Recession.

Table C.1: Inequality Measures Relative to 2000 for Workers in Household Surveys

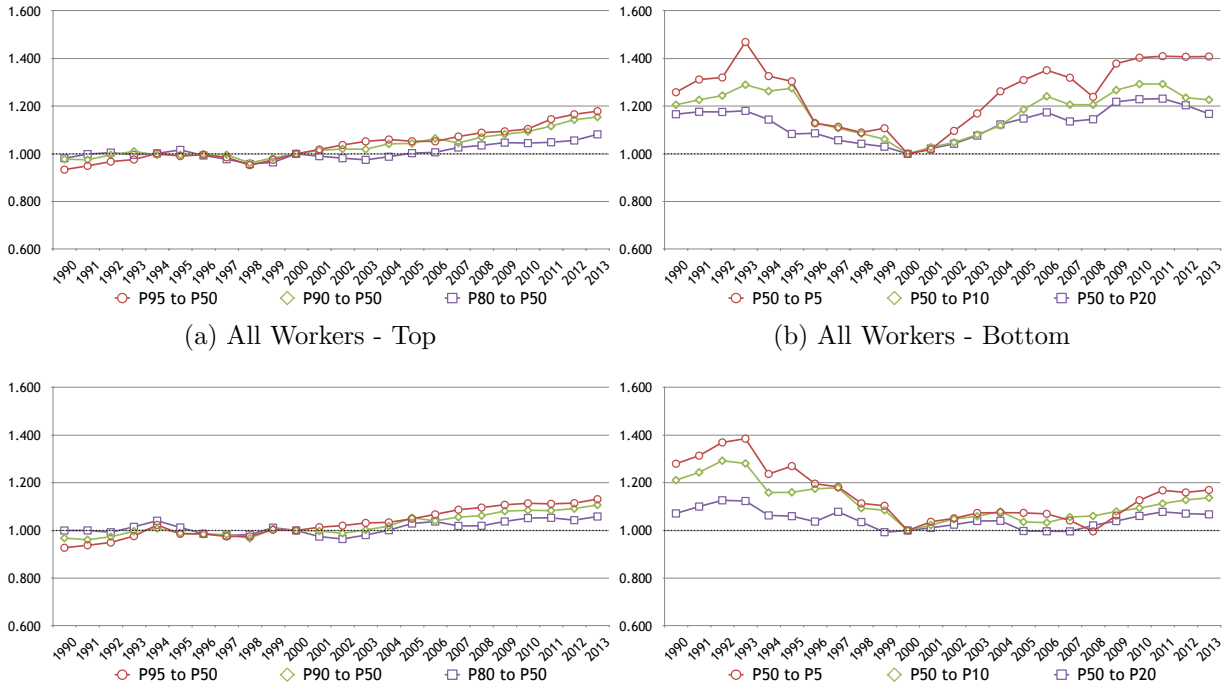
	Inequality Measures			
	95th/5th	*90th/10th*	*80th/20th*	*Variance*
		All Workers		
Pre-2000	1.126	1.118	1.041	1.099
Post-2000	1.406	1.273	1.174	1.001
Pre-GR	1.280	1.171	1.099	0.976
GR	1.429	1.331	1.231	0.998
Post-GR	1.616	1.422	1.278	1.047
		Covered Workers		
Pre-2000	1.156	1.122	1.035	1.082
Post-2000	1.168	1.129	1.056	1.064
Pre-GR	1.106	1.071	1.016	1.040
GR	1.135	1.147	1.060	1.044
Post-GR	1.293	1.221	1.125	1.117

Notes: The first panel presents measures of earnings inequality for *all* workers in CPS/ACS, while the second panel presents the same measures for *covered* workers. The measures of earnings inequality considered are (i) $95^{th}/5^{th}$: the ratio of the 95^{th} to the 5^{th} percentile; (ii) $90^{th}/10^{th}$: the ratio of the 90^{th} to the 10^{th} percentile; (iii) $80^{th}/20^{th}$ the ratio of the 80^{th} to the 20^{th} percentile; and (iv) *Variance*: the variance of log annual earnings. The values in the table are averages before and after 2000: (i) *pre-2000*: 1995-1999; and (ii) *post-2000*: 2001-2013. The *post-2000* years are further subdivided into three periods: (i) *pre-GR*: 2001-2007; (ii) *GR*: 2008-2009; and (iii) *post-GR* 2010-2013. All measures are 1.00 in 2000.

To see whether it is changes in the top or bottom half of the earnings distribution that are driving these trends, we decompose these ratios around the median, as we did using the two worker frames from LEHD. Notice that since 2000 the ratio of the top percentiles to the median has been gradually increasing for both the all-workers sample and the covered-workers sample (Figures C.3a and C.3c). The bottom of the earnings distribution, however, has evolved differently across these two samples. In the all-workers sample, there has been a substantial rise in inequality (Figure C.3b). Among the covered workers, the rise has been much more mild (Figures C.3d). In fact, the trends in earnings inequality among the covered workers is very similar to those observed among the eligible workers in LEHD both in terms of the correlation of the times series and the magnitude of the changes. However, one notable difference is the change in earnings inequality around the Great Recession. In LEHD, inequality increased dramatically during the Great Recession. In CPS/ACS, inequality actually drops substantially just prior to the onset of the Great Recession. However, these gains are lost during the recession years as inequality quickly increases back to trend. Thus, while both the household surveys and the administrative data highlight the

sensitivity of the bottom of the earnings distribution to the Great Recession, the precise cyclical patterns are not consistent across these two data sources.

Figure C.3: Selected Inequality Measures 1990-2013 for the Top and Bottom of the Earnings Distribution, Relative to 2000 (CPS/ACS)



(a) All Workers - Top

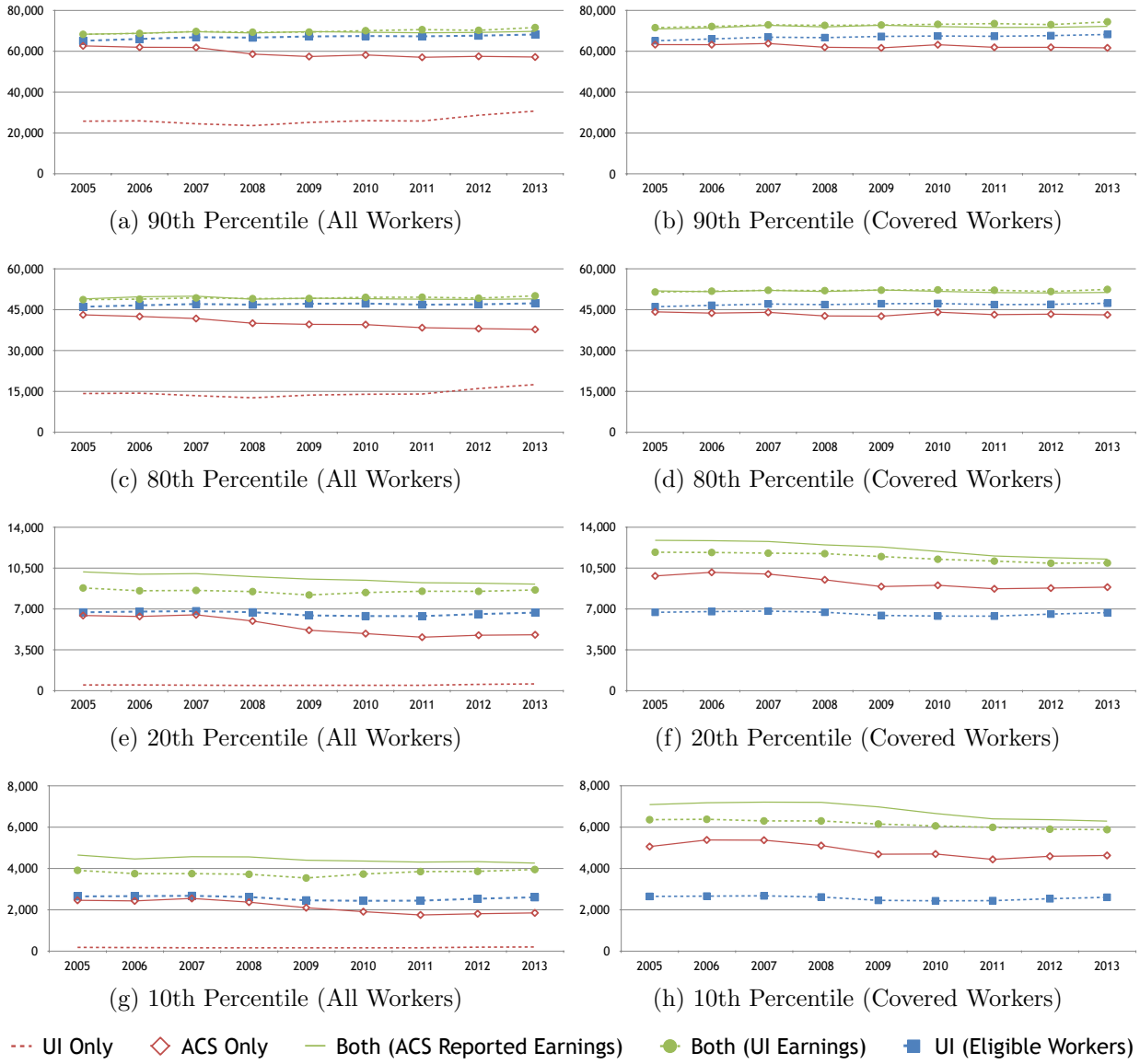
(b) All Workers - Bottom

(c) Covered Workers - Top

(d) Covered Workers - Bottom

Notes: Subplots (a) and (b) decompose the 99/1 ratio, the 95/5 ratio, the 90/10 ratio, and the 80/20 for *all* workers in CPS/ACS relative to 2000 from 1990-2013 relative to the median. Subplot (a) plots the following ratios for the top half of the earnings distribution: (i) *P95 to P50*: the ratio of the 95^{th} to the 50^{th} percentile; (ii) *P90 to P50*: the ratio of the 90^{th} to the 50^{th} percentile; and (iii) *P80 to P50* the ratio of the 80^{th} to the 50^{th} percentile. Subplot (b) plots the following ratios for the bottom half of the earnings distribution: (i) *P50 to P5*: the ratio of the 50th to the 5th percentile; (ii) *P50 to P10*: the ratio of the 50^{th} to the 10^{th} percentile; and (iii) *P50 to P20* the ratio of the 50^{th} to the 20^{th} percentile. Subplots (c) and (d) present the same decomposition for *covered* workers in CPS/ACS relative to 2000 from 1990-2013.

Figure C.4: Comparison of Percentiles in the ACS and LEHD



Notes: This figure plots the 10th, 20th, 80th, and 90th percentiles of the earnings distributions from four samples of the ACS: (i) individuals with positive UI earnings, but no reported ACS earnings (dotted red line); (ii) individuals with positive reported ACS earnings, but no UI earnings (solid red line with diamonds); (iii) individuals with positive reported ACS earnings and positive UI earnings using ACS earnings to compute the earnings distribution (solid green line); and (iv) individuals with positive reported ACS earnings and positive UI earnings using UI earnings to compute the earnings distribution (dotted green line with circles). Subplots (a), (c), (e), and (g) are the percentiles for all workers in ACS. Subplots (b), (d), (f), and (h) are the percentiles for the covered workers in ACS. These are compared to the same percentiles from the eligible workers frame in LEHD (dotted blue line with squares).

C.3 Detailed Analysis of Linked Records

To understand where the discrepancies between the administrative and household survey earnings distributions occur, we analyze the individual ACS records from 2005-2013–linking them to LEHD UI records from the eligible-workers frame using a crosswalk between the two person identifiers

developed and maintained by the Census Bureau. This allows us to see how earnings differ among workers who do and do not match to the LEHD individual data. We focus on records from 2005 forward because, for these years, both the ACS and LEHD are fully national.

For an individual in the ACS, there are three types of matches to the eligible-workers frame in the LEHD data: (i) reported earnings are positive in ACS, but UI earnings are zero; (ii) no reported earnings in ACS, but UI earnings are positive; (iii) both ACS reported earnings and UI earnings are positive. We present these match results in Table C.2. The left panel presents the statistics for all individuals in ACS and the right panel presents the same statistics for covered workers in ACS. When we include all individuals in ACS, about 96% report positive earnings when surveyed in ACS. However, 21% have no UI earnings. A very small fraction of the individuals in ACS, the remaining 4%, have positive UI earnings but did not report any earnings when surveyed. When we consider only covered ACS workers, all these individuals should report positive earnings in ACS. Of these covered workers, 85% also have positive UI earnings and 15% do not match to any UI records.

Table C.2: ACS/UI Match Comparison

All Individuals			**Covered Workers**		
ACS	*UI*	*Percent*	*ACS*	*UI*	*Percent*
earn > 0	earn > 0	75%	earn > 0	earn > 0	85%
earn $= 0$	earn > 0	4%	earn $= 0$	earn > 0	0%
earn > 0	earn $= 0$	21%	earn > 0	earn $= 0$	15%

Notes: The first row reports the fraction of individuals in ACS that report positive earnings when surveyed and match to the eligible-workers frame in the LEHD data, and therefore have positive UI earnings. The second row reports the fraction of individuals in ACS that do not report earnings when surveyed, but match to the LEHD data, and, therefore, have positive UI earnings. The third row reports the fraction of individuals in ACS who do not match to LEHD data. The left-side panel presents the statistics for all individuals in ACS and the right-side panel presents the same statistics for covered workers in ACS.

Using these matched records, we compare the earnings distribution of four samples of ACS individuals in Figure C.5:

- Individuals with positive UI earnings, but no reported ACS earnings (dotted red line)
- Individuals with positive reported ACS earnings, but no UI earnings (solid red line with diamonds)
- Individuals with positive reported ACS earnings and positive UI earnings using ACS earnings to compute the earnings distribution (solid green line)
- Individuals with positive reported ACS earnings and positive UI earnings using UI earnings to compute the earnings distribution (dotted green line with circles).

We compute these distributions for both all workers and covered workers in the ACS. Note that for covered workers, having only UI earnings is vanishingly rare since all covered workers should report positive earnings in the ACS. The earnings distributions from these samples are compared to the one constructed from the eligible-workers frame in LEHD (dotted blue line with squares). Figure C.5 plots the 5^{th}, 50^{th}, and 95^{th} percentiles of these various earnings distributions.[56]

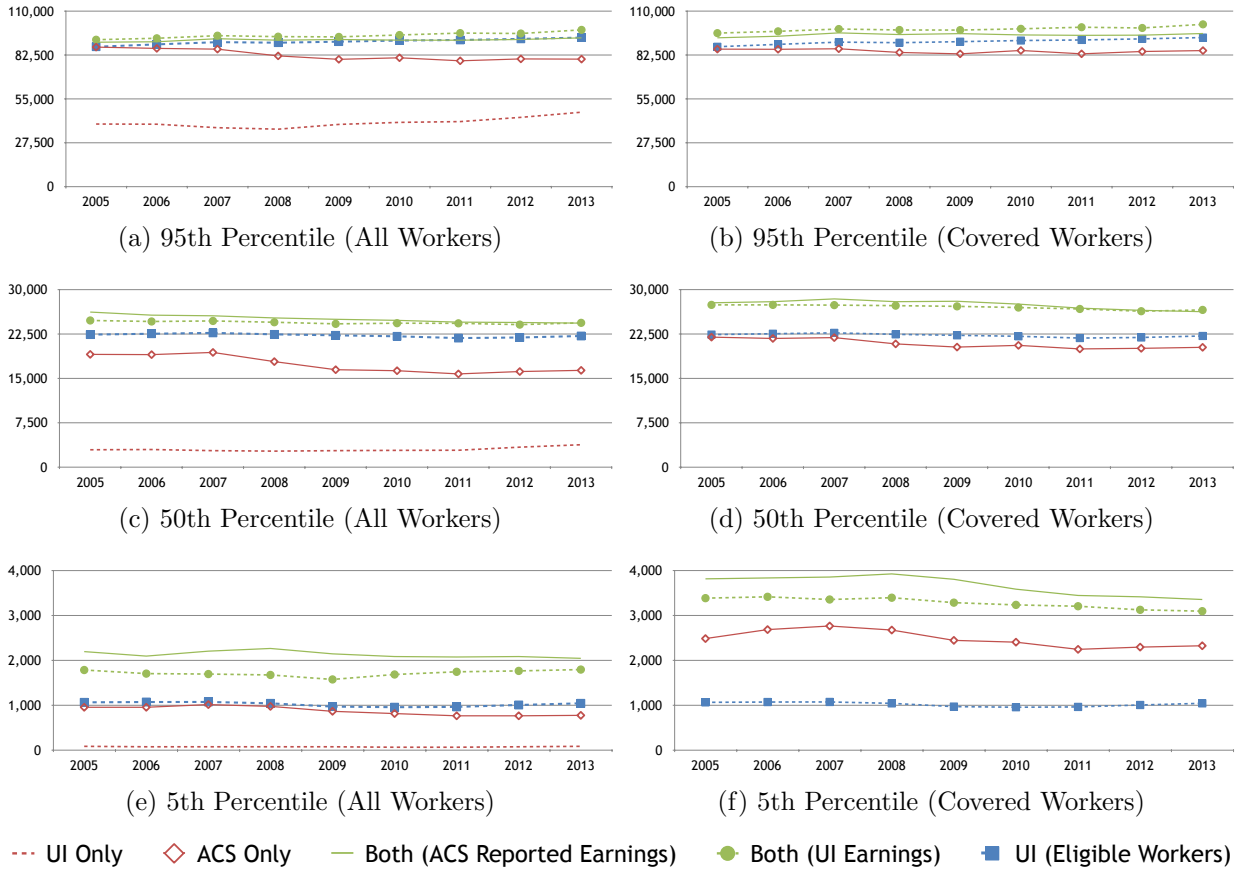
For workers whose earnings are both reported in the ACS and found in LEHD (matched workers), the percentiles computed using the ACS earnings are nearly identical to those computed using UI earnings. Specifically, notice that in Figure C.5 the solid green line and the dotted

[56]For similar comparisons of the 10th, 20th, 80th, and 90th percentiles, see Figure C.4.

green line with circles are very close to each other in all subplots, and especially at and above the median. The differences in the CPS/ACS percentiles and the LEHD percentiles in Figure C.2 are, therefore, very unlikely to be due to misreporting in household surveys. Instead, they must be due to differences in the workers who are surveyed and report earnings in the ACS and those who are found in LEHD. Workers who report positive ACS earnings, but do not match to LEHD (ACS-only) tend to have lower earnings than the workers who do match (solid red lines with diamonds in Figure C.5). However, this gap is less pronounced for workers at the top of the earnings distribution for both the all-workers and covered-workers samples in ACS.

While the ACS-only workers do not earn as much as the matched workers, they do earn significantly more than a large portion of workers in LEHD. This means that the LEHD eligible-workers frame captures more workers in the bottom half of the earnings distribution than the ACS. To see this, notice in Figure C.5b, the 95^{th} percentiles of both the matched and ACS-only samples are nearly identical to the 95^{th} percentile of the eligible-workers earnings distribution in LEHD. However, for the median and lower percentiles, the differences are not trivial. The median matched worker tends to make about 21.5% more than the median eligible-worker in LEHD (≈\$4,770), while the median ACS-only worker makes about 6.4% less (≈\$1,417). At the bottom, the differences in the 5^{th} percentiles are most stark. A matched worker at the 5^{th} percentile tends to make about 3.22 times as much as an eligible worker at the 5^{th} percentile in LEHD (≈\$2,649). Even an ACS-only worker at the 5^{th} percentile makes about 2.44 times as much as a corresponding eligible worker (≈\$1,459). Thus, the left tail of the earnings distribution in ACS is much shorter than the one for eligible workers in the LEHD data, resulting in the LEHD percentiles being less than those computed from household surveys.

Figure C.5: Comparison of Percentiles in the ACS and LEHD



(a) 95th Percentile (All Workers)

(b) 95th Percentile (Covered Workers)

(c) 50th Percentile (All Workers)

(d) 50th Percentile (Covered Workers)

(e) 5th Percentile (All Workers)

(f) 5th Percentile (Covered Workers)

Notes: This figure plots the 5^{th}, 50^{th}, and 95^{th} percentiles of the earnings distributions from four samples of the ACS: (i) individuals with positive UI earnings, but no reported ACS earnings (dotted red line); (ii) individuals with positive reported ACS earnings, but no UI earnings (solid red line with diamonds); (iii) individuals with positive reported ACS earnings and positive UI earnings using ACS earnings to compute the earnings distribution (solid green line); and (iv) individuals with positive reported ACS earnings and positive UI earnings using UI earnings to compute the earnings distribution (dotted green line with circles). Subplots (a), (c), and (e) are the percentiles for all workers in ACS. Subplots (b), (d), and (f) are the percentiles for the covered workers in ACS. These are compared to the same percentiles from the eligible-workers frame in the LEHD data (dotted blue line with squares).

D Inactive Workers and Inequality

In Section 2.2 we tracked both active and inactive workers in our eligible-workers frame. In Section 3.2 we briefly discussed how the treatment of inactivity affects measures of earnings inequality. This Appendix presents details supporting those analyses and conclusions.

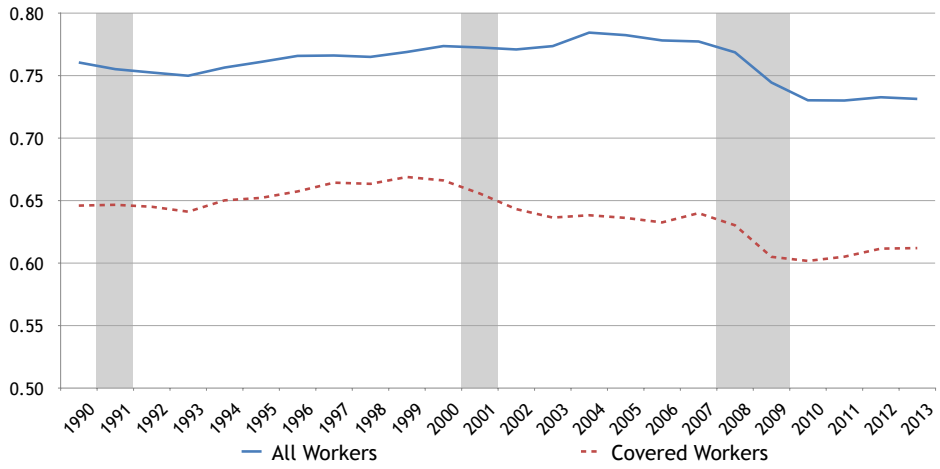
In Appendix B, we excluded inactive workers from the analysis so that we could focus on trends in the ratios of top and bottom percentiles over time. While some inactive workers, given the wages and employment terms on offer, choose to be nonparticipants, others are involuntarily excluded from the labor market. In this section, we present an analysis of how including inactive workers, especially those who were recently employed, affects earnings inequality measures. We begin by analyzing how inactivity has changed in recent years considering comparisons with the employment to population ratio from the CPS/ACS data. Next, we turn our attention to the eligible workers in the LEHD data.

D.1 The Employment-to-Population Ratio

If the U.S. labor market tends to stay relatively close to full employment except for brief periods after the start of a recession, the resultant implied rapid employment growth during a recovery should generate a quick increase in the employment-to-population ratio and a quick decline in the unemployment rate to pre-recession levels. However, our results using annual CPS/ACS survey data show a different pattern around the Great Recession.

Figure D.1 shows the estimated employment-to-population ratio by year from 1990-2013 for all workers (solid blue line) and covered workers (dashed red line) in the CPS/ACS. The NBER identifies three recessions during this period, beginning in the following years: 1990, 2001, and 2008 (December, 2007). Both CPS/ACS series show a dip in 1993 and then a sustained increase until 1999, when the covered-worker sample employment-population ratio begins to decline while the all-workers sample remains relatively flat. Until 1999, the trends for both series are similar, but then the two series diverge, with a decline in the covered workers as a proportion of all workers– suggesting a movement of workers into self-employment. At the beginning of the Great Recession, all three series show a large sustained drop in the employment to population ratio, bottoming out in 2009/2010, with only a mild recovery during the ensuing years. These results suggest that unlike previous recessions, substantial numbers of persons employed prior to the Great Recession did not return to employment even five or more years after the start of the Great Recession. While previous research focused only on employed persons, the large and persistent decrease in the employment-to-population ratio for all workers and for covered workers only, during and after the Great Recession, argues strongly for an expansion of inequality measures to include at least some inactive but eligible workers.

Figure D.1: Employment to Population Ratio (Household Surveys)



Notes: This figure plots the estimated employment-to-population ratio by year from 1990-2013 for all workers (solid blue line) and covered workers (dashed red line) in CPS/ACS. Estimates based on the authors' calculations from the microdata. These are not the official statistics as released by the BLS from the CPS.

D.2 Inactivity-Adjusted Inequality Measures

We estimate three traditional measures of inequality (Gini, Hoover, and Theil), both with and without a category for inactive workers. Deciles of the earnings distribution, estimated as discussed in section 2.4, were used to compute each statistic, with an additional category added for eligible workers with no reported earnings (the inactive category). The earnings value for each person in the inactive category was set to \$1, a modification necessary to facilitate the consistent calculation of all measures (particularly the Theil index, which uses logarithms).[57] We create three samples, each with a different set of eligible but inactive workers:

1. *All eligible workers each year*: This sample assumes all eligible workers are at risk to be employed. Note that this sample is complete and has no dependence on previous years, but the majority of the inactive eligible workers are probably not in the labor force.

2. *Active workers and eligible workers most at risk to be employed*: This sample includes all active workers and workers not active in the current year, but who were active in at least one of the past 4 years. For years prior to 2008 we do not have complete data for every state. In particular, workers with jobs in MA, DC, AR, NH, and MS will be slightly under-represented (see Table A.1). Some of these workers will have earnings in the previous fours year that we do not observe. An upper bound of the impact of this exclusion might be 5% of the jobs in 2004, but the actual impact is likely much less since the largest state, MA, entered in 2002Q1, and is therefore missing only two years of history in 2004Q1. In addition, employment in every state is at risk at some point during 2003, the year a worker not employed in 2004 is most likely to have previously been employed.

3. *Only active workers*: This sample includes only active workers, so no modifications are made to the standard formulas.

[57] The results for the Gini and Hoover measures using \$0 show very small differences in levels and identical trends compared with setting the earnings value to \$1.

Table D.1 shows the results. The first panel is for all eligible workers, while the second panel shows results for workers most at risk to be eligible workers. The last panel includes only active eligible workers.

All three of the inequality measures increase substantially as the proportion of eligible workers included in the calculation increases. Not surprisingly, including a large block of workers with only $1 of annual earnings greatly increases measured inequality. Comparing our results with another administrative data source, estimates of inequality using SSA data, we find that ours are somewhat larger, although the exact source of the difference is unclear due to coverage differences imposed on the SSA estimation sample. For example, in 2004 the estimated Gini coefficient using a restricted sample of currently eligible SSA recipients is 0.471, while in our data the estimated Gini is 0.510 (Kopczuk et al., 2010).

Table D.1: Inequality Measures with and without Inactive Workers

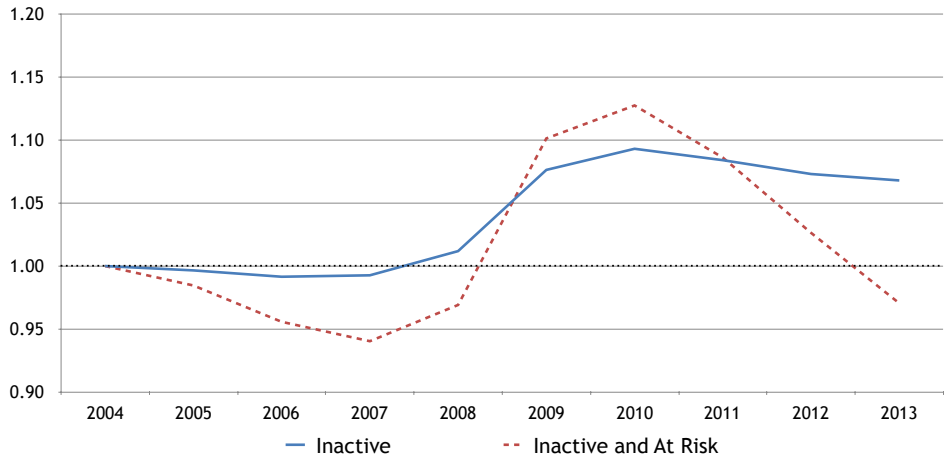
Year	Persons	Gini	Hoover	Theil
All Eligible Workers				
2004	219,763,469	0.696	0.538	2.379
2005	222,160,089	0.697	0.538	2.379
2006	224,721,578	0.698	0.539	2.377
2007	227,553,012	0.699	0.540	2.386
2008	230,355,015	0.702	0.544	2.416
2009	232,813,313	0.714	0.558	2.535
2010	234,304,705	0.720	0.564	2.576
2011	235,429,997	0.720	0.563	2.563
2012	236,484,312	0.719	0.560	2.547
2013	237,816,938	0.716	0.558	2.532
Eligible Workers with Earnings in the Current or at Least One of the Past 4 Years				
2004	164,243,214	0.593	0.437	1.352
2005	165,892,505	0.594	0.438	1.346
2006	167,417,542	0.594	0.438	1.331
2007	168,988,105	0.595	0.439	1.327
2008	170,229,709	0.597	0.441	1.351
2009	170,241,870	0.609	0.452	1.472
2010	170,617,692	0.616	0.458	1.509
2011	171,015,983	0.615	0.457	1.480
2012	170,986,772	0.611	0.454	1.437
2013	170,735,917	0.604	0.448	1.387
Eligible Workers with Earnings in the Current Year				
2004	136,562,515	0.510	0.369	0.529
2005	138,340,770	0.513	0.372	0.535
2006	140,363,860	0.516	0.375	0.541
2007	142,034,418	0.519	0.378	0.546
2008	142,109,590	0.517	0.377	0.543
2009	137,948,364	0.517	0.378	0.546
2010	137,345,658	0.522	0.382	0.557
2011	138,810,297	0.525	0.385	0.562
2012	140,415,325	0.527	0.386	0.563
2013	141,665,611	0.523	0.384	0.555

Notes: This table presents traditional measures of inequality (Gini, Hoover, and Theil) for three samples of persons: (i) all eligible workers (top panel), (ii) most at-risk eligible workers (middle panel), and (iii) only active workers (bottom panel).

Figure D.2 shows the share of eligible workers who are inactive (solid blue line) and the share who are most at risk to be active (dashed red line) relative to the base year 2004. The solid blue

line represents the share of eligible workers not currently working each year–the difference between the number of workers in the first panel of Table D.1 and the number of workers in the third panel. The dashed red line represents the share of workers most at risk to be active not currently working each year–the difference between the number of workers in the second panel of Table D.1 and the number of workers in the third panel. The dashed red line is noticeably more responsive to changes in labor demand, suggesting that we chose a reasonable group to represent the workers most at risk to be active. However, a closer look at the source of the decline in the most at risk group (red line) from 2011 forward shows that the decline is due both to the growth in employment during the recovery and a lack of growth in the number of workers most at risk to be active. Many of the at risk workers who had positive earnings just prior to or at the start of the Great Recession have not had positive earnings in the subsequent four years. By 2011, they are dropping out of the at risk group. It is difficult to know the labor force status of these workers due to limitations of administrative data, it does highlight the benefit of having multiple measures of inactive status for the eligible-workers population.

Figure D.2: Inequality Measures



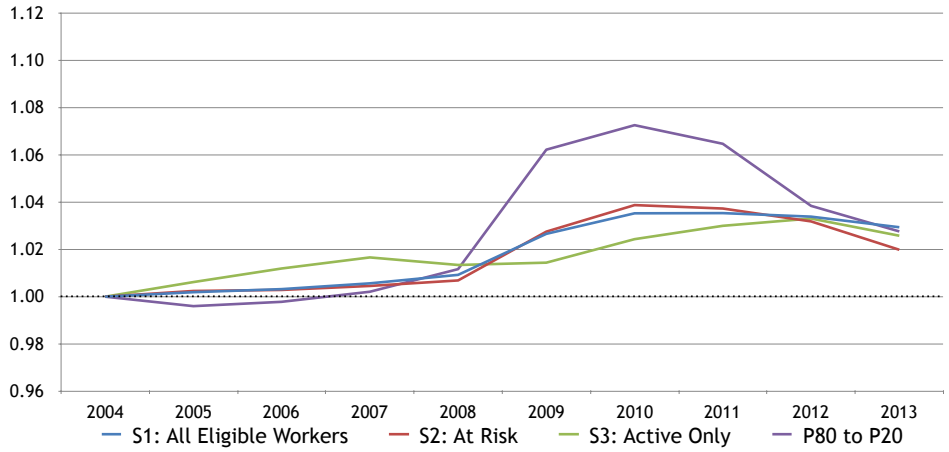
Notes: This figure shows the share of eligible workers who are inactive (solid blue line) and the share who are "at risk" (dashed red line) relative to the base year 2004. In a given year, a person is *inactive* if that person did not make positive earnings that year. In a given year, a person is "*at risk*" if that person did not make positive earnings that year, but did make positive earnings sometime in the last four years.

In Section 3 we documented the increase in inequality post-2000 using ratios of various percentiles of the earnings distribution. For the eligible-workers frame, the increase in earnings of the top 20% relative to the bottom 20% of earners accelerates during the Great Recession, with annual earnings increases for workers at the 80th percentile and small declines or no increases for at the 20th percentile. The increases for the 99/1 ratio, the 95/5 ratio, and the 90/10 ratio are even larger, with the ratios increasing faster the more extreme the comparison (Figure 3). Here we have taken an alternative approach. Instead of comparing two specific points in the earnings distribution, the portmanteau inequality measures presented here weight the changes occurring across the earnings distribution and combine them to produce a single measure of overall inequality. Each measure uses different weights and combining rules, therefore it is useful to compare each approach.

The relative changes in the Gini coefficients for each of the three samples are presented in Figure D.3. The results for the first two samples are almost identical. The Gini coefficients for

the third sample, only active workers, grow faster before the Great Recession, but do not show the increase in inequality at the start of the Great Recession present for the eligible-workers and at-risk-workers samples. Part of the reason for this difference is that the Gini coefficient is very sensitive to changes in earnings at the top of the distribution. At the beginning of the recession, earnings at the top of the distribution declined or stagnated. In spite of the large number of workers moving from active to inactive status at the beginning of the Great Recession, the Gini coefficient for the active-only sample shows inequality declining, although it does start to climb as earnings growth at the top of the distribution resumes in 2009. In contrast, the Gini coefficients for the all-eligible, and most at-risk samples show increasing inequality at the start of the Great Recession, similar to the 80/20 ratio (also shown in the figure).

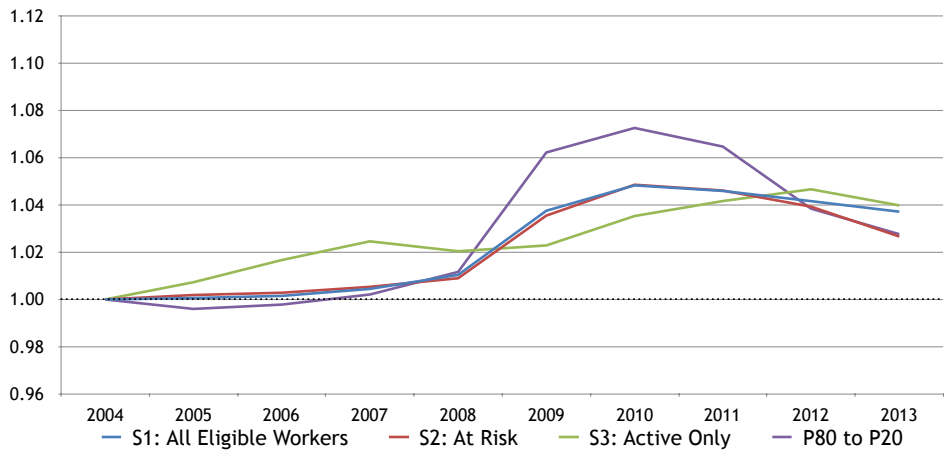
Figure D.3: Inequality Measures – Gini Coefficient



Notes: This figure plots the Gini coefficient for three samples of eligible workers: (i) *All Eligible Workers*: includes active and all inactive workers, (ii) *At Risk*: includes active workers and inactive workers who made positive earnings sometime in the last four years, and (iii) *Active Only*: includes only active workers. The ratio of the 80^{th} to the 20^{th} percentile is also plotted for reference.

The results for the Hoover index, shown in Figure D.4, are similar to the Gini coefficient, although the relative increase in inequality during the Great Recession is larger when measured using the Hoover index. The increase before the Great Recession is also larger when using only active workers.

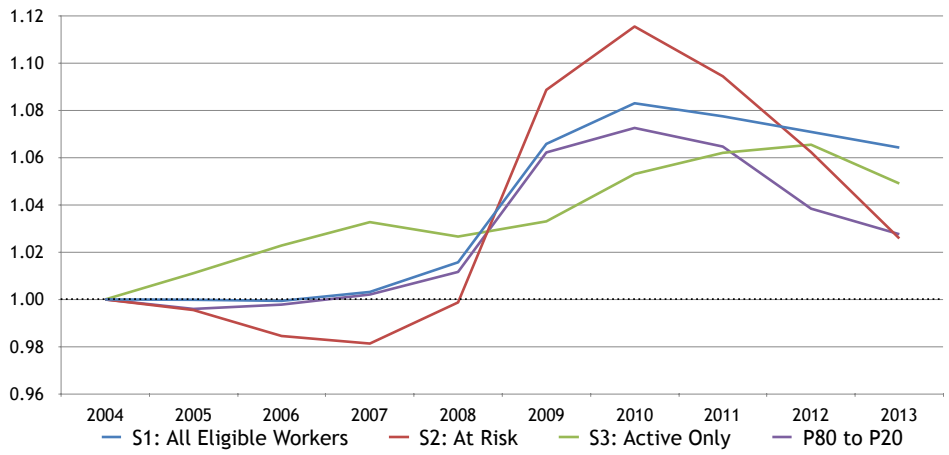
Figure D.4: Inequality Measures – Hoover Index



Notes: This figure plots the Hoover index for three samples of eligible workers: (i) *All Eligible Workers*: includes active and all inactive workers, (ii) *At Risk*: includes active workers and inactive workers who made positive earnings sometime in the last four years, and (iii) *Active Only*: includes only active workers. The ratio of the 80^{th} to the 20^{th} percentile is also plotted for reference.

The final measure we consider is the symmetric Theil index. The results using this measure are shown in Figure D.5. Over the entire period, the Theil measure is more responsive to earnings distribution changes than either the Gini coefficient or the Hoover index, but it is especially responsive to the addition of inactive workers. The relative change in the Theil index computed using all eligible workers (sample one) is almost identical to the 80/20 ratio through 2009, with greater inequality after that 2009 reflecting the slow decline in inactive workers during the recovery. The relative change in the Theil index computed using only the most at-risk workers (sample two) could arguably be viewed as an exaggerated version of the 80/20 ratio. The inclusion of inactive at-risk workers in sample two introduces additional information into the Theil index calculation, magnifying the decline in inequality prior to the Great Recession, the increase during the Great Recession, and the decline during the recovery.

Figure D.5: Inequality Measures – Theil Index



Notes: This figure plots the Theil index for three samples of eligible workers: (i) *All Eligible Workers*: includes active and all inactive workers, (ii) *At Risk*: includes active workers and inactive workers who made positive earnings sometime in the last four years, and (iii) *Active Only*: includes only active workers. The ratio of the 80^{th} to the 20^{th} percentile is also plotted for reference.

Introducing information about inactive but at-risk workers into the calculation of the Gini coefficient and Hoover index changes the trend, but the inequality levels in 2013 are largely the same relative to 2004 using either measure. The Theil index changes in similar ways with the addition of inactive but at-risk workers; however, the Theil index is much more sensitive to both changes in the earnings distribution and the addition of inactive workers. The growth in the Theil index using only active workers is larger than either the Gini or Hoover index. Similar to the Gini and the Hoover indices, by not including inactive workers the Theil index fails to capture the increase in inequality at the start of the Great Recession. Adding inactive workers to the Theil index (sample one), results in a measure similar to the 80/20 ratio through 2009; after 2009 the two measures diverge due to the slow decline in the number inactive workers during the recovery from the Great Recession. The Theil index for the most at-risk workers (sample two) shows the largest changes in inequality.

Although it is unclear which of the adjusted inequality measures correctly weights the inactive workers, it is worthwhile to consider adjusted measures that count at least some of the zero-earning workers as part of any general analysis of changes in earnings inequality.

E Decomposing Changes in the Earnings Distribution

In Section 4 we presented the evolution of the earnings/inactivity distribution in terms of the year to year flows of workers across different parts of the earnings distributions and into and out of active status.

E.1 Worker Flows

Starting in 2005, each year we calculate the change in the number of workers between the current and the previous year for the four earnings/inactivity categories. The year-to-year change in the number of workers in a specific category is driven by changes in the number of workers entering (inflows) and the number of workers leaving (outflows). Specifically, to compute the flows between two employment states, let A and B be arrays of counts for each category in years $t-1$ and t, respectively:

$$\begin{aligned} \text{year } t-1: \quad A &= [\ a_0 \quad a_1 \quad a_2 \quad a_3 \quad a_4\ ] \\ \text{year } t: \quad B &= [\ b_0 \quad b_1 \quad b_2 \quad b_3 \quad b_4\ ] \end{aligned}$$

In order to complete the decomposition and capture all possible transitions we must add an additional category, zero, representing workers not eligible to work in one of the two periods, but who are eligible to work in the other. Let C_{AB} be the transition matrix of counts:

$$C_{AB} = \begin{bmatrix} c_{00} & c_{01} & c_{02} & c_{03} & c_{04} \\ c_{10} & c_{11} & c_{12} & c_{13} & c_{14} \\ c_{20} & c_{21} & c_{22} & c_{23} & c_{24} \\ c_{30} & c_{31} & c_{32} & c_{33} & c_{34} \\ c_{40} & c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix}$$

The rows of the transition matrix represent the origin state (A) and the columns represent the destination state (B). For example, c_{21} is the number of workers who were in the bottom 20% of the overall-earnings distribution in year $t-1$ and transition to the eligible but no reported UI earnings category in year t.

To compute the total net inflows into an employment category, we first introduce some notation. Let ι be a (5×1) column vector of ones. Then:

$$\begin{aligned} C_{A\bullet} &= C_{AB} \cdot \iota = \text{outflows + stayers} \\ C_{\bullet B} &= C_{AB}^T \cdot \iota = \text{inflows + stayers} \end{aligned}$$

Net inflows into each employment state Δ_{AB}^C are defined as:

$$\begin{aligned} \Delta_{AB}^C &\equiv B - A \\ &= C_{\bullet B} - C_{A\bullet} \\ &= \underbrace{C_{AB}^T \cdot \iota}_{\text{inflows + stayers}} - \underbrace{C_{AB} \cdot \iota}_{\text{outflows + stayers}} \\ &= \underbrace{(C_{AB}^T - C_{AB}) \cdot \iota}_{\text{inflows} - \text{outflows} = \text{net inflows}} \end{aligned} \tag{E-1}$$

Note the position of the stayers on the main diagonal. When we take the difference between C_{AB}^T and C_{AB}, the resulting matrix will have zeros on the main diagonal, showing that the stayers do not directly affect the earnings distribution except through changes in average earnings. It should also be noted there is a direct relationship between the number of outflows and the number of stayers. If more workers leave a given category then there will be fewer stayers, ceteris paribus.

Table E.1 provides descriptive statistics on the individuals in each earnings category. It is expanded in the main text in Table 5, which shows the net change in workers between the previous year and the current year from 2005-2013.

The flows of workers affect the earnings distribution, but the average earnings of each category and the change in average earnings for stayers also affect the change in the earnings distribution. Here we show the complete decomposition of the change in the earnings distribution. Table E.2 shows the earnings changes we decompose here. Unlike Table 5 in the main text, the decomposition for earnings does not include net inflows into the eligible-worker frame or net inflows to inactive status. As we show below, these flows have no associated earnings and therefore have a weight of zero.

The corresponding earnings transition matrix for a given transition matrix of counts C_{AB} is:

$$E_{AB} = \begin{bmatrix} 0 & 0 & e_{02} & e_{03} & e_{04} \\ 0 & 0 & e_{12} & e_{13} & e_{14} \\ e_{20} & e_{21} & e_{22} & e_{23} & e_{24} \\ e_{30} & e_{31} & e_{32} & e_{33} & e_{34} \\ e_{40} & e_{41} & e_{42} & e_{43} & e_{44} \end{bmatrix}$$

Unlike the transition matrix of counts, each element of the transition matrix of earnings has two associated total earnings values, the total earnings for the workers in period A and the total earnings for those same workers in period B. Each element of the earnings transition matrix is an ordered pair of elements. For example, $e_{23} = \{e_{23}^A, e_{23}^B\}$ represents the earnings of workers moving from the bottom 20% to the middle 60% of the earnings distribution. The first element is the total earnings in the previous period (when each worker is in the bottom 20%) and the second element is the total earnings in the current period (when each worker is in the middle 60%. Elements with an ordered pair of two zeros are shown as zeros in the earnings transition matrix.

Applying the net inflow formulas for the counts to the earnings transition matrix,

$$\Delta_{AB}^E = \underbrace{(E'_{AB} - E_{AB}) \cdot i}_{\text{net inflows}} \tag{E-2}$$

and choosing the appropriate earnings value from each tuple, using an A or B superscript to indicate the first or second element chosen, respectively, we have:

$$\Delta_{AB}^E = \begin{bmatrix} (0-0) & (0-0) & (e_{20}^B - e_{02}^A) & (e_{30}^B - e_{03}^A) & (e_{40}^B - e_{04}^A) \\ (0-0) & (0-0) & (e_{21}^B - e_{12}^A) & (e_{31}^B - e_{13}^A) & (e_{41}^B - e_{14}^A) \\ (e_{02}^B - e_{20}^A) & (e_{12}^B - e_{21}^A) & (e_{22}^B - e_{22}^A) & (e_{32}^B - e_{23}^A) & (e_{42}^B - e_{24}^A) \\ (e_{03}^B - e_{30}^A) & (e_{13}^B - e_{31}^A) & (e_{23}^B - e_{32}^A) & (e_{33}^B - e_{33}^A) & (e_{43}^B - e_{34}^A) \\ (e_{04}^B - e_{40}^A) & (e_{14}^B - e_{41}^A) & (e_{24}^B - e_{42}^A) & (e_{34}^B - e_{43}^A) & (e_{44}^B - e_{44}^A) \end{bmatrix} \cdot i.$$

The sum of each row in the matrix is the net inflow for each category of the earnings/inactivity

Table E.1: Descriptive Statistics by Earnings Categories

	1: Eligible, No Earn	**2: Bottom 20%**	**3: Middle 60%**	**4: Top 20%**	**Total**
Year		*Number of Eligible Workers*			
2004	83,200,954	27,062,314	82,821,341	26,678,860	219,763,469
2005	83,819,319	27,376,301	84,079,363	26,885,106	222,160,089
2006	84,357,718	27,598,826	84,946,369	27,818,665	224,721,578
2007	85,518,594	27,800,774	85,576,064	28,657,580	227,553,012
2008	88,245,425	28,120,283	85,548,690	28,440,617	230,355,015
2009	94,864,949	28,119,169	81,894,162	27,935,033	232,813,313
2010	96,959,047	28,154,014	81,314,722	27,876,922	234,304,705
2011	96,619,700	28,498,111	82,538,961	27,773,225	235,429,997
2012	96,068,987	28,269,636	83,930,862	28,214,827	236,484,312
2013	96,151,327	28,119,381	84,707,469	28,838,761	237,816,938
Year		*Total Earnings (Millions of Real (2000) Dollars)*			
2004	—	76,178	1,959,201	2,351,882	4,387,260
2005	—	77,118	1,984,925	2,407,259	4,469,302
2006	—	77,653	2,006,111	2,529,269	4,613,033
2007	—	78,142	2,021,497	2,636,516	4,736,155
2008	—	78,716	2,012,397	2,576,185	4,667,298
2009	—	77,793	1,923,326	2,488,291	4,489,410
2010	—	77,788	1,901,588	2,524,307	4,503,683
2011	—	79,000	1,918,544	2,542,238	4,539,782
2012	—	78,880	1,947,808	2,625,836	4,652,524
2013	—	78,850	1,969,953	2,657,238	4,706,041
Year		*Average Earnings per Worker* ($e_{it} > 0$)			
2004	—	2,815	23,656	88,155	32,126
2005	—	2,817	23,608	89,539	32,306
2006	—	2,814	23,616	90,920	32,865
2007	—	2,811	23,622	92,001	33,345
2008	—	2,799	23,523	90,581	32,843
2009	—	2,767	23,486	89,074	32,544
2010	—	2,763	23,386	90,552	32,791
2011	—	2,772	23,244	91,536	32,705
2012	—	2,790	23,207	93,066	33,134
2013	—	2,804	23,256	92,141	33,219
Variable		*Cumulative Change (2004-2013)*			
Number of	12,950,373	1,057,067	1,886,128	2,159,901	18,053,469
Workers	14.4%	3.8%	2.3%	7.8%	7.9%
Total	—	2,671	10,752	305,357	318,780
Earnings	—	3.4%	0.5%	12.2%	7.0%
Average	—	-11	-400	3,986	-175
Earnings	—	-0.4%	-1.7%	4.4%	-0.9%

Notes: The cumulative change in average earnings includes workers with $e_{it} = 0$ (column 1) in the denominator. The overall change for the entire period for workers with $e_{it} > 0$ is 3.3%.

Table E.2: Earnings Associated with Exit from and Entry to Each Earnings Category

Year	Earn $t-1$	Earn t	Net Change	Stayers	Outflows	Inflows	Inflows− Outflows	Net Change
Bottom 20% of the Overall UI Earnings Distribution								
2005	76,178	77,118	939	1,625	41,849	41,164	-685	939
2006	77,118	77,653	535	1,752	42,340	41,123	-1,217	535
2007	77,653	78,142	489	1,553	42,415	41,351	-1,065	489
2008	78,142	78,716	575	337	41,662	41,900	237	575
2009	78,716	77,793	-923	-1,193	41,681	41,951	270	-923
2010	77,793	77,788	-5	1,401	42,571	41,165	-1,406	-5
2011	77,788	79,000	1,212	1,948	42,359	41,622	-736	1,212
2012	79,000	78,880	-120	2,680	43,350	40,550	-2,800	-120
2013	78,880	78,850	-30	2,637	42,914	40,246	-2,668	-30
Middle 60% of the Overall UI Earnings Distribution								
2005	1,959,201	1,984,925	25,725	37,258	278,555	267,021	-11,534	25,725
2006	1,984,925	2,006,111	21,186	55,382	292,830	258,634	-34,196	21,186
2007	2,006,111	2,021,497	15,386	53,012	296,600	258,975	-37,626	15,386
2008	2,021,497	2,012,397	-9,101	15,411	288,018	263,506	-24,512	-9,101
2009	2,012,397	1,923,326	-89,071	4,842	331,453	237,541	-93,912	-89,071
2010	1,923,326	1,901,588	-21,738	23,095	289,271	244,438	-44,833	-21,738
2011	1,901,588	1,918,544	16,956	22,643	263,326	257,639	-5,687	16,956
2012	1,918,544	1,947,808	29,264	47,349	266,666	248,581	-18,085	29,264
2013	1,947,808	1,969,953	22,144	58,469	273,520	237,196	-36,324	22,144
Top 20% of the Overall UI Earnings Distribution								
2005	2,351,882	2,407,259	55,377	64,813	245,494	236,058	-9,436	55,377
2006	2,407,259	2,529,269	122,010	88,284	227,727	261,453	33,726	122,010
2007	2,529,269	2,636,516	107,247	86,390	240,848	261,705	20,857	107,247
2008	2,636,516	2,576,185	-60,330	-15,291	271,995	226,955	-45,040	-60,330
2009	2,576,185	2,488,291	-87,894	-22,790	291,186	226,082	-65,104	-87,894
2010	2,488,291	2,524,307	36,016	67,434	246,006	214,587	-31,418	36,016
2011	2,524,307	2,542,238	17,931	44,185	230,451	204,197	-26,254	17,931
2012	2,542,238	2,625,836	83,598	78,243	214,172	219,527	5,355	83,598
2013	2,625,836	2,657,238	31,403	28,123	217,801	221,081	3,280	31,403

Notes: The estimates are based on the authors' calculations using transitions into and out of the eligible-workers frame and between categories of the earnings distributions, including inactive workers.

distribution. The sum of the first two rows is zero; each element of the first two rows is zero, there are no earning when not eligible or eligible but inactive. Multiplying each element of the next three rows by a conformable vector of ones we can separate each total earnings value into the product of average earnings and the counts for that value. For example, the net inflows between period A and period B for earnings category two is:

$$\begin{aligned}\Delta_{AB}^{E2} = &(\bar{e}_{02}^{B} \cdot c_{02} - \bar{e}_{20}^{A} \cdot c_{20}) + (\bar{e}_{12}^{B} \cdot c_{12} - \bar{e}_{21}^{A} \cdot c_{21}) \\ &+ (\bar{e}_{22}^{B} - \bar{e}_{22}^{A}) \cdot c_{22} + (\bar{e}_{32}^{B} \cdot c_{32} - \bar{e}_{23}^{A} \cdot c_{23}) \\ &+ (\bar{e}_{42}^{B} \cdot c_{42} - \bar{e}_{24}^{A} \cdot c_{24}) \end{aligned} \tag{E-3}$$

The year-to-year change in the earnings associated with a given part of the earnings distribution is a linear function (weighted sum) of the average earnings and the transition counts. Table E.2 shows the results, after first grouping the stayers, inflows, and outflows together for the bottom 20%, middle 60%, and top 20% categories.

The change in earnings reduces to a simple (signed) sum of the counts if the average earnings

is the same for each flow, i.e. $(\bar{e}_2^* = \bar{e}_{02}^B = \bar{e}_{20}^A = \bar{e}_{12}^B = \bar{e}_{21}^A = \bar{e}_{22}^B = \bar{e}_{22}^A = \bar{e}_{32}^B = \bar{e}_{23}^A = \bar{e}_{42}^B = \bar{e}_{24}^A)$.

$$\Delta_{AB}^{E2} = \bar{e}_2^* \cdot \big[\underbrace{(c_{02} + c_{12} + c_{32} + c_{42})}_{\text{inflows}} - \underbrace{(c_{20} + c_{21} + c_{23} + c_{24})}_{\text{outflows}} \big] \tag{E-4}$$

Although the simple formula will rarely hold in practice, it is useful as the earnings change for each category is now a scaled function of the counts. For the data in this paper a different constant average earnings value for each category does a reasonable job approximating the gross outflows and inflows. However, when using a constant the individual flows are not always scaled correctly since the weights (average earnings) differ substantially in some cases. Even though there are level differences across flows, the average earnings values are for the most part stable over time, allowing the counts to proxy for the change in the earnings distribution over time, once the appropriate scale factor is known for a given flow. The table below shows the average earnings and measures of variability for each of the flows.

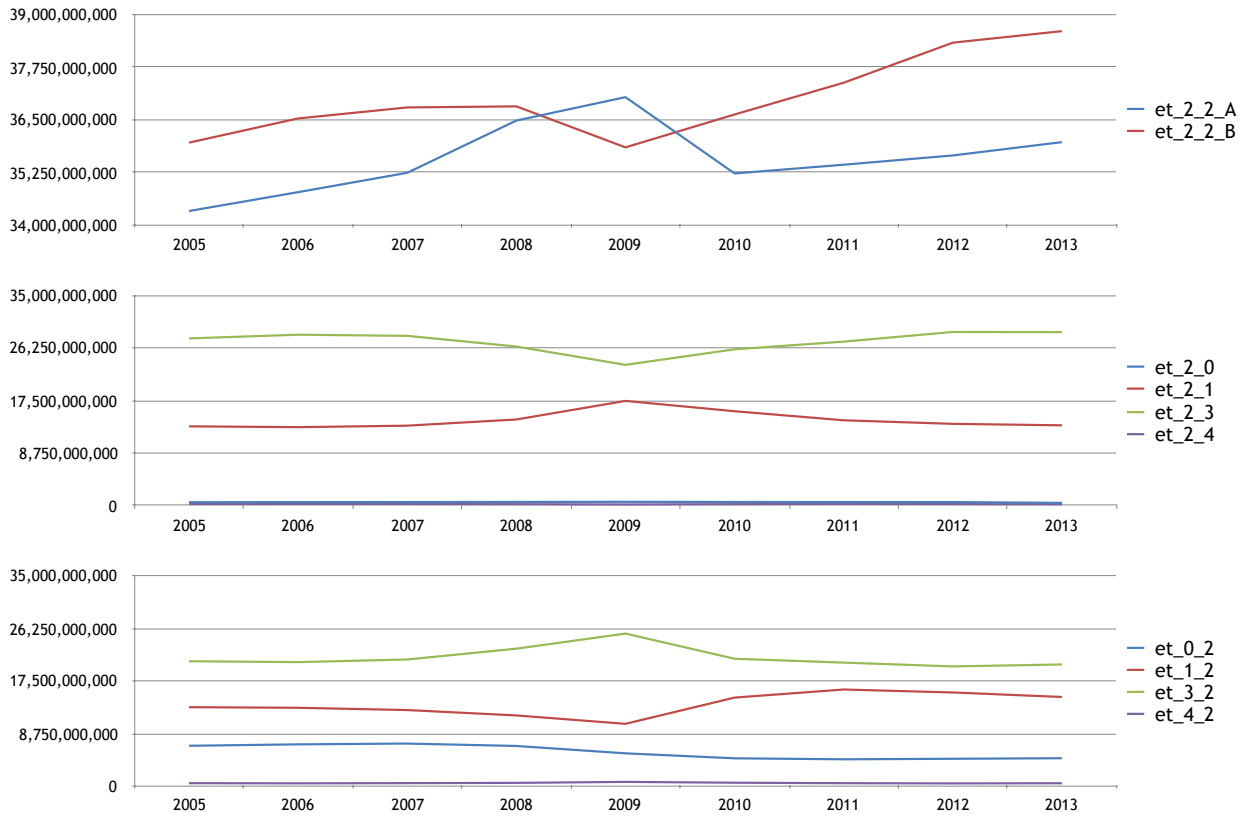
Table E.3: Average Earnings and Variability by Transition Type

	Flows from Bottom 20%					*Flows to Bottom 20%*				
	et_2_2_A	et_2_0	**et_2_1**	**et_2_3**	et_2_4	**et_2_2_B**	et_0_2	**et_1_2**	**et_3_2**	et_4_2
Mean	**2,706**	2,620	**2,053**	**3,631**	3,377	**2,814**	2,427	**2,202**	**3,474**	2,963
IQR	**22**	17	**30**	**48**	15	**24**	244	**14**	**20**	67
Minimum	**2,657**	2,600	**2,015**	**3,571**	3,362	**2,712**	2,267	**2,150**	**3,365**	2,804
Maximum	**2,802**	2,678	**2,087**	**3,747**	3,399	**2,873**	2,569	**2,227**	**3,518**	3,041
	Flows from Middle 60%					*Flows to Middle 60%*				
	et_3_3_A	et_3_0	**et_3_1**	**et_3_2**	**et_3_4**	**et_3_3_B**	et_0_3	**et_1_3**	**et_2_3**	**et_4_3**
Mean	**23,940**	18,680	**16,720**	**14,220**	**37,430**	**24,450**	11,672	**15,240**	**12,980**	**35,340**
IQR	**160**	442	**166**	**181**	**228**	**297**	1,685	**429**	**277**	**507**
Minimum	**23,540**	18,110	**16,560**	**13,850**	**36,910**	**24,220**	10,370	**14,950**	**12,540**	**34,260**
Maximum	**24,160**	19,391	**16,910**	**14,950**	**38,200**	**24,720**	12,290	**15,510**	**13,210**	**35,970**
	Flows from Top 20%					*Flows to Top 20%*				
	et_4_4_A	et_4_0	et_4_1	et_4_2	**et_4_3**	**et_4_4_B**	et_0_4	et_1_4	et_2_4	**et_3_4**
Mean	**94,160**	113,200	107,220	80,970	**60,900**	**96,080**	96,160	94,320	73,770	**57,510**
IQR	**1,922**	3,219	8,788	2,596	**1,117**	**2,086**	6,893	1,409	3,017	**152**
Minimum	**91,810**	107,440	97,650	78,800	**59,930**	**93,720**	78,540	91,320	69,450	**56,820**
Maximum	**96,100**	118,500	117,400	82,800	**61,790**	**98,010**	108,200	98,800	81,490	**57,780**

Notes: Dominant flows are in bold. The estimates are the weighted annual mean, inter-quartile Range (IQR), minimum, and maximum of the mean annual earnings in each category. Statistics are over nine pairs of years from 2004-2005 to 2012-2013.

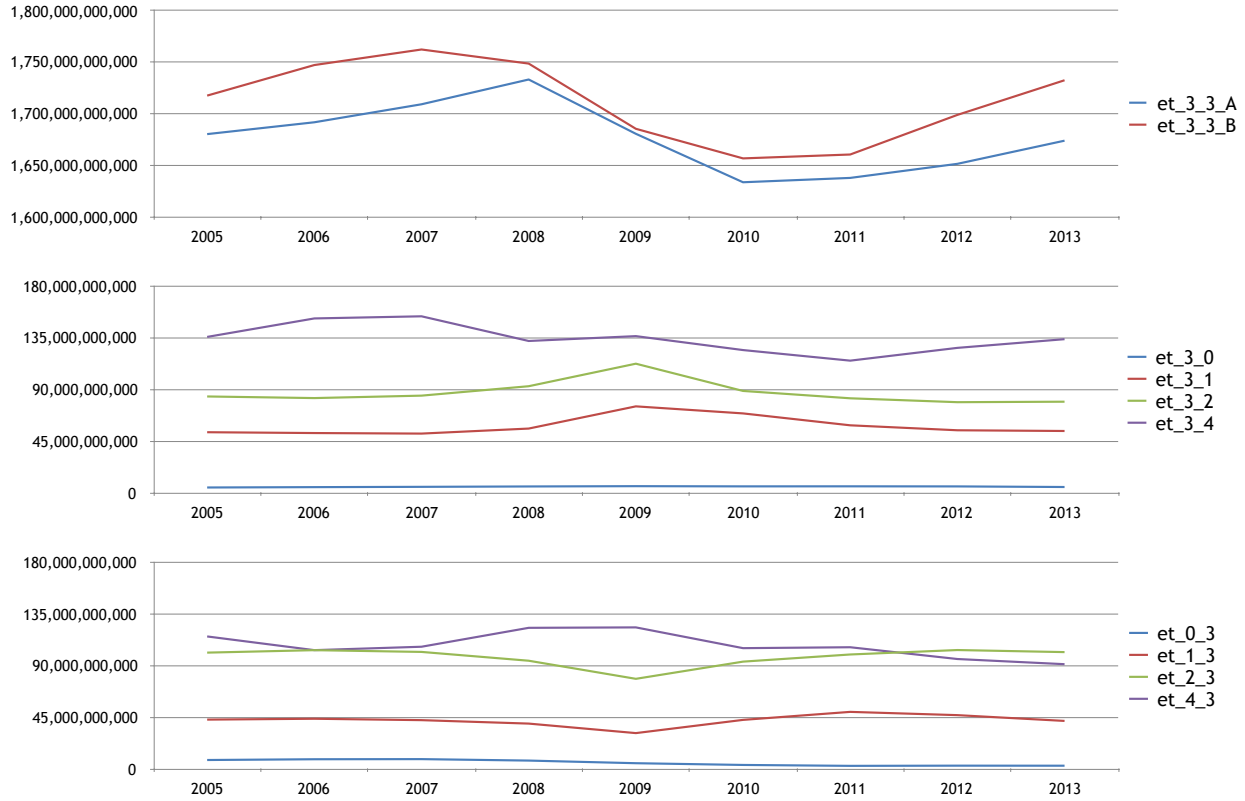
Appendix Figures E.1-fig:TE4 repeat the analysis shown in the main text in Figures 10-12.

Figure E.1: Earnings Flows out from and in to the Bottom 20%



Notes: The estimates are based on the authors' calculations using transitions into and out of categories of the earnings distribution, including inactivity.

Figure E.2: Earnings Flows out from and in to the Middle 60%



Notes: The estimates based on the authors' calculations using transitions into and out of categories of the earnings distribution, including inactivity.

Figure E.3: Earnings Flows out from and in to the Top 20%

Notes: The estimates are based on the authors' calculations using transitions into and out of categories of the earnings distribution, including inactivity.

E.2 AKM Decomposition

We estimate the following AKM model:

$$\ln y_{ijt} = x_{it}\beta + \theta_i + \psi j + \varepsilon_{ijt} \tag{E-5}$$

where y_{ijt} is log real annual earnings of person i, employed at firm j in year t; θ_i is individual i's person effect; ψ_j is firm j's fixed effect; and x_{it} includes controls for experience, labor force attachment, and aggregate labor market conditions detailed in Table E.4. Estimates of all these controls are in Table E.5.

Table E.4: AKM Model Specification

Actual Labor-Force Experience
$[exp, exp^2/10, exp^3/100, exp^4/1000]$
$\mathbb{1}\{\text{female}\} * [exp, exp^2/10, exp^3/100, exp^4/1000]$
$\mathbb{1}\{\text{black}\} * [exp, exp^2/10, exp^3/100, exp^4/1000]$
$\mathbb{1}\{\text{Hispanic}\} * [exp, exp^2/10, exp^3/100, exp^4/1000]$
$\mathbb{1}\{\text{foreign born}\} * [exp, exp^2/10, exp^3/100, exp^4/1000]$
$\mathbb{1}\{\text{foreign born}\} * \mathbb{1}\{\text{female}\} * [exp, exp^2/10, exp^3/100, exp^4/1000]$
$\mathbb{1}\{\text{foreign born}\} * \mathbb{1}\{\text{black}\} * [exp, exp^2/10, exp^3/100, exp^4/1000]$
$\mathbb{1}\{\text{foreign born}\} * \mathbb{1}\{\text{Hispanic}\} * [exp, exp^2/10, exp^3/100, exp^4/1000]$
Labor-Force Attachment
weeks by hours categories (41 total, 40 hours by 50-52 weeks excluded)
sixq dummies (9 total: sixq2-sixq6, sixq_4th, sixq_left, sixq_right, sixq_inter)
$\mathbb{1}\{\text{female}\} *$ [sixq dummies]
$\mathbb{1}\{\text{black}\} *$ [sixq dummies]
$\mathbb{1}\{\text{Hispanic}\} *$ [sixq dummies]
$\mathbb{1}\{\text{foreign born}\} *$ [sixq dummies]
$\mathbb{1}\{\text{foreign born}\} * \mathbb{1}\{\text{female}\} *$ [sixq dummies]
$\mathbb{1}\{\text{foreign born}\} * \mathbb{1}\{\text{black}\} *$ [sixq dummies]
$\mathbb{1}\{\text{foreign born}\} * \mathbb{1}\{\text{Hispanic}\} *$ [sixq dummies]
Aggregate Labor-Market Conditions
$[u_t, \mathbb{1}\{u_t > u_{t-1}\} * u_t]$
$\mathbb{1}\{\text{female}\} * [u_t, \mathbb{1}\{u_t > u_{t-1}\} * u_t]$
$\mathbb{1}\{\text{black}\} * [u_t, \mathbb{1}\{u_t > u_{t-1}\} * u_t]$
$\mathbb{1}\{\text{Hispanic}\} * [u_t, \mathbb{1}\{u_t > u_{t-1}\} * u_t]$
$\mathbb{1}\{\text{foreign born}\} * [u_t, \mathbb{1}\{u_t > u_{t-1}\} * u_t]$
$\mathbb{1}\{\text{foreign born}\} * \mathbb{1}\{\text{female}\} * [u_t, \mathbb{1}\{u_t > u_{t-1}\} * u_t]$
$\mathbb{1}\{\text{foreign born}\} * \mathbb{1}\{\text{black}\} * [u_t, \mathbb{1}\{u_t > u_{t-1}\} * u_t]$
$\mathbb{1}\{\text{foreign born}\} * \mathbb{1}\{\text{Hispanic}\} * [u_t, \mathbb{1}\{u_t > u_{t-1}\} * u_t]$
Incomplete 2014Q1 Data Controls
[right: indicator for incomplete data in 2014Q1 in one state and DC]
$\mathbb{1}\{\text{female}\} *$[right]
$\mathbb{1}\{\text{black}\} *$[right]
$\mathbb{1}\{\text{Hispanic}\} *$[right]
$\mathbb{1}\{\text{foreign born}\} *$[right]
$\mathbb{1}\{\text{foreign born}\} * \mathbb{1}\{\text{female}\} *$[right]
$\mathbb{1}\{\text{foreign born}\} * \mathbb{1}\{\text{black}\} *$[right]
$\mathbb{1}\{\text{foreign born}\} * \mathbb{1}\{\text{Hispanic}\} *$[right]

Notes: The specification also includes a fixed worker effect for each individual in the eligible-workers frame and a fixed firm effect for each employer in that frame. The AKM estimation occurs only during date regime 4, which is the complete population; however, our labor-force attachment variables require an additional quarter to calculate (2014Q1), which is missing for one state and DC. The "right" variable controls for the case where a sixq variable is set to zero due to data availability instead of actual labor-force attachment.

Table E.5: AKM Model Estimates

1	experience	0.0973	86	female_sixq5	0.0474
2	experience_2	-0.0379	87	female_sixq6	0.0545
3	experience_3	0.0070	88	female_sixq_4th	-0.0149
4	experience_4	-0.0006	89	female_sixq_left	-0.0028
5	female_experience	-0.0122	90	female_sixq_right	0.0333
6	female_experience_2	0.0044	91	female_sixq_inter	0.0265
7	female_experience_3	-0.0005	92	black_sixq2	0.0896
8	female_experience_4	0.0000	93	black_sixq3	0.1388
9	black_experience	-0.0470	94	black_sixq4	0.1584
10	black_experience_2	0.0211	95	black_sixq5	0.1445
11	black_experience_3	-0.0046	96	black_sixq6	0.2161
12	black_experience_4	0.0004	97	black_sixq_4th	0.0250
13	hispanic_experience	-0.0372	98	black_sixq_left	-0.1004
14	hispanic_experience_2	0.0201	99	black_sixq_right	-0.0761
15	hispanic_experience_3	-0.0049	100	black_sixq_inter	-0.0829
16	hispanic_experience_4	0.0005	101	hispanic_sixq2	0.0962
17	fbstat_experience	-0.0424	102	hispanic_sixq3	0.1280
18	fbstat_experience_2	0.0238	103	hispanic_sixq4	0.1386
19	fbstat_experience_3	-0.0058	104	hispanic_sixq5	0.1101
20	fbstat_experience_4	0.0005	105	hispanic_sixq6	0.1990
21	female_fbstat_experience	0.0007	106	hispanic_sixq_4th	0.0269
22	female_fbstat_experience_2	0.0008	107	hispanic_sixq_left	-0.0927
23	female_fbstat_experience_3	-0.0004	108	hispanic_sixq_right	-0.0933
24	female_fbstat_experience_4	0.0000	109	hispanic_sixq_inter	-0.1123
25	black_fbstat_experience	0.0179	110	fbstat_sixq2	-0.0108
26	black_fbstat_experience_2	-0.0086	111	fbstat_sixq3	-0.0361
27	black_fbstat_experience_3	0.0021	112	fbstat_sixq4	-0.0602
28	black_fbstat_experience_4	-0.0002	113	fbstat_sixq5	-0.1155
29	hispanic_fbstat_experience	0.0146	114	fbstat_sixq6	-0.1533
30	hispanic_fbstat_experience_2	-0.0140	115	fbstat_sixq_4th	0.0377
31	hispanic_fbstat_experience_3	0.0044	116	fbstat_sixq_left	0.0292
32	hispanic_fbstat_experience_4	-0.0005	117	fbstat_sixq_right	0.0079
33	WKSHRS1	-0.3017	118	fbstat_sixq_inter	0.0724
34	WKSHRS2	-0.2561	119	female_fbstat_sixq2	-0.0245
35	WKSHRS3	-0.2044	120	female_fbstat_sixq3	-0.0390
36	WKSHRS4	-0.1260	121	female_fbstat_sixq4	-0.0308
37	WKSHRS5	-0.0625	122	female_fbstat_sixq5	-0.0262
38	WKSHRS6	0.0782	123	female_fbstat_sixq6	-0.0374
39	WKSHRS7	0.1381	124	female_fbstat_sixq_4th	0.0067
40	WKSHRS8	-0.2907	125	female_fbstat_sixq_left	0.0100
41	WKSHRS9	-0.1951	126	female_fbstat_sixq_right	-0.0028
42	WKSHRS10	-0.1122	127	female_fbstat_sixq_inter	0.0041
43	WKSHRS11	-0.0100	128	black_fbstat_sixq2	0.0007
44	WKSHRS12	0.0831	129	black_fbstat_sixq3	0.0243
45	WKSHRS13	0.1570	130	black_fbstat_sixq4	0.0403
46	WKSHRS14	0.1734	131	black_fbstat_sixq5	0.0770
47	WKSHRS15	-0.3176	132	black_fbstat_sixq6	0.0787
48	WKSHRS16	-0.1633	133	black_fbstat_sixq_4th	-0.0270
49	WKSHRS17	-0.0929	134	black_fbstat_sixq_left	0.0341
50	WKSHRS18	-0.0090	135	black_fbstat_sixq_right	0.0483
51	WKSHRS19	0.0628	136	black_fbstat_sixq_inter	0.0437
52	WKSHRS20	0.1167	137	hispanic_fbstat_sixq2	0.0099
53	WKSHRS21	0.1404	138	hispanic_fbstat_sixq3	-0.0025
54	WKSHRS22	-0.3661	139	hispanic_fbstat_sixq4	0.0027
55	WKSHRS23	-0.2028	140	hispanic_fbstat_sixq5	0.0529
56	WKSHRS24	-0.1196	141	hispanic_fbstat_sixq6	0.0141
57	WKSHRS25	-0.0685	142	hispanic_fbstat_sixq_4th	-0.0252
58	WKSHRS26	-0.0223	143	hispanic_fbstat_sixq_left	0.0414
59	WKSHRS27	0.0011	144	hispanic_fbstat_sixq_right	0.0278
60	WKSHRS28	0.0161	145	hispanic_fbstat_sixq_inter	0.0434
61	WKSHRS29	-0.3451	146	urate	-0.0095
62	WKSHRS30	-0.1839	147	urate_up	0.0017
63	WKSHRS31	-0.0999	148	female_urate	0.0034
64	WKSHRS32	-0.0550	149	female_urate_up	0.0006
65	WKSHRS33	-0.0145	150	black_urate	0.0045
66	WKSHRS34	0.0028	151	black_urate_up	-0.0001
67	WKSHRS35	0.0183	152	hispanic_urate	0.0015
68	WKSHRS36	-0.3237	153	hispanic_urate_up	0.0005
69	WKSHRS37	-0.1716	154	fbstat_urate	-0.0000
70	WKSHRS38	-0.0929	155	fbstat_urate_up	0.0003
71	WKSHRS39	-0.0361	156	female_fbstat_urate	-0.0004
72	WKSHRS41	0.0223	157	female_fbstat_urate_up	-0.0001
73	WKSHRS42	0.0320	158	black_fbstat_urate	-0.0059
74	sixq2	1.1170	159	black_fbstat_urate_up	0.0009
75	sixq3	2.2170	160	hispanic_fbstat_urate	-0.0032
76	sixq4	2.7750	161	hispanic_fbstat_urate_up	-0.0003
77	sixq5	3.2910	162	right	0.2083
78	sixq6	3.6920	163	female_right	0.0319
79	sixq_4th	0.0323	164	black_right	-0.0181
80	sixq_left	-0.2940	165	hispanic_right	-0.0051
81	sixq_right	-0.1401	166	fbstat_right	0.0060
82	sixq_inter	-0.7029	167	female_fbstat_right	-0.0273
83	female_sixq2	0.0250	168	black_fbstat_right	0.0545
84	female_sixq3	0.0563	169	hispanic_fbstat_right	-0.0139
85	female_sixq4	0.0544			

Notes: The table presents the coefficient estimates of all the controls listed in Table E.4. $N = 2,014,000,000$, $Jobs = 825,900,000$, $Persons = 200,700,000$, $Firms = 14,650,000$. $Intercept = 6.098$, calculated after estimation. The equation includes one person effect for each person and firm effects for all firms, save one. Estimation and identification performed as described in Abowd et al. (2002). All observations in the complete frame, which has universal coverage over the period 2004-2013, were used. Finite population standard errors are zero. The estimates and their associated standard errors have not been corrected for edit, imputation, and post-processing uncertainty.

E.3 Analyzing Earnings Inequality Changes Using Only Firm-Type and Non-Firm-Type

In the main text Section 5, we use the AKM decomposition to create firm, non-firm, and skill components of earnings. These components are used to create firm-type, non-firm-type and skill-type bins that we subsequently employ to characterize the worker and firm contributions to changes in earnings inequality.

An earlier version of this paper used the non-firm-type bins in a manner similar to the use of the skill-type bins in the main text. We discuss these results here for each non-firm-type separately. We remind the reader that the non-firm-component contains the effects of changes in the labor-force attachment, macroeconomic conditions, date regime boundaries, and the residual, all of which are excluded from the skill-type in the main text.

Table E.6 presents outcomes for workers in the bottom bin of the non-firm component. Table E.7 presents outcomes for workers in the middle bin of the non-firm component distribution. Table E.8 presents outcomes for workers in the top bin of the non-firm component distribution.

The tables were created as follows. They are based on classifying workers in the previous year, i.e., year $t-1$. Beginning in 2004 and ending in 2012, for every year that an eligible worker has positive earnings a single observation is added to one of the three tables. The appropriate table classification for each observation is determined by the non-firm type for that year, which can vary over time as workers accumulate experience, work more/less hours during the quarter, receive a positive or negative aggregate demand shock, or have a large positive or negative residual. Within each non-firm type, the earnings record is further classified based on the firm type, resulting in each earnings observation being classified into one of nine possible cells.[58] Within each of the non-firm-type $\times$ firm-type cells, we break down the results by the three possible overall-earnings outcomes (bottom, middle and top). There are, thus, twenty-seven cells for which we present information on the number of workers, average earnings for the previous year $(t-1)$, and average earnings for the current year (t) by flow type.[59]

To fix ideas, we will take a detailed look at two rows in Table E.6. To be recorded in this table, the person must have been in the bottom bin (lowest bin) of the non-firm-component distribution in the "Previous Year," i.e., $t-1$.

Consider the first row of the table. This row is in the panel labeled "Bottom Firm," indicating that this person is employed at a firm in the bottom bin of the firm component distribution in $t-1$. Persons in this row are also in the bottom bin of the overall-earnings distribution in year $t-1$, and the share of such persons (relative to those in the middle or top of the overall-earnings distribution) is 1.000, indicating that no person in the bottom of the non-firm component distribution and the bottom of the firm component distribution is employed outside of the bottom bin of the overall-earnings distribution. The flow labeled "2_0" is the movement from the bottom of the overall-earnings distribution (bin 2) to ineligible; that is, this is the flow out of the frame for persons at the bottom of the overall-earnings distribution. There were, on average, 59,554 such persons each "previous year" $(t-1)$. They represent 0.7% of the flows from bin 2 of the overall-earnings distribution. Average earnings in $t-1$ were \$1,381 of which $-\$1,463$ are attributed to the firm component of our decomposition and \$2,844 are attributed to the non-firm component of our decomposition. There were no earnings in the current year (t), because the person has moved out of the frame in t.

Next, consider the row labeled "Middle" in the "All Earnings" column in the "Middle Firm" panel with a "3_3" flow. All persons in this row were, once again, at the bottom of the non-

[58] The estimated AKM firm effects do not vary during the period, but workers can and do change employers.

[59] The earnings observation we used for classification are labeled "previous year" in the tables.

firm component distribution in year $t-1$. Of all such persons, 56% are employed by a firm in the middle of the firm component distribution. Of all persons at the bottom of the non-firm component distribution and in the middle of the firm component distribution in year $t-1$, the proportion 0.159 were in the middle of the overall-earnings distribution. Among such persons, the "3_3" row shows those who remain in the middle of the overall-earnings distribution in the current year, t, of which there were, on average, 1,470,659 in the 9 pairs of years for which the table was constructed. Those who stayed in the middle of the overall-earnings distribution represented 58.9% of all persons who were in the middle of the overall-earnings distribution in year $t-1$, on average. In year $t-1$, their earnings averaged \$8,498 of which \$2,180 is attributed to the firm component in our decomposition and \$6,318 is attributed to the non-firm component. In the current year, year t, average earnings were \$15,688 of which \$3,555 is associated with the firm component and \$12,132 is associated with the non-firm component.

We use these tables to investigate worker sorting directly by looking at the interaction of the non-firm and firm type for each worker-year-earnings observation. If there were no sorting, the distribution of earnings observations across firm types would be similar for all three tables, because outcomes would be unaffected by which part of the non-firm component distribution an individual occupied, given his place in the overall-earnings distribution. This hypothesis is clearly not supported by the data. For example, again using Table E.6 showing the bottom of the non-firm-type distribution, about 33% of the earnings observations are in firms at the bottom of the firm-type distribution, 56% are in firms of the middle type, and only 11% are in top firms. In comparison, Tables E.7 and E.8 show that persons in the middle and top of the non-firm type distributions are much less likely to be employed at firms in the bottom type (14% and 24% respectively), and much more likely to be employed at top firms (23% and 20% respectively). Interestingly, the relationship is not monotonic; workers in the middle are more likely to work at both middle and top firms relative to top workers.

Next, we focus on each non-firm-type in turn, starting with the earnings observations for workers in the bottom of the non-firm component distribution in Table E.6. For workers at the bottom of the non-firm-component distribution, working at a high-paying firm has two advantages: higher earnings than otherwise and a greater chance of moving to a higher bin in the overall-earnings distribution. For example, a worker at the bottom of the non-firm-component and overall-earnings distributions has a probability of moving to the middle of the overall-earnings distribution of 18% at a low paying firm, 29.5% at a middle paying firm, and 27.5% at a high paying firm. Prior to the transition the average worker with a low non-firm component at a low-, middle- and high-paying firm earns \$2,084, \$3,556, and \$3,806, respectively.[60] After the transition the average worker at a low-, middle- and high-paying firm earns \$11,640, \$13,752, and \$18,017, respectively. Most of the additional increase in earnings for workers employed at a top-paying employer in the previous year is due to working at a top-paying employer in the next year.

[60] Notice that the non-firm component of earnings declines as we move up the firm type distribution. Although it is unclear exactly which covariate is primarily responsible for this decline (fewer hours worked during the year perhaps), the impact of working at a higher paying firm would be much greater if the non-firm component of earnings were the same across firm types.

Table E.6: Earnings Associated with Flows by Firm Bin for Persons in the Bottom-Type Non-Firm Category)

All Earnings	Share	Flow	Average Count	Pct	Previous Year			Current Year		
					Total	*Firm*	*Non-Firm*	*Total*	*Firm*	*Non-Firm*
Bottom Firm (33%)										
Bottom	1.000	2_0	59,554	0.7%	1,381	-1,463	2,844	—	—	—
		2_1	2,441,375	26.8%	1,102	-1,099	2,201	—	—	—
		2_2	4,962,828	54.5%	1,635	-1,588	3,223	2,466	-1,981	4,447
		2_3	1,641,446	18.0%	2,084	-1,738	3,823	11,640	-4,017	15,657
		2_4	8,640	0.1%	1,513	-1,558	3,071	78,157	8,958	69,199
Middle	0.000	3_0	0	0.0%	—	—	—	—	—	—
		3_1	0	0.0%	—	—	—	—	—	—
		3_2	0	0.0%	—	—	—	—	—	—
		3_3	0	0.0%	—	—	—	—	—	—
		3_4	0	0.0%	—	—	—	—	—	—
Top	0.000	4_0	0	0.0%	—	—	—	—	—	—
		4_1	0	0.0%	—	—	—	—	—	—
		4_2	0	0.0%	—	—	—	—	—	—
		4_3	0	0.0%	—	—	—	—	—	—
		4_4	0	0.0%	—	—	—	—	—	—
Middle Firm (56%)										
Bottom	0.841	2_0	116,724	0.9%	2,660	72	2,588	—	—	—
		2_1	3,613,606	27.3%	2,289	42	2,247	—	—	—
		2_2	5,565,538	42.0%	2,784	-145	2,929	2,799	-594	3,392
		2_3	3,911,555	29.5%	3,556	-27	3,583	13,752	561	13,191
		2_4	36,073	0.3%	3,392	469	2,923	69,402	19,672	49,730
Middle	0.159	3_0	21,191	0.8%	8,381	2,189	6,193	—	—	—
		3_1	428,729	17.2%	8,384	2,249	6,135	—	—	—
		3_2	554,068	22.2%	8,153	1,893	6,260	3,321	-14	3,336
		3_3	1,470,659	58.9%	8,498	2,180	6,318	15,688	3,555	12,132
		3_4	21,549	0.9%	8,955	2,823	6,132	64,566	22,919	41,647
Top	0.000	4_0	0	0.0%	—	—	—	—	—	—
		4_1	0	0.0%	—	—	—	—	—	—
		4_2	0	0.0%	—	—	—	—	—	—
		4_3	0	0.0%	—	—	—	—	—	—
		4_4	0	0.0%	—	—	—	—	—	—
Top Firm (11%)										
Bottom	0.396	2_0	17,420	1.4%	2,913	1,598	1,314	—	—	—
		2_1	469,324	38.3%	2,758	1,515	1,243	—	—	—
		2_2	377,303	30.8%	3,174	1,740	1,433	2,905	806	2,099
		2_3	337,787	27.5%	3,806	2,034	1,771	18,017	7,482	10,535
		2_4	24,607	2.0%	3,701	2,058	1,642	76,278	41,316	34,962
Middle	0.596	3_0	16,910	0.9%	12,121	7,299	4,822	—	—	—
		3_1	375,155	20.3%	12,082	7,280	4,802	—	—	—
		3_2	243,668	13.2%	11,573	6,756	4,817	3,134	774	2,360
		3_3	1,108,356	60.0%	13,551	8,484	5,067	20,030	10,749	9,281
		3_4	102,240	5.5%	15,785	10,529	5,255	70,117	42,636	27,481
Top	0.008	4_0	172	0.7%	97,790	93,786	4,005	—	—	—
		4_1	1,924	7.6%	96,408	92,599	3,809	—	—	—
		4_2	498	2.0%	93,690	89,588	4,103	2,376	-53	2,430
		4_3	4,217	16.7%	65,091	60,669	4,422	32,200	28,313	3,887
		4_4	18,478	73.1%	108,698	104,482	4,216	117,522	110,839	6,683

Notes: The estimates are based on the authors' calculations using the nine paired years from 2004-2005 to 2012-2013. The first year in the pair is the "previous year" in the table, and the second year in the pair is the "current year." Bins associated with the flows are "0" inflow/outflow from the eligible-workers frame, "1" inactive but eligible, "2" bottom of the overall-earnings distribution, "3" middle of the overall-earnings distribution, and "4" top of the overall-earnings distribution. "Average count" is the average number of persons in the row during the year labeled "previous year" $(t-1)$. Pct is the percent distribution of transitions for all persons who started the year in the same overall-earnings distribution bin. For "Previous Year" and "Current Year," "Total" is the average real earnings in 2000 dollars, "Firm" is the average real earnings associated with the firm component in our decomposition, and "Non-Firm" is the average real earnings associated with the non-firm component in our decomposition.

The vast majority (63%) of workers in the middle of the non-firm component distribution are employed at middle-paying firms, as Table E.7 shows. The next most prevalent outcomes for such workers are employment at top- and bottom-paying firms, 23% and 14% respectively. Similar to workers at the bottom of the non-firm type distribution, who also generally appear at the bottom of the overall-earnings distribution (84%) when employed by middle-paying firms, the majority of workers in the middle of the non-firm type distribution, no matter the firm type, are in the middle of the overall-earnings distribution. However, in spite of the majority of earnings observations being in the middle of the overall-earnings distribution, average earnings differ substantially across firm types. A middle-type worker in bin 3 of the overall-earnings distribution who stays in bin 3 of the overall distribution (a "3_3" flow) at a bottom-type firm has $t-1$ earnings of \$12,356, a middle-type worker in a middle-type firm has $t-1$ earnings of \$22,978, and a middle-type worker at a top firm has earnings of \$32,321. Most of the difference is due to a larger firm effect, although the non-firm component declines somewhat as a middle-type person is found in increasing firm types, giving back some of the gains. Similar to bottom-type workers, one of the additional benefits of finding employment at a high-paying firm is a greater probability of moving to the top of the earnings distribution (0.2% vs. 2.7% vs. 11.9% in rows 5, 25, and 40, respectively).

Similar to bottom and middle non-firm-type workers, Table E.8 shows that about 64% of top non-firm type workers are also in the top of the overall-earnings distribution, but there is also a substantial minority in the middle. The differences between working at a middle- compared to a bottom-type firm are relatively small, but the gains from working at a top-type firm are very large. Somewhat surprisingly perhaps, there are a relatively large number of top-type workers at bottom- and middle-type firms. On average, these workers, especially in the middle, are employed at worse-paying firms than middle non-firm type workers.

Table E.7: Earnings Associated with Flows by Firm Bin for Persons in the Middle-Type Non-Firm Category

All Earnings	Share	Flow	Average Count	Pct	Previous Year			Current Year		
					Total	*Firm*	*Non-Firm*	*Total*	*Firm*	*Non-Firm*
Bottom Firm (14%)										
		2_0	26,241	0.7%	4,596	-7,583	12,179	—	—	—
		2_1	352,516	9.4%	4,627	-6,141	10,768	—	—	—
Bottom	0.313	2_2	2,005,303	53.6%	4,676	-6,627	11,303	3,590	-4,861	8,452
		2_3	1,352,780	36.2%	5,199	-5,937	11,137	11,278	-7,815	19,093
		2_4	4,255	0.1%	4,649	-7,524	12,173	79,652	-16,894	96,546
		3_0	39,797	0.5%	11,198	-9,855	21,053	—	—	—
		3_1	312,400	3.8%	10,551	-8,857	19,408	—	—	—
Middle	0.687	3_2	1,331,161	16.2%	9,726	-8,550	18,275	3,798	-3,326	7,124
		3_3	6,493,717	79.2%	12,356	-9,762	22,118	14,200	-9,400	23,600
		3_4	18,706	0.2%	14,081	-10,636	24,717	70,391	-30,622	101,013
		4_0	0	0.0%	—	—	—	—	—	—
		4_1	0	0.0%	—	—	—	—	—	—
Top	0.000	4_2	0	0.0%	—	—	—	—	—	—
		4_3	0	0.0%	—	—	—	—	—	—
		4_4	0	0.0%	—	—	—	—	—	—
Middle Firm (63%)										
		2_0	3,160	0.6%	6,108	-2,046	8,154	—	—	—
		2_1	56,529	10.7%	6,093	-2,039	8,132	—	—	—
Bottom	0.010	2_2	211,504	39.9%	6,079	-2,081	8,160	3,753	-1,489	5,242
		2_3	257,664	48.7%	6,122	-2,062	8,185	12,008	-2,577	14,585
		2_4	730	0.1%	6,121	-2,031	8,152	68,020	5,440	62,580
		3_0	170,775	0.3%	18,829	1,971	16,858	—	—	—
		3_1	1,789,911	3.6%	16,909	2,210	14,699	—	—	—
Middle	0.958	3_2	3,467,732	6.9%	15,078	884	14,194	3,439	-520	3,958
		3_3	43,259,502	86.4%	22,978	3,475	19,503	23,517	3,506	20,012
		3_4	1,370,036	2.7%	35,902	10,130	25,772	57,122	16,550	40,572
		4_0	2,532	0.2%	51,159	18,825	32,335	—	—	—
		4_1	17,159	1.0%	51,191	19,006	32,185	—	—	—
Top	0.031	4_2	13,212	0.8%	50,902	18,745	32,156	3,202	569	2,632
		4_3	437,317	26.6%	49,933	17,999	31,934	37,792	13,081	24,711
		4_4	1,174,019	71.4%	51,694	19,249	32,445	55,357	20,583	34,775
Top Firm (23%)										
		2_0	0	0.0%	—	—	—	—	—	—
		2_1	0	0.0%	—	—	—	—	—	—
Bottom	0.000	2_2	0	0.0%	—	—	—	—	—	—
		2_3	0	0.0%	—	—	—	—	—	—
		2_4	0	0.0%	—	—	—	—	—	—
		3_0	30,130	0.3%	29,438	15,459	13,980	—	—	—
		3_1	445,548	4.0%	27,758	14,834	12,923	—	—	—
Middle	0.569	3_2	343,349	3.1%	27,352	14,186	13,166	3,111	789	2,322
		3_3	8,891,952	80.7%	32,321	16,549	15,772	31,657	15,654	16,003
		3_4	1,306,028	11.9%	38,938	20,918	18,021	58,297	31,112	27,185
		4_0	12,388	0.1%	64,410	38,198	26,213	—	—	—
		4_1	129,141	1.5%	64,268	39,009	25,258	—	—	—
Top	0.431	4_2	69,540	0.8%	61,384	35,782	25,602	2,939	974	1,965
		4_3	1,055,443	12.6%	56,162	31,142	25,020	34,895	17,929	16,965
		4_4	7,085,455	84.8%	64,675	37,688	26,987	68,632	39,649	28,983

Notes: The estimates are based on the authors' calculations using the nine paired years from 2004-2005 to 2012-2013. The first year in the pair is the "previous year" in the table, and the second year in the pair is the "current year." Bins associated with the flows are "0" inflow/outflow from the eligible-workers frame, "1" inactive but eligible, "2" bottom of the overall-earnings distribution, "3" middle of the overall-earnings distribution, and "4" top of the overall-earnings distribution. "Average count" is the average number of persons in the row during the year labeled "previous year" $(t-1)$. Pct is the percent distribution of transitions for all persons who started the year in the same overall-earnings distribution bin. For "Previous Year" and "Current Year," "Total" is the average real earnings in 2000 dollars, "Firm" is the average real earnings associated with the firm component in our decomposition, and "Non-Firm" is the average real earnings associated with the non-firm component in our decomposition.

Table E.8: Earnings Associated with Flows by Firm Bin for Persons in the High-Type Non-Firm Category

All Earnings	Share	Flow	Average Count	Pct	Previous Year			Current Year		
					Total	*Firm*	*Non-Firm*	*Total*	*Firm*	*Non-Firm*
Bottom Firm (24%)										
		2_0	679	2.0%	4,353	-62,991	67,344	—	—	—
		2_1	2,485	7.2%	4,316	-64,774	69,090	—	—	—
Bottom	0.005	2_2	23,108	67.2%	4,275	-64,476	68,750	3,484	-51,189	54,673
		2_3	7,784	22.7%	5,099	-60,143	65,241	12,205	-94,469	106,674
		2_4	307	0.9%	3,928	-91,664	95,592	193,962	-307,444	501,406
		3_0	16,791	0.3%	24,421	-41,367	65,788	—	—	—
		3_1	93,523	1.8%	25,157	-38,303	63,460	—	—	—
Middle	0.768	3_2	162,321	3.2%	21,323	-37,554	58,877	3,476	-8,143	11,619
		3_3	4,657,816	90.6%	28,233	-34,400	62,633	27,573	-31,979	59,552
		3_4	211,635	4.1%	39,258	-46,667	85,925	56,165	-60,567	116,732
		4_0	4,499	0.3%	92,980	-133,477	226,457	—	—	—
		4_1	15,036	1.0%	93,061	-171,801	264,862	—	—	—
Top	0.227	4_2	9,282	0.6%	80,095	-133,162	213,257	3,008	-6,393	9,401
		4_3	192,250	12.6%	58,951	-75,598	134,549	36,197	-41,533	77,730
		4_4	1,299,404	85.5%	79,502	-109,491	188,992	80,704	-107,069	187,773
Middle Firm (56%)										
		2_0	0	0.0%	—	—	—	—	—	—
		2_1	0	0.0%	—	—	—	—	—	—
Bottom	0.000	2_2	0	0.0%	—	—	—	—	—	—
		2_3	0	0.0%	—	—	—	—	—	—
		2_4	0	0.0%	—	—	—	—	—	—
		3_0	9,579	0.2%	37,680	-5,362	43,042	—	—	—
		3_1	58,065	1.2%	37,365	-5,593	42,957	—	—	—
Middle	0.310	3_2	61,221	1.3%	36,524	-5,940	42,464	3,195	-1,035	4,230
		3_3	4,173,530	85.7%	37,519	-5,487	43,005	35,402	-5,028	40,430
		3_4	570,086	11.7%	42,684	-2,446	45,130	54,161	-1,779	55,940
		4_0	25,565	0.2%	111,359	17,448	93,911	—	—	—
		4_1	103,830	1.0%	96,076	16,375	79,701	—	—	—
Top	0.690	4_2	60,924	0.6%	76,673	11,055	65,618	3,037	-346	3,383
		4_3	1,192,613	11.0%	61,122	7,031	54,091	35,535	2,971	32,564
		4_4	9,463,943	87.3%	88,559	15,037	73,522	89,570	15,434	74,136
Top Firm (20%)										
		2_0	0	0.0%	—	—	—	—	—	—
		2_1	0	0.0%	—	—	—	—	—	—
Bottom	0.000	2_2	0	0.0%	—	—	—	—	—	—
		2_3	0	0.0%	—	—	—	—	—	—
		2_4	0	0.0%	—	—	—	—	—	—
		3_0	0	0.0%	—	—	—	—	—	—
		3_1	0	0.0%	—	—	—	—	—	—
Middle	0.000	3_2	0	0.0%	—	—	—	—	—	—
		3_3	0	0.0%	—	—	—	—	—	—
		3_4	0	0.0%	—	—	—	—	—	—
		4_0	9,962	0.2%	203,735	115,510	88,225	—	—	—
		4_1	73,693	1.4%	214,392	127,870	86,521	—	—	—
Top	1.000	4_2	27,036	0.5%	155,772	88,384	67,388	2,733	610	2,123
		4_3	163,477	3.0%	121,408	66,733	54,675	29,337	13,458	15,879
		4_4	5,145,974	94.9%	158,370	90,525	67,845	158,948	90,228	68,720

Notes: The estimates are based on the authors' calculations using the nine paired years from 2004-2005 to 2012-2013. The first year in the pair is the "previous year" in the table, and the second year in the pair is the "current year." Bins associated with the flows are "0" inflow/outflow from the eligible-workers frame, "1" inactive but eligible, "2" bottom of the overall-earnings distribution, "3" middle of the overall-earnings distribution, and "4" top of the overall-earnings distribution. "Average count" is the average number of persons in the row during the year labeled "previous year" $(t-1)$. Pct is the percent distribution of transitions for all persons who started the year in the same overall-earnings distribution bin. For "Previous Year" and "Current Year," "Total" is the average real earnings in 2000 dollars, "Firm" is the average real earnings associated with the firm component in our decomposition, and "Non-Firm" is the average real earnings associated with the non-firm component in our decomposition.