NBER WORKING PAPER SERIES

EARNINGS INEQUALITY AND MOBILITY TRENDS IN THE UNITED STATES: NATIONALLY REPRESENTATIVE ESTIMATES FROM LONGITUDINALLY LINKED EMPLOYER-EMPLOYEE DATA

John M. Abowd
Kevin L. McKinney
Nellie L. Zhao

Working Paper 23224
http://www.nber.org/papers/w23224

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2017

Abowd acknowledges direct support from NSF Grants SES-0339191, CNS-0627680, SES-0922005, TC-1012593, and SES-1131848. This paper was written while the the third author was a Pathways Intern at the U.S. Census Bureau. We have benefitted from discussions with David Card, John Eltinge, Patrick Kline, Francis Kramarz, Kristin McCue, Ian Schmutte, Lars Vilhuber, participants at the NBER conference that preceded this volume, the editors of this volume, Edward Lazear and Kathryn Shaw, and two anonymous referees. Sara Sullivan edited the final manuscript. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau, other sponsors, or the National Bureau of Economic Research. All results have been reviewed to ensure that no confidential information is disclosed. This research uses data from the Census Bureau's Longitudinal Employer-Household Dynamics Program, which was partially supported by the following National Science Foundation Grants: SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation. An archive of the code used to prepare analysis samples and conduct estimation along with the released results is available directly from http://digitalcommons.ilr.cornell.edu/ldi/34.

Earnings Inequality and Mobility Trends in the United States: Nationally Representative Estimates from Longitudinally Linked Employer-Employee Data
John M. Abowd, Kevin L. McKinney, and Nellie L. Zhao
NBER Working Paper No. 23224
March 2017
JEL No. J3,J6,M5

ABSTRACT

Using earnings data from the U.S. Census Bureau, this paper analyzes the role of the employer in explaining the rise in earnings inequality in the United States. We first establish a consistent frame of analysis appropriate for administrative data used to study earnings inequality. We show that the trends in earnings inequality in the administrative data from the Longitudinal Employer-Household Dynamics Program are inconsistent with other data sources when we do not correct for the presence of misused SSNs. After this correction to the worker frame, we analyze how the earnings distribution has changed in the last decade. We present a decomposition of the year-to-year changes in the earnings distribution from 2004-2013. Even when simplifying these flows to movements between the bottom 20%, the middle 60%, and the top 20% of the earnings distribution, about 20.5 million workers undergo a transition each year. Another 19.9 million move between employment and non-employment. To understand the role of the firm in these transitions, we estimate a model for log earnings with additive fixed worker and firm effects using all jobs held by eligible workers from 2004-2013. We construct a composite log earnings firm component across all jobs for a worker in a given year and a non-firm component. We also construct a skill-type index. We show that, while the difference between working at a low-or middle-paying firm are relatively small, the gains from working at a top-paying firm are large. Specifically, the benefits of working for a high-paying firm are not only realized today, through higher earnings paid to the worker, but also persist through an increase in the probability of upward mobility. High-paying firms facilitate moving workers to the top of the earnings distribution and keeping them there.

John M. Abowd
U.S. Census Bureau
4600 Silver Hill Road
Office 8H120
Washington, DC 20233
and Cornell University
john.maron.abowd@census.gov

Kevin L. McKinney
California Census Research Data Center
University of California, Los Angeles
4250 Public Policy Building
Los Angeles, CA 90095
kevinm@ccrdc.ucla.edu

Nellie L. Zhao
Cornell University
nlz6@cornell.edu

A curated digital commons paper and code repository is available at
http://digitalcommons.ilr.cornell.edu/ldi/34

1 Introduction

A growing body of work studies the rise in earnings inequality in the United States[1] with many studies focus on assessing the role of the employer in explaining these trends.[2] Virtually all of these papers use administrative data to analyze employer effects on earnings inequality. Information from administrative sources, unlike data collected in household surveys, is found or organic data.[3] That is, these data come from a convenient frame neither designed nor ensured to be representative of the population under study. Using found data to study features of a population, in particular features that evolve over time like earnings inequality, requires additional effort to determine what types of individuals are included in the found data and how those who are excluded from the analysis affect the results.

In this paper, we analyze the importance of firms in explaining the evolution of earnings inequality in the U.S. using administrative data from the Longitudinal Employer-Household Dynamics (LEHD) infrastructure files of the U.S. Census Bureau. We do two important analyses before turning to earnings inequality. First, unlike other studies of earnings inequality that use administrative data, we supplement the LEHD frame with data from the Social Security Administration (SSA) to create a consistent frame of workers over time. Specifically, we identify records in the administrative LEHD data that are either (a) associated with an invalid SSN or (b) used in a manner indicating possibly fraudulent labor market activity associated with a valid SSN. We remove both of these types of records from our analysis and document the consequences. Second, we systematically compare the cross-sectional estimates of earnings inequality in our administrative data to estimates based on conventional household surveys and document the differences. Then, we link the administrative and household survey data and use the linked data to understand where important features of the earnings distribution diverge.

By converting the found LEHD administrative data frame on workers to a designed frame, we can produce reliable estimates of features of the worker population. Conversion to a designed frame matters because trends in inequality differ between the designed frame and the original found frame. While our designed frame shows a widening of the earnings distribution post-2000 that is consistent with findings from other data sources, the found frame does not. Instead, the found frame shows no growth in earnings inequality since 2000.[4] We show that the workers excluded from the found frame to create the designed frame tend to be very low earnings individuals. In general, it is

[1] See Katz and Autor (1999) for a summary of the extensive body of work that analyzed the rise in U.S. wage inequality from the late 1970s to the mid-1990s and the forces behind this change in the wage structure.

[2] In particular, see Card et al. (2013) and Song et al. (2015). Aside from the increase in firm-specific wage premiums as a source of earnings inequality, other analyses have focused on skill-biased technological change (Acemoglu and Autor, 2011) and the rise in the returns to education (Autor, 2014).

[3] The term "found data" is commonly used in the data science literature, but the first formal use is Whither Biometrics Committee (2010, p. 72). Former U.S. Census Director Robert Groves coined the term "organic data"(Groves, 2010).

[4] Blackburn and Bloom (1987) made a related point when analyzing earnings inequality from 1967-1985 in CPS-ASEC. They noted that the patterns in earnings inequality observed in the data depended on various factors, including which individuals were included in the earnings distribution. More recently, Spletzer (2014) noticed a similar difference when comparing trends in inequality observed in the CPS vs. the LEHD data.

the bottom of the earnings distribution that is most affected by different frame-selection criteria.

Using our designed worker frame, we analyze how the earnings distribution has changed over time. We include inactive workers in our analysis, defined as individuals who are eligible to work, but report no positive earnings. While most studies of earnings inequality limit the analysis to employed individuals, we highlight the importance of accounting for the mass of nonemployed workers when analyzing the patterns in earnings inequality.[5] We show that by excluding these inactive workers in the analysis of earnings inequality, traditional measures of inequality like the Gini coefficient actually show a decline in earnings inequality during the Great Recession as workers, mostly from the bottom 80% of the earnings distribution, moved into nonemployment.[6] While we do not receive a direct report of the labor-force participation status of these inactive workers from the administrative records, about 30% of them have reported positive earnings within the last four years, indicating that many have had some recent attachment to the labor market.

With both active and inactive workers in our analysis sample, we decompose the year-to-year changes in the earnings distribution into flows of workers across five employment outcomes from 2004-2013: (i) ineligibility, (ii) nonemployment, (iii) employment with annual earnings in the bottom 20% of the earnings distribution, (iv) employment in the middle 60%, and (v) employment in the top 20%. From this decomposition we find that mobility, both upward and downward, generally occurs between neighboring parts of the inactivity/earnings distribution. Specifically, inactive workers are most likely to transition into employment at the bottom of the earnings distribution, and rarely do employed workers jump from the bottom of the earnings distribution to the top. Furthermore, worker flows explain almost all of the changes in the earnings distribution in the last decade, with the average real earnings of workers in the bottom, middle and top of the earnings distribution remaining fairly constant.

The Great Recession had a very large impact on these flows, with over nine million workers moving into nonemployment and almost four million workers falling from the middle of the earnings distribution to the bottom. While gross outflows from inactivity increased during the recovery, the flows have not been large enough to noticeably reduce the large pool of inactive workers accumulated during the early years of the Great Recession. Furthermore, overall mobility declined, with workers more likely to remain in their current employment state.

To understand the role of the firm in moving workers to the various parts of the earnings distribution, we decompose the above flows by worker and firm types. Specifically, we estimate the

[5] Numerous studies have documented that workers face large earnings losses upon job loss (Jacobson et al., 1993; Stevens, 1997), and that layoffs are highly counter-cyclical (Hall, 2005). However, while there are studies that allow the probability of experiencing unemployment spells to differ across income groups (Castañeda et al., 1998) and some that explicitly take into account the effect of job loss when estimating models of earnings dynamics (Altonji et al., 2013), there is no consensus in the literature on the best method for incorporating inactive workers into a study of earnings inequality.

[6] While many studies have documented a rise in earnings inequality during recessions (for example, Castañeda et al., 1998), other studies have noted that this relationship may not always hold. In particular, Barlevy and Tsiddon (2006) note that the cyclical variation in earnings inequality may also be connected to long-run trends in inequality. More recently, Piketty and Saez (2013) focus on the affect of business cycles on the top of the income distribution, where transitions into unemployment are less of an issue.

statistical model developed in Abowd, Kramarz and Margolis (1999), AKM hereafter, to decompose annual earnings into firm and non-firm components. Then, we characterize the firm component by its position in the overall distribution of firm components. For the person side of the analysis, we consider only the AKM worker effect and the estimated effect of the skill-related regressors. We label this component the skill type. Using the firm-type/skill-type dichotomy, we classify workers by their skill-type location in the earnings distribution and by their employer's location in the firm-component distribution. We examine the mobility patterns and earnings changes for different skill-types of workers employed at different pay-types of firms. We find that while the difference between working at a bottom- or middle-paying firm is relatively small, the gains from working at a top-paying firm are large. In particular, workers benefit from working at a top-paying firm in two ways. First, workers employed at a top-paying firm earn more than similar workers employed at bottom- and middle-paying firms. Second, workers at top-paying firms experience a higher probability of moving up the earnings distribution in the following year. Thus, top-paying firms are associated with increases in both the probability of upward mobility for a worker and the probability of remaining at the top of the earnings distribution.

Related Literature

This paper contributes to two main strands of literature on earnings inequality. The first documents the trends in earnings inequality. The second analyzes the sources of earnings inequality, with a particular emphasis on understanding the role of firms.

Trends in Earnings Inequality There is a large body of work that documents the trends in inequality and tries to identify the sources of these changes. Many of these early studies analyzed the public-use micro-data from the Current Population Survey (CPS). These studies documented a dramatic rise in wage-rate and earnings inequality that started in the late 1970s and continued until the mid-1990s.[7] This widening of the earnings distribution, while still present, has slowed since the mid-1990s. Much of the focus of this literature has shifted towards understanding the polarization of the wage rate distribution.[8] Piketty and Saez (2003) directed the attention of researchers to analyzing the change in the share of income received by the top 1%.

More recent studies also use administrative data to study earnings inequality. Unlike data collected from household surveys, administrative earnings data are generally free from measurement error and top-coding (Abowd and Stinson, 2013), but often lack direct information on labor supply, which makes the distinction between earnings and wage-rate analyses salient. In the U.S., given the absence of information on hours worked on most administrative records, the focus has shifted from wage rates to earnings or income.[9] Kopczuk et al. (2010) use micro-data from the Social

[7] See, for example, Katz and Murphy (1992), Levy and Murnane (1992), Juhn, Murphy and Pierce (1993), and Katz and Autor (1999).

[8] See Autor et al. (2008) and Acemoglu and Autor (2011) for a summary of more recent work on wage polarization. These studies usually define the "wage rate" as the earnings of full-time, full-year workers.

[9] Researchers have also used administrative data from other countries to study earnings inequality. Baker and

Security Administration (SSA) with information on taxable social security earnings to analyze the evolution of earnings inequality and mobility from 1937 to 2004.[10] They find that earnings inequality over this period is U-shaped, decreasing until about 1953, then increasing thereafter.[11] Finally, Spletzer (2014) analyzes the trends in earnings equality in the LEHD data. The trend in earnings inequality depends on the workers included in the sample for analysis. Specifically, he shows that quarterly earnings inequality has been increasing among workers who are very attached to the labor market.[12] However, annual earnings inequality has not changed since around 2000 when Spletzer (2014) includes all workers age 15 and older in his analysis sample, a result consistent with our findings when we use the data on all available jobs, but inconsistent with our findings when we correct the estimation frame to exclude job records associated with what we call ineligible workers.

With the availability of all these data sources for studying earnings inequality, Spletzer (2014) also analyzes how the patterns in inequality vary across data sources. While he conducts detailed comparison across several sources, we focus on the comparison between Current Population Survey's Annual Social and Economic Supplement (CPS-ASEC) and the LEHD data.[13] In particular, he compares the evolution of the ratio of the 90^{th} to the 10^{th} percentile in CPS-ASEC and in LEHD. When comparing these two sources, Spletzer documents different patterns in the 90/10 ratio across the two that are very similar to our findings. In particular, the 90/10 ratio in CPS-ASEC has been increasing over the last 15 years, while the 90/10 ratio in LEHD has remained relatively flat. While Spletzer (2014) identifies that it is differences in the evolution of the bottom half of the earnings distribution causing this discrepancy, he does not further analyze the differences in the sample of workers covered by CPS-ASEC (a designed frame) and those covered by LEHD (a found frame), as we do.[14]

The Role of Firms in Earnings Inequality To evaluate the role of firms in the rise in inequality, many papers estimate a variation of the statistical model presented in AKM.[15] The data

Solon (2003) use longitudinal income tax records to decompose the growth in earnings inequality in Canada into its permanent and transitory components. They find that both of these components are important.

[10] Kopczuk et al. (2010) use several data sources from SSA for their analysis. See section II.B of their paper for the precise details.

[11] Guvenen et al. (2014) also use micro-data from SSA; however, their earnings measure comes from W-2 forms (box 1) submitted directly by the employers. They focus on earnings risk changes during recessions; in their results, the trend in earnings inequality, which can be computed from their reported earnings percentiles, shows a small decline in earnings inequality during the 1990s until about 2000. Since 2000, earnings inequality has been on the rise. See Table A.3 of their Appendix A for the percentiles of the earnings distribution.

[12] To proxy for full-time workers, Spletzer (2014) only includes full-quarter workers: individuals with at least three consecutive quarters of positive earnings, for whom the interior quarter is studied.

[13] Spletzer (2014) also compares LEHD data to the data from the outgoing rotation group contained in the CPS basic monthly files and the IRS data used in Saez (2015).

[14] Spletzer (2014) notes that the scope of coverage of CPS-ASEC is previous year income from all jobs for all persons aged 15 and over currently residing in the household. To get comparable estimates from LEHD, he computes annual earnings from all jobs for all workers aged 15 and over. While we impose no age restriction on our found frame, the results in Spletzer (2014) are very comparable. Further, it is interesting to note that just imposing an age restriction, as is done in many studies, is not sufficient to reconcile the differences in earnings inequality trends observed in CPS-ASEC and LEHD when using the underlying LEHD data without adjustments to their frame.

[15] A number of papers in this volume use or analyze the AKM decomposition. Freeman et al. (forthcoming) directly apply the AKM method to the LEHD data, using the all-workers frame we discuss below. Card et al. (forthcoming)

requirements that allow for the identification of separate worker and firm fixed effects–longitudinal links in both dimensions with sufficient network connectivity–almost always restrict such studies to administrative data. Card et al. (2013) applied this statistical technique to administrative data from Germany. Their analysis focused on full-time, full-year male employees. Their estimates suggest that, for this group of workers, the rise in German wage-rate inequality was in part due to the increase in the dispersion in wage premiums paid by firms. Song et al. (2015) take a nonparametric approach to measurement of the firm's contribution to the rise in earnings inequality. Using SSA earnings data from the W-2 tax information forms from 1978-2012, they decompose the rise in earnings inequality into the part attributed to rising dispersion between firms in the average earnings they pay their employees and the part attributed to rising earnings dispersion among workers within a firm. They find that virtually all of the rise in earnings inequality is accounted for by an increase in the dispersion in the average earnings paid by firms. In their data, earnings differences among workers at the same firm have remained fairly constant over this period.

The remainder of the paper is organized as follows. Section 2 discusses the Unemployment Insurance (UI) based individual- and job-level data we use in our analysis and discusses the creation of our all-workers and eligible-workers frames. Section 3 presents and compares inequality trends in found versus designed data. This section also briefly discusses the results of comparing our analysis frames to data from the Current Population Survey and the American Community Survey. Section 4 analyzes the evolution of the earnings distribution over time and presents a worker flow decomposition of the year-to-year change in the earnings distribution. Section 5 analyzes the importance of firms in understanding these changes in the earnings distribution. Section 6 concludes.

2 Data Sources and Methods

The empirical work in this paper uses three different sources of earnings information. The primary data source is the Longitudinal Employer-Household Dynamics (LEHD) infrastructure files, developed and maintained by the U.S. Census Bureau.[16] We also use two household surveys: the Current Population Survey (CPS) and the American Community Survey (ACS). From these three data sources, we create two annual person-level earnings files covering the period 1990-2013. We construct the first file using the LEHD data and the second file using repeated cross-sectional CPS and ACS microdata.

In the LEHD data infrastructure, a "job" is the statutory employment of a worker by a statutory employer as defined by the Unemployment Insurance (UI) system in a given state. Man-

develop a full economic model that interprets AKM in the equilibrium, and apply that model to data from Portugal. Van Reenen et al. (forthcoming) use AKM worker and firm effects to augment their analysis of German employers. Haltiwanger et al. (forthcoming) use the LEHD data in the all-workers frame, relying on observable firm characteristics rather than the AKM decomposition to study the employer's contribution. Juhn et al. (forthcoming) measure earnings volatility from the LEHD data, using the all-workers frame and requiring two consecutive years of positive earnings to study whether firm revenue shocks are transmitted to workers.

[16]See Abowd et al. (2009) for a detailed summary of the construction of the LEHD infrastructure.

dated reporting of UI-covered wage and salary payments between one statutory employer and one statutory employee is governed by the state's UI system. Reporting covers private employers, state and local government. There are no self-employment earnings unless the proprietor drew a salary, which is indistinguishable from other employees in this case.

The Office of Personnel Management (OPM) supplied federal jobs data, included from 2000Q1 forward. The OPM data were edited as part of the LEHD infrastructure processing to produce records containing quarterly earnings reports comparable to those reported directly in the UI wage and salary payments. As part of this processing, pseudo-UI account numbers were created using the observed combinations of duty station state and agency/sub-element.[17] The result is a set of state-level employer identifiers conceptually similar to those found on the UI data for private firms.

Due to national security regulations, which suppress certain jobs from the ones released by OPM to the public and other agencies, the coverage of the OPM extract varies by agency. Under-coverage is particularly severe for the Department of Defense (including the Air Force, Army, and Navy), Department of Justice, Department of State, and the Department of Treasury. Although the federal jobs data are typically not included as part of the state-based UI system, going forward in this paper, when we say "UI-covered" employment, we mean "statutory employment" as defined by the UI system or a statutory federal employee.

2.1 Date Regimes

States and the federal government joined the partnership that supplies input data to the LEHD program at different dates. When a state or the federal government joined, the data custodians were asked to produce historical data for as many quarters in the past, back to 1990Q1, as could be reasonably recovered from their information storage systems. As a result, the date that a data-supplying entity joined the partnership is not the same as the first quarter in which that entity's data appear in the system. The start date for any state or the federal government depends primarily on the amount of historical data the state or federal government could recover at the time it joined. This potential ignorability (in the sense of Rubin (1987) or Imbens and Rubin (2015)) of the start data for a segment of the LEHD data is the basis for our methods of constructing nationally representative estimates back to the early 1990s.[18]

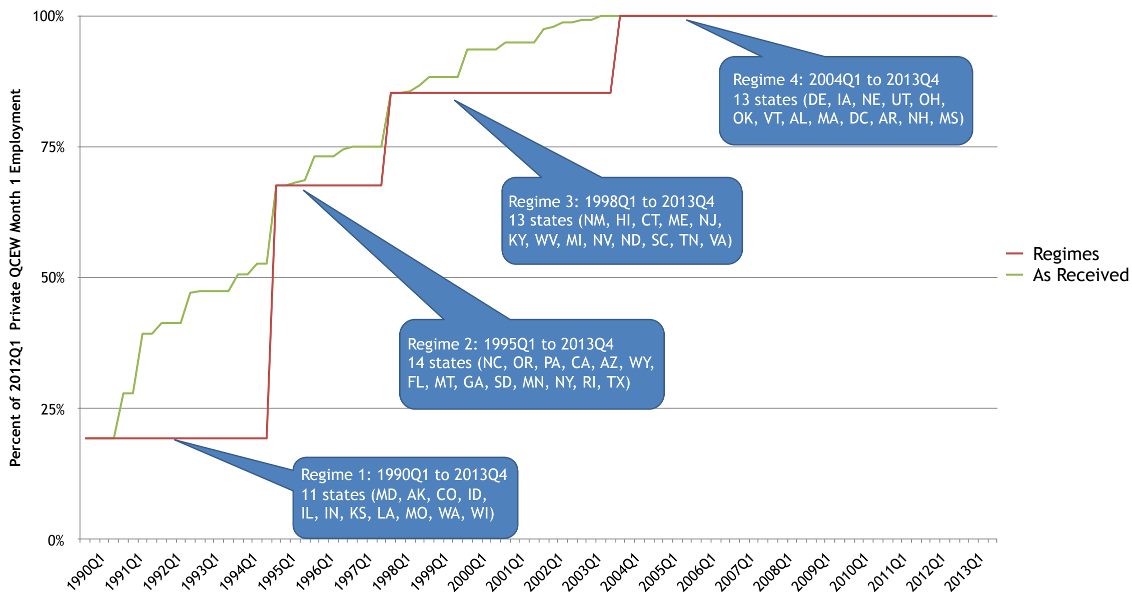
To understand how state entrance into the LEHD program affects the trends in earnings inequality, we analyze four regimes that correspond to different epochs of data availability:

- Regime 1: 1990-2013 [19% of 2012Q1 QCEW private employment]
  (AK CO MD ID IL IN KS LA MO WA WI);
- Regime 1 and 2: 1995-2013 [68%]

[17]See `https://www.opm.gov/policy-data-oversight/data-analysis-documentation/data-policy-guidance/reporting-guidance/part-a-human-resources.pdf` for a list of agency codes.

[18]See Abowd and Vilhuber (2011) for a description of how these modeling assumptions were used to construct national gross worker and job flow estimates.

Figure 1: LEHD Infrastructure Data as Received and by Date Regimes



Notes: The figure shows the coverage of the LEHD infrastructure data expressed as the percent of 2012Q1 private QCEW employment as received (green line) and by date regime (red line). OPM data for federal workers are not shown in this figure, but are available beginning in 2000Q1.

(+ AZ CA FL GA MN NC NY OR MT PA RI SD TX WY);

- Regime 1, 2, and 3: 1998-2013 [85%]
  (+ CT HI KY MI ND NJ NM NV SC TN VA WV);
- Regime 1, 2, 3, and 4: 2004-2013 [100%]
  (+ AL AR DC DE IA MA MS NE NH OH OK UT VT).

The number in brackets represents the percentage of private employment as reported in the BLS Quarterly Census of Employment and Wages (QCEW) in 2012Q1 for the regimes listed. The construction of the different date regimes is shown graphically in Figure 1. Appendix Table A.1 presents a detailed tabulation of the coverage of each date regime.

We did not create a separate date regime when the OPM data begin in 2000Q1 for two reasons. First, the proportion of **reported** federal jobs in 2000Q1 is small, no more than four percent of 2012Q1 UI covered employment in a state and, on average, less than one percent, except for DC, where the proportion of OPM jobs is about nineteen percent. Second, although the proportion of OPM jobs in DC is high, jobs in DC are part of regime four, which does not begin until 2004Q1.

2.2 Job and Person Sampling Frames

By 2004Q1 the LEHD data represent the complete universe of statutory jobs in the U.S.: all fifty states, the District of Columbia, and the federal government are reporting regularly. Before this date, LEHD data provide a complete frame for the states in each date regime (excluding the federal government before 2000Q1). After this date, the LEHD data, in theory, provide a complete frame for the national population of UI covered jobs (including federal employees). Studying job-level inequality, the task for which having a complete job frame is well suited, as a proxy for person-level inequality may be misleading due to the time-varying many-to-one assignment of jobs to workers. The number of employers per worker varies over the business cycle. Lower-earning workers tend to have more employers, complicating the interpretation of job-level results.[19] Therefore, when studying inequality, it is preferable to have a person frame that covers a known population of interest, such as all persons legally eligible to work in the United States. For our analyses, we use two different approaches to creating frames of jobs and workers. The first approach relies only on the employer-employee links present in the UI data. This method captures all reported jobs. The second approach uses the Census Bureau's edited version of the Social Security Administration's master SSN database (the "Numident"), capturing all reported employment-eligible workers but removing jobs associated with ineligible workers, as we elaborate below.

LEHD earnings records are reported quarterly by the employing firm. These records contain a nine-digit person identifier, typically assumed to be a Social Security Number. However, at the time the report is received by the state UI office, the nine-digit person identifiers are not verified, resulting in records both with and without a valid SSN. Using the Numident, we ascertain if each earnings record is associated with a valid SSN. Records not associated with a valid SSN may have an alternate, valid person identifier, such as an IRS-issued Taxpayer Identification Number; nevertheless, we can only distinguish between valid and invalid SSNs. If the SSN is valid, we have access to demographic characteristics, such as sex and birth date, from the Numident and other Census sources. We also have an employment history from the UI wage records. If the SSN is not valid, we only have access to the employment history.

Our first 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.

Our second frame, the "eligible-workers frame" is also delineated by these date regimes, but includes jobs only for the subset of the all-workers frame that meets the following criteria:

- has a valid SSN on the Numident;
- appears in the frame every year from 1990 to 2013 in which that individual is between the ages of 18 and 70, inclusive;
- the year of the recorded data is greater than or equal to the SSN year of issue, and less than or equal to the year of death (if available); and
- has an SSN that was associated with fewer than 12 jobs during the data year.

An eligible worker is labeled as "active" in the labor market in a given year when UI earnings are positive, and "inactive" otherwise. The valid SSN, age-range, date of death, number of jobs,

[19]See Hirsch and Husain (2013) and the references therein for a summary of the literature on the cyclicality of multiple job-holding. The authors find that the multiple job-holding rate is pro-cyclical, declining during recessions because of the increased slackness in the labor market.

and active worker restrictions remove about 7% of the worker-year records found in the all-workers frame.

The purpose of the eligible-workers frame is twofold. The Numident allows us to consistently identify a set of persons legally eligible to work each year, while at the same time implicitly removing earning records from our analysis sample that are not associated with individuals in the covered population. However, we also go a step further, and remove earnings records with a valid SSN where the available data strongly suggest that the SSN is not being used by the original owner.[20] These two types of suspect nine-digit person identifiers—invalid SSNs that do not match to the Census Numident and valid SSNs apparently being used by multiple persons and/or for whom the age of the person issued the SSN is inconsistent with labor-market activity—we call "immigrant candidates."

Table 1 presents basic counts of persons and jobs in the eligible-workers frame and in the immigrant candidate file. While we present some analysis of the immigrant candidate jobs in Appendix A, we do not have sufficient information to convert the collection of these jobs into an inter-temporally consistent frame for this population of individuals. We have no plausible means of determining how many immigrant candidates are using each SSN in this collection of UI wage records.

Even accounting for the increasing coverage of each date regime, there is a clear, strong upward trend in the number of immigrant-candidate records until the Great Recession, and a strong downward trend thereafter. Because there are data in the system for each of these records, the associated nine-digit person identifier represents at least one worker, but may represent many. A great deal of supplemental research would be required to estimate the relation between how many jobs are reported for a "worker" in the immigrant-candidate records and the number of individuals employed. Even more research would be required to estimate their characteristics. We do not attempt such research here. We note that when we use the all-workers frame, each associated nine-digit person identifier counts as one individual, but there is no adjustment for being inactive; that is, we do not assume zero earnings when this "worker" has no reported earnings. When we use the eligible-workers frame, all immigrant-candidate records are excluded. Eligible inactive workers are assumed to have zero earnings when there is no reported activity in any job.

Appendix A contains additional detailed analyses of the construction of the eligible-workers frame. In particular, Appendix Figure A.1 plots the count of earnings records excluded from the eligible-workers frame each year, disaggregated by earnings records each year associated with individuals who have invalid SSNs, who are either too young or too old, and/or who report working 12 or more jobs in a year. Detailed counts are reported in Appendix Table A.2.

[20] The use of SSNs not originally issued to the person using the SSN has been documented and studied by Brown et al. (2013) and others.

Table 1: Observations per Year for Eligible Workers and Immigrant Candidates

	Eligible Workers				Immigrant Candidates	
Year	**Persons**	**Inactive**	**Active**	**Jobs**	**Active**	**Jobs**
1990	190,814,228	167,224,366	23,589,862	34,936,872	1,585,988	2,173,054
1991	192,605,253	169,444,287	23,160,966	33,446,875	1,503,127	2,029,041
1992	194,341,714	171,098,747	23,242,967	33,588,257	1,485,173	2,024,225
1993	196,043,096	172,414,178	23,628,918	34,733,064	1,590,507	2,227,908
1994	197,726,878	173,562,792	24,164,086	36,705,168	1,737,988	2,546,460
1995	199,524,643	116,578,567	82,946,076	128,077,314	6,666,437	9,875,811
1996	201,276,549	117,660,582	83,615,967	129,681,393	6,725,830	10,144,571
1997	203,229,484	117,863,679	85,365,805	134,003,889	6,925,825	10,560,373
1998	205,266,723	94,856,664	110,410,059	173,778,794	8,919,168	13,680,138
1999	207,478,545	94,930,531	112,548,014	178,813,085	9,473,798	14,850,424
2000	209,895,465	94,701,668	115,193,797	184,243,425	9,975,102	15,909,402
2001	212,479,460	96,709,036	115,770,424	178,433,884	9,703,887	15,142,444
2002	214,891,408	99,804,179	115,087,229	172,249,424	9,109,574	13,646,946
2003	217,298,533	102,254,299	115,044,234	169,454,044	8,715,459	13,105,529
2004	219,763,469	83,200,954	136,562,515	202,935,084	10,218,971	15,254,789
2005	222,160,089	83,819,319	138,340,770	207,737,171	10,577,475	16,109,360
2006	224,721,578	84,357,718	140,363,860	212,227,031	10,943,861	16,830,576
2007	227,553,012	85,518,594	142,034,418	213,889,946	10,818,763	16,464,027
2008	230,355,015	88,245,425	142,109,590	206,906,286	9,845,415	14,509,746
2009	232,813,313	94,864,949	137,948,364	188,220,068	8,358,069	11,701,711
2010	234,304,705	96,959,047	137,345,658	188,740,259	7,919,195	11,019,697
2011	235,429,997	96,619,700	138,810,297	193,351,447	7,813,411	10,942,606
2012	236,484,312	96,068,987	140,415,325	197,879,809	8,254,250	11,556,277
2013	237,816,938	96,151,327	141,665,611	202,277,055	9,310,868	13,216,695

Notes: The table presents counts of the number of persons and jobs in the eligible-workers frame and in the immigrant candidates file. The sum of these two components is the all-workers frame. The persons in the eligible-workers frame are disaggregated into those who report positive earnings (*Active*) and those who do not (*Inactive*). The frame is complete and covers the entire U.S. from 2004 forward.

2.3 Earnings Definitions and Coverage

In this section we define our earnings measures for both the all-workers and the eligible-workers frames. Our primary measure of earnings is based on annual UI job-level earnings reports. We adjust nominal earnings to real earnings using the Consumer Price Index (CPI-U), with 2000 as the base year. Let y_{ijt} be the real earnings for worker i employed at firm j in year t. Person-level annual earnings sum all jobs for each worker in each year:

$$e_{it} = \sum_{j} y_{ijt}. \tag{1}$$

Using e_{it}, we estimate total annual earnings for the eligible-workers frame in year t using:

$$E_t^{EW} = \sum_{i|(i,t) \in EW_t} e_{it}, \tag{2}$$

where EW_t is the set of workers in the eligible-workers frame in year t. For the period 2004-2013, when our frames contain data for the entire U.S., the eligible-workers frame is approximately 90% of wage and salary compensation as defined in the National Income and Product Accounts.[21]

2.4 Estimation of the Earnings/Inactivity Distribution

We begin by calculating deciles of eligible-workers person-level earnings, e_{it}, pooled across the years 2004-2013. Using these deciles, we create three earnings bins as shown in Table 2. The bins are designed to capture the bottom, middle, and top of the earnings distribution over the entire ten-year period. For example, the first two columns in Table 2 show the results for bin 2, workers in the bottom 20% of the earnings distribution.[22] Workers in this bin have a minimum annual earnings value of \$2, a maximum value of \$6,600, and a mean log earnings in real 2000 dollars of \$7.473 (implied geometric mean real earnings \$1,760).

The low mean log earnings in bins 2 and 3 suggests that a large proportion of workers in these bins are employed for only part of the year. In Table 3 we present information about the labor-force activity of workers in each of the three earnings bins. Each row in the table (except the residual category "All Other") represents a specific combination of quarters worked and number of quarters in the longest job, truncated at a maximum of six quarters. The number of quarters in the longest job takes on values from one to six. 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*. Thus a six-quarter active job is active at the beginning of a calendar year, the end of a calendar year, and all quarters in the middle. The most prevalent pattern in each earnings category is listed first, followed by the next most prevalent, and continuing until the table contains the patterns for approximately 80% of the workers in a typical year.

[21] Appendix A provides a detailed analysis of the earnings coverage for each of our frames in comparison with BEA National Income and Product Accounts (NIPA) annual wages and salary estimates.

[22] Bin 1 is reserved for eligible but inactive workers, who are not included in the summaries described in Table 2.

Table 2: Descriptive Statistics by Real Earnings Bins

	Earnings Bins (e_{it})					
	2: *Bottom 20%*		**3**: *Middle 60%*		**4**: *Top 20%*	
Minimum	0.693	($2)	8.795	($6,600)	10.750	($46,800)
Mean	7.473	($1,760)	9.938	($20,700)	11.240	($75,810)
Maximum	8.795	($6,600)	10.750	($46,800)	16.140	($10,230,000)

Notes: The table presents statistics on real earnings for three categories of workers: (i) *Bin 2*: bottom 20% of the pooled annual-earnings distribution of eligible workers from 2004-2013, (ii) *Bin 3*: the middle 60%, and (iii) *Bin 4*: the top 20%. The rows show the minimum, mean, and maximum log earnings, $\log(e_{it})$, of each bin. The exponentiated values (implied geometric means) are listed in parentheses. All earnings are in real 2000 dollars (adjusted using CPI-U). Note that the minimum and maximum values of each bin are rounded to four significant digits.

The dominant labor-force attachment pattern varies substantially across bins: 31% of the workers in the bottom 20% of the earnings distribution work only one quarter. In contrast, for workers in the top 20% of the earnings distribution, the most common labor force status is employment with at least one firm for the entire year. Although there are almost certainly large differences in average wages for workers at the bottom, middle and top of the earnings distribution, one of the primary reasons average annual earnings for workers in the bottom earnings category are so low is they are employed for only a small portion of the year.

Finally, we combine the earnings bins discussed above with the active/inactive status information available for the eligible-workers frame to create four mutually exclusive earnings/inactivity categories:

1. eligible to work, but no reported UI earnings (inactive);
2. working and in the bottom 20% of the overall UI earnings distribution;
3. working and in the middle 60% of the overall UI earnings distribution; and
4. working and in the top 20% of the overall UI earnings distribution.

We analyze these four bins comprehensively in Sections 4 and 5, when we study the dynamics of earnings distribution changes and the role of the firm.

3 Inequality Trends in the LEHD Infrastructure Data (1990-2013)

The determination of an appropriate frame for studying changes in earnings inequality led us to analyze our eligible-workers frame, which is the complete population for the United States during the critical period from 2004-2013. The results in this section, and the rest of the paper, relate exclusively to the eligible-workers frame. Where we have also analyzed comparable data for the all-workers frame, we reference results in Appendix B.

We begin by examining the evolution of the percentiles of the earnings distribution. To

Table 3: Labor-Force Activity of Workers in Each Earnings Bin

Quarters Worked	**Longest Job**	**Workers** *Counts*	*Pct All*	*Pct Bin*	**Jobs** (*Avg*)	**Earnings** (*Avg*)
		Bottom 20% of Earnings Distribution				
1	1	8,543,957	6.1%	30.6%	1.066	$1,366
2	2	5,806,138	4.2%	20.8%	1.213	$2,824
4	6	2,893,038	2.1%	10.4%	1.251	$4,227
3	3	2,591,936	1.9%	9.3%	1.263	$3,726
3	2	2,467,851	1.8%	8.8%	2.297	$3,480
All Other		5,608,961	4.0%	20.1%	2.306	$3,472
		Middle 60% of Earnings Distribution				
4	6	52,012,001	37.3%	62.1%	1.212	$26,110
4	5	8,869,511	6.4%	10.6%	1.602	$22,410
4	3	7,105,740	5.1%	8.5%	2.592	$20,570
All Other		15,748,549	11.3%	18.8%	1.786	$16,626
		Top 20% of Earnings Distribution				
4	6	22,653,328	16.2%	81.2%	1.181	$88,450
All Other		5,258,632	3.8%	18.8%	1.756	$91,630

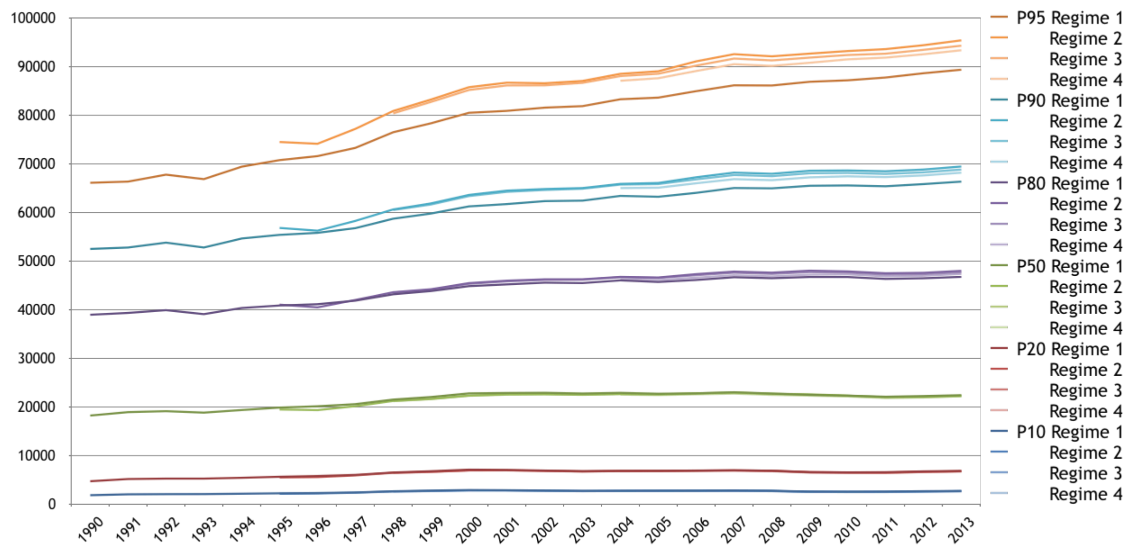
Notes: Each row in the table represents a specific combination of quarters worked and number of quarters in the longest job. The number of quarters in the longest job takes on values from one to six. A five-quarter longest job is active in *either* the fourth quarter of the previous year or the first quarter of the subsequent year, while a six-quarter longest job is active in *both*. The most prevalent pattern in each earnings bin is listed first with the next most prevalent second, continuing until the table contains the patterns for approximately 80% of the workers in a typical year. *All Other* is the residual category. The counts are averages per year from the eligible-workers frame. The full table of counts is available in Appendix Table A.4.

understand how state entry into the LEHD partnership might affect the earnings distribution over time, we estimate percentiles of the earnings distribution for workers in each of the four date regimes by year. Figure 2 plots the 10^{th}, 20^{th}, 50^{th}, 80^{th}, 90^{th}, and 95^{th} percentiles of the cumulative earnings distribution by date regime. The cumulative distribution contains data for all regimes less than or equal to the sampling regime shown in the legend.

Notice that differences in the 10^{th}, 20^{th}, and 50^{th} percentiles are virtually indistinguishable across date regimes. Above the median, however, there is some variation in the 80^{th}, 90^{th}, and 95^{th} percentiles across date regimes. In particular, the inclusion of data from high earnings and populous states, like CA and NY, in date regime 2 increases the estimated quantile values, especially for the 95^{th} percentile. However, once these states have entered, the effect of the states entering in regimes 3 and 4 is much less pronounced. The level differences in the percentiles are relatively minor, especially after 1995 (regime 2). The trends in inequality are very similar across date regimes. We conclude that the date of state entry does not appear to significantly affect earnings inequality measures, especially after 1995.[23]

[23] Appendix Figure B.1 plots the same percentiles comparing the two worker frames for all available years, equivalent to the regime 4 cumulative distribution in Figure 2. That comparison indicates that the levels of all percentiles are greater in every year in the eligible-worker frame, as compared to the all-worker frame, but the trends are nearly identical. For both frames, there is no evidence of differences that are due to the dates in which states entered the frames, except for the jump associated with the entry of CA and NY in regime 2. Appendix 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. The fact that

Figure 2: Percentiles of the Earnings Distribution for Eligible Workers by Date Regime



Notes: The figure plots the 10^{th}, 20^{th}, 50^{th}, 80^{th}, 90^{th}, and 95^{th} percentiles of the earnings distribution of eligible workers by date regime and year. The cumulative distribution plots the data for all regimes less than or equal to the indicated regime. For example, "P95 Regime 2" indicates the 95^{th} percentile for all states in regimes 1 and 2.

We next turn to the evolution of earnings inequality over the past 23 years. Figure 3 presents the following measures of earnings inequality: (i) the ratio of the 99^{th} and the 1^{st} percentiles (99/1 ratio, and so forth), (ii) the ratio of the 95^{th} and the 5^{th} percentiles, (iii) the ratio of the 90^{th} and the 10^{th} percentiles, (iv) the ratio of the 80^{th} and the 20^{th} percentiles, and (v) the variance of log earnings.[24] These measures are all reported relative to their value in 2000.[25] After 2000, there is a persistent increase in earnings inequality according to all measures. On average, the 99/1 ratio is 15.4% higher; the 95/5 ratio is 13.6% higher; and the 90/10 ratio is 11.4% higher than in 2000. Much of the rise in earnings inequality occurs during the Great Recession and persists into the recovery. For example, the 90/10 ratio was on average 6.7% higher from 2001-2007 than in 2000. Then, during the Great Recession, the 90/10 ratio was 15.1% higher from 2008-2009 than in 2000. This increase does not peak until 2010, resulting in inequality being 17.8% higher from 2010-2013 than in 2000. Except for the 99/1 ratio, post-2000 trends do not appear in the all-workers frame.[26]

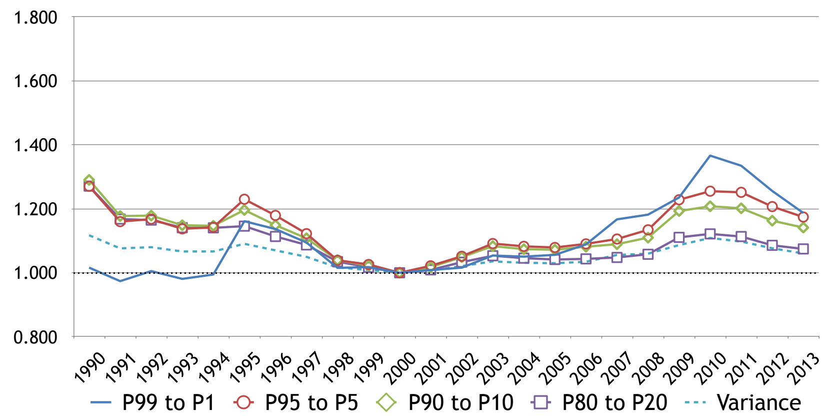
the earnings distribution is unaffected by state entry is very strong evidence that the date of entry of a state into the LEHD infrastructure can be modeled as ignorably missing (Rubin, 1987; Imbens and Rubin, 2015).

[24] We also analyze how state entry into the LEHD data affects these measures of earnings inequality. Appendix Figure B.2 plots the 90/10 ratio across date regimes. There are only minor differences across these regimes. Thus, the date of state entry does not affect the analysis of earnings inequality. Therefore, in the main discussion, we present results from the overall distribution where we include data from all available states in a given year.

[25] For a comparison of the levels of these ratios across the two frames, see Appendix Figures B.5a, B.5b, B.6a, and B.6b.

[26] The statistics quoted in this paragraph are summarized in Appendix Table B.2, where they are also compared to ratios obtained when the all-workers frame is used. Removing the immigrant candidates from the frame materially alters the estimated trends in earnings inequality. While there is a similar decline in earnings inequality from 1995 to 2000 in both frames, post-2000, the trends in earnings inequality among eligible workers diverge from those observed among all workers. Appendix Table B.2 shows that, on average, while the 99/1 ratio was 15.4% higher post-2000

Figure 3: Selected Inequality Measures 1990-2013, Relative to 2000



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. Results are based on the eligible-workers frame from the LEHD infrastructure files.

Figure 4 presents changes in the top and bottom halves of the earnings distribution by decomposing the ratios around the median.[27] At the top of the distribution, we compute the ratio of the 99^{th} to the 50^{th} percentile (99/50 ratio, and so forth), the ratio of the 95^{th} to the 50^{th} percentile, the ratio of the 90^{th} to the 50^{th} percentile, and the ratio of the 80^{th} to the 50^{th} percentile. At the bottom, we analyze the ratio of the 50^{th} to the 1^{st} percentile, the ratio of the 50^{th} to the 5^{th} percentile, the ratio of the 50^{th} to the 10^{th} percentile, and the ratio of the 50th to the 20^{th} percentile. These ratios are all reported relative to the their value in 2000.[28] The top and the bottom of the earnings distribution have evolved quite differently. Since 2000, the ratios of the top percentiles to the median have been increasing very gradually, as shown in Figure 4a. However, this increase at the top has been small compared to the rise in inequality at the bottom of the earnings distribution as shown in Figure 4b. The ratios of the median to the bottom percentiles have been increasing dramatically, indicating earnings growth at the bottom of the distribution has not kept up with earnings growth at the median. The bottom of the earnings distribution is more cyclically sensitive with much of the rise in inequality occurring during the Great Recession.[29]

in the eligible-workers frame compared to only 5.0% higher in the all-workers frame, none of the other post-2000 inequality measures are rising in the all-workers frame.

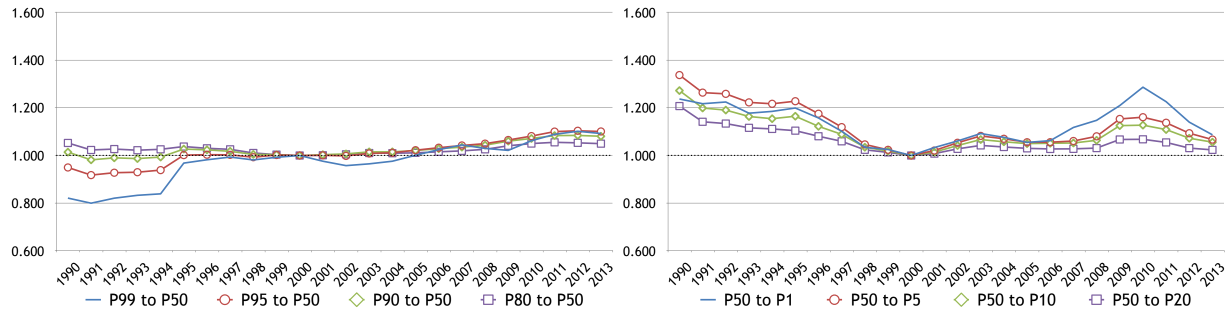
[27]For example, decomposing the 90/10 ratio in year t (relative to 2000) around the median:

$$\frac{90_t^{th}}{10_t^{th}} \Big/ \frac{90_{2000}^{th}}{10_{2000}^{th}} = \left(\frac{90_t^{th}}{50_t^{th}} \cdot \frac{50_t^{th}}{10_t^{th}}\right) \Big/ \left(\frac{90_{2000}^{th}}{50_{2000}^{th}} \cdot \frac{50_{2000}^{th}}{10_{2000}^{th}}\right) = \left(\frac{90_t^{th}}{50_t^{th}} \Big/ \frac{90_{2000}^{th}}{50_{2000}^{th}}\right) \cdot \left(\frac{50_t^{th}}{10_t^{th}} \Big/ \frac{50_{2000}^{th}}{10_{2000}^{th}}\right)$$

[28]For a comparison of the levels of these ratios across the two frames, see the bottom four subplots in Appendix Figures B.5 and B.6.

[29]These conclusions are materially different in the all-workers frame, which shows modest changes in inequality of roughly equal magnitudes at the top and bottom of the distribution. See Appendix Figures B.4a and B.4b.

Figure 4: Selected Inequality Measures for the Top and Bottom of the Earnings Distribution 1990-2013, Relative to 2000



(a) Eligible Workers - Top

(b) Eligible 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 eligible-workers frame from the LEHD infrastructure files. See Appendix Figure B.4 for comparable data using the all-worker frame.

To summarize, earnings inequality has been on the rise since 2000 and spiked during the Great Recession. This conclusion depends materially on our use of the eligible-workers frame. Direct use of the all-workers frame–that is, using all the job records in the LEHD infrastructure files– produces earnings inequality measures and trends that appear not to change from 2000 through the Great Recession to 2013. Constructing the eligible-workers frame by excluding immigrant candidate records shifts the earnings distribution of eligible workers to the right of the earnings distribution of all workers. While this shift in the earnings distribution is not particularly surprising, the resulting change in the trends in earnings inequality between the two frames is. When studying changes in earnings inequality over time, especially when using administrative data, the choice of worker frame matters substantially.

3.1 Comparison of LEHD Data to the CPS and ACS

To understand the differences between analyses using respondent-provided earnings data in large-scale household surveys and those using administrative data on earnings in the unemployment insurance system, we constructed detailed analysis samples from the Current Population Survey-Annual Social and Economic Supplement (CPS-ASEC, 1990-2004) and the American Community Survey (ACS, 2000-2013). We also linked the LEHD earnings data to the ACS data at the individual record level (2005-2013). We used the linked data to study differences between individuals found only in the ACS data and those found in both sources.[30] The details of the data construction can

[30]Because the ACS is a sample survey while the LEHD data are essentially the population over this period, we cannot distinguish between individuals found only in the LEHD data who should have linked to the ACS and those

be found in Appendix C.1. The details of the data linkage can be found in Appendix C.3.

In the household survey data, we defined "covered workers" to be individuals who worked in a job that should have appeared in the LEHD data. Table 4 show that the estimated percentile values of the earnings distribution tend to be greater for covered workers than for all workers in the household survey data.[31] For percentiles at or above the median, the values from the LEHD eligible-workers frame are close to the ones from the covered workers in 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 the LEHD eligible-workers frame. For example, earnings associated with the 10^{th} percentile in the CPS/ACS data are close to the 20^{th} percentile in the LEHD data.[32] We conclude that the differences between survey-reported and administrative earnings data at or above the median are minimal, but that, below the median, household surveys capture earnings that are not captured by the administrative data.[33]

Table 4: Average Percentiles of the Earnings Distribution from Household Surveys vs. LEHD

	Percentiles						
Frame	*5th*	*10th*	*20th*	*50th*	*80th*	*90th*	*95th*
CPS/ACS: Covered Workers	3,419	6,412	11,703	26,345	49,059	67,872	89,323
LEHD: Eligible Workers	1,005	2,527	6,463	21,762	45,343	64,021	86,108
Difference: HHLD Survey − UI	2,414	3,884	5,240	4,583	3,716	3,850	3,215
Ratio of HHLD Survey to UI	3.4027	2.5370	1.8108	1.2106	1.0819	1.0601	1.0373

Notes: The first row presents the average percentile values from the earnings distribution of covered workers in the combined CPS/ACS data 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 third row computes the difference between the average percentiles in the household surveys and the same ones computed from the LEHD eligible-workers frame. The last row computes the ratio of each percentile from the covered workers in the household surveys to the LEHD eligible- workers frame.

Figures 5a and 5b show that the trends in earnings inequality in the household survey data are consistent with our findings from the eligible-worker frame in the LEHD data, and inconsistent with the findings from the all-worker frame. This conclusion also holds for the analysis of trends in the top and bottom separately.[34] The pattern of inequality ratios around the Great Recession is also similar to the one found in the eligible-workers frame, and dissimilar to the the analysis based on the all-workers frame.[35] Additional results and discussion are in Appendix C.2.

who were not sampled by the ACS. For this reason, we don't study the records of those in the LEHD-only group.

[31] Also compare Appendix Figure C.1 with Appendix Figure B.1.

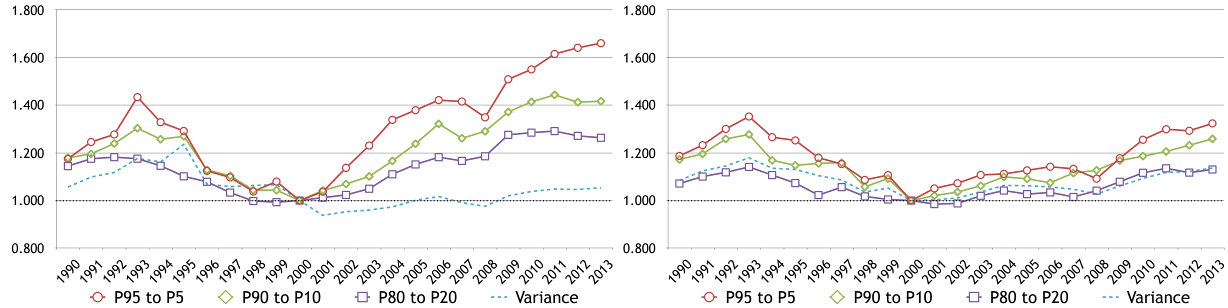
[32] See also Appendix Figure C.2.

[33] This conclusion is not sensitive to using the eligible-workers frame. There are many more very low earnings records among the immigrant candidates, which would only exacerbate the differences below the median.

[34] See Appendix Figures C.3a and C.3c.

[35] See Appendix Table C.1.

Figure 5: Selected Inequality Measures 1990-2013, Relative to 2000 (CPS/ACS)



(a) All Workers from Household Surveys

(b) Covered Workers from Household Surveys

Notes: Subplot (a) presents measures of earnings inequality for *all* workers in CPS/ACS relative to 2000 from 1990-2013. Subplot (b) presents measures of earnings inequality for *covered* workers in CPS/ACS relative to 2000 from 1990-2013. The measures of earnings inequality considered are (i) *P95 to P5*: the ratio of the 95^{th} to the 5^{th} percentile; (ii) *P90 to P10*: the ratio of the 90^{th} to the 10^{th} percentile; (iii) *P80 to P20* the ratio of the 80^{th} to the 20^{th} percentile; and (iv) *Variance*: the variance of log annual earnings.

To better understand the discrepancies at the bottom of the earnings distribution, we analyzed linked ACS and LEHD individual records. We find that 85% of the individuals we expect to find in both the ACS and LEHD data are present in both with positive earnings. The other 15% are not found in the LEHD data, and, therefore, have zero administrative earnings.[36] The result suggests that survey respondents are giving answers that imply that their job should be covered in the administrative data, but there are not corresponding administrative records to match the survey-reported income. The reported income of these individuals appears to be the reason for the discrepancy between the survey and administrative data at the bottom of the earnings distribution.[37] Additional results and discussion are in Appendix C.3.

Our analysis of the trends in earnings inequality using the various samples of workers in the CPS/ACS and the two worker frames in the LEHD data shows that, whether the data were designed or found, an understanding of contributions of individuals included or excluded from the sample is essential. Conclusions regarding fundamental trends in inequality depend upon these decisions.

3.2 Effects of Inactivity on Earnings Inequality

We used our household survey to confirm that the trends in the estimated employment-to-population ratio produced directly from the micro-data match those of the official estimates. In particular, the employment-to-population ratio fell during the Great Recession, and had not recovered through 2013.[38]

We document that around 30 million eligible workers per year had no earnings in the current

[36] See Appendix Table C.2.
[37] See Appendix Figure C.5 and discussion.
[38] See Appendix Figure D.1.

year but positive earnings in at least one of the past four years.[39] Their treatment materially affects the distribution of earnings in any given year. It also materially affects the cyclical features of inequality measures such as the Gini coefficient.[40] Additional results and discussion are in Appendix D.2.

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.

4 Decomposing Changes in the Earnings Distribution

4.1 Evolution of the Earnings/Inactivity Distribution

Beginning in 2004 our eligible-workers frame is complete, including all states, DC, and the federal government. The adventitious timing of the start of the complete-data period presents us with the opportunity to study the evolution of the earnings distribution and the dynamics of non-employment during three distinct epochs of labor market conditions. According to the NBER, the Great Recession began in December 2007 and ended in June 2009. Applying this definition to annual data we have a pre-recession epoch spanning 2004-2007, a recession epoch running from 2008-2009, and a post-recession epoch beginning in 2010 and ending in 2013.

As we did in Section 2.4, we simplify the earnings distribution by assigning each active worker to one of three earnings categories (bottom, middle and top) and assign inactive, but eligible workers to a fourth category. Using the estimated annual earnings/inactivity distributions, we start by comparing the distribution in 2005 with the distribution in 2004, repeating this process for each subsequent year until 2013. Each year, the earnings/inactivity distribution may change relative to the previous year. The extent of this change depends on the number of workers entering and exiting the eligible-workers frame; the number of workers moving between earnings/inactivity categories; and changes in average earnings within each category.[41]

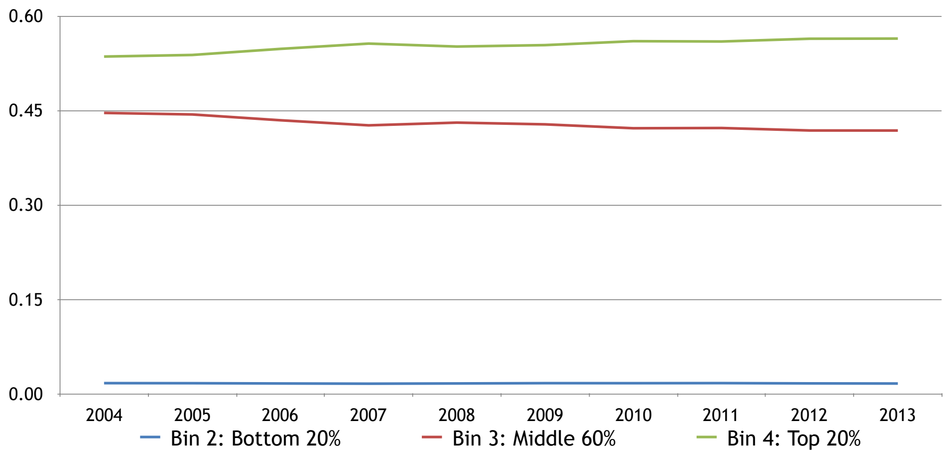
From 2004 through 2013 there is relatively strong growth in the number of eligible workers, averaging about 1% per year through 2009, declining to about half that rate after 2009, although growth within each category is uneven. For example, the largest growth in the number of workers occurs in the among those eligible to work but inactive (i.e., no earnings in the indicated year), a category with a growth rate almost twice that of any other group. The growth rate of workers in the bottom and middle earnings categories is less than half the overall growth rate. When we examine total earnings, most of the growth is found in the top 20% of the earnings distribution, with relatively little growth in the bottom 80%. Given that average earnings are falling for the

[39] See Appendix Table D.1.

[40] See Appendix Figures D.2-D.5.

[41] The distribution of eligible workers across the four earnings/inactivity categories by year is shown in Appendix Table E.1. The table also shows the total earnings per year, average earnings per year (total earnings/eligible workers with positive earnings). The last panel shows the cumulative change over the entire period for each of the three previous panels.

Figure 6: Share of Total Earnings in each Earnings Category



Notes: The figure plots the share of total earnings in each earnings category by year for the following categories: (i) *Bin 2*: the bottom 20% of the earnings distribution in blue, (ii) *Bin 3*: the middle 60% of the earnings distribution in red, and (iii) *Bin 4*: the top 20% of the earnings distribution in green.

bottom 80% of workers, it should not be surprising that total earnings for this group fails to keep pace with the growth in the number of workers. The situation for workers at the top, however, is much brighter: the top 20% of workers have relatively strong earnings growth of 4.4% over the period, resulting in growth in total earnings (12.2%) that outpaces the growth in the number of workers (7.8%).

The relatively high growth in both the number of eligible workers and average earnings in the top 20% of the earnings distribution has a strong effect on the distribution of total earnings. Figure 6 shows the share of total earnings attributed to each earnings category by year. We see a large amount of earnings inequality: earnings for the top 20% of workers are greater than both the bottom and middle combined, with the relative share of the top increasing almost continuously except for a brief pause in 2008 during the height of the Great Recession. We can also see the declining share of income accruing to the middle 60% of workers. Although the number of workers in the middle recovered after the Great Recession, average earnings continued to decline while higher growth in the number of workers at the bottom and top resulted in a declining share of earnings for workers in the middle 60% of the earnings distribution.

4.2 Earnings/Inactivity Distribution: Decomposition of the Changes

In the previous section we discussed the changing structure of the earnings/inactivity distribution around the time of the Great Recession. Several key facts stand out. First, there is enormous growth in the number of eligible workers with no reported earnings from 83,200,954 in 2004 to 96,959,047 in 2010, and still 96,151,327 in 2013. Second, average earnings stagnate (bottom 20%) or decline (middle 60%) for active workers. Third, the growth in the share of earnings accruing to the top 20% results from growth in the number of workers and average earnings in that category.[42]

In order to better understand these changes, we turn our focus to the flows of workers moving between active and inactive status as well as between different earnings categories. When interpreting these results, we implicitly assume average earnings within each category are stable between 2004-2013. Although changes in average earnings have a role to play, the data suggest that we focus on the worker flows because the percentage change in the number of workers dominates the change in average earnings for each category.[43]

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). Appendix Section E.1 provides the derivation of the flow accounting that we summarize here.

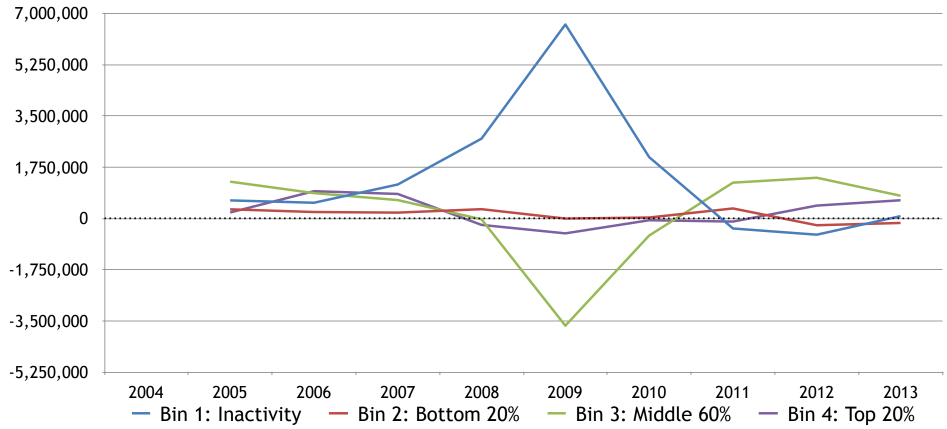
Figure 7 shows the change in the number of workers for each earnings/inactivity category by year. The data used to calculate the year-to-year differences can be found in the first panel of Appendix Table E.1. Each line in the graph represents the year-to-year change in the number of workers for one of the four earnings/inactivity categories. For example, in 2009 there were 94,864,949 eligible workers with no reported UI earnings while in 2008 there were only 88,245,425 workers in the same category, resulting in an increase of 6,619,524 workers. This increase in the number of eligible but inactive workers occurred during the height of the Great Recession and dwarfs the change in any other period, even the relatively large increases in the preceding and subsequent years. The area between the inactive line (blue) and the x-axis represents the cumulative increase in the number of eligible workers with no earnings over the entire period of 12,950,373. We can clearly see that the recovery has largely failed to reduce the number of eligible workers with no earnings through 2013. In contrast, the middle 60% (green line) faced a large reduction in numbers during the Great Recession, but, unlike the inactive category, the number of workers in the middle has returned to pre-recession levels, as reflected by the area between middle 60% line and the x-axis.

Table 5 presents an overview of the flow analysis, flows we will decompose further using the transition count matrix. The Count $(t-1)$, Count (t), and Net Change columns are shown in Appendix Table E.1 and Figure 7. They are reproduced here for comparison with sums of the transition counts. The stayers (i.e., c_{22}) are also included, and while they do not directly affect the net change in the flows, they represent the number of workers who remain in a given category for at least two years–giving an indication of the earnings stability of the typical worker. The outflows, inflows, and net change columns show the results of using Appendix equation E-1. The difference between the inflows and the outflows equals the net change, which should also equal the difference in counts between the current and the previous period. For example, returning to

[42] See Appendix Table E.1.

[43] When examining only the flows, we also implicitly assume average earnings are the same for each flow. Although this assumption is false it does not affect trends, since average earnings for a given flow are typically stable over time. It does, however, affect the scale or magnitude of each flow relative to another flow. Additional earnings change results and discussion, including the average earnings for each flow, are shown in Appendix E.

Figure 7: Year-to-Year Change in the Number of Workers in Each Earnings Bin



Notes: The estimates are based on the authors' calculations using year-to-year changes in the distribution of inactive and active workers among the activity/earnings bins.

the eligible workers with no earnings in 2009, we can see that the inflows were 16,166,420 and the outflows were 9,546,896. The large increase in the number of workers in this category was due primarily to a large increase in the inflows relative to the previous year and a small decrease in the number of outflows. That is, a relatively large number of workers who had a job in the previous year were unable to find an employer in the current year, while a relatively small number of workers without a job in the previous year were able to find one in the current year or moved out of the eligible-workers frame.

The table also shows a relatively large increase in the number of eligible workers with no earnings for at least two years over the entire period (stayers in the "Eligible, but no Reported UI Earnings" Panel). The change in the number of stayers equals the difference between the inflows in the previous period (the candidates to become stayers) minus the outflows in the current period. For example, in 2008 there were 13,271,459 eligible workers with at least one year without reported earnings (inflows), and 9,546,896 of these workers transitioned to another category in 2009 (outflows), resulting in an increase in the stayers between 2008 and 2009 of 3,724,563 workers. The stayers are useful for understanding the short term (two-year) volatility differences between each of the categories. For example, the bottom 20% of the UI earnings distribution has relatively few stayers compared to the middle and the top, consistent with the results presented earlier that most of these jobs are of relatively short duration. The results also imply that a large number of workers in the bottom 20% of the earnings distribution only stay there for a year or two before moving to another category, frequently inactivity.

Table 6 presents demographic characteristics for the 24 possible year-to-year transitions, excluding workers not eligible to work in both year $t-1$ and year t. The transitions labeled 0 represent workers moving into or out of the eligible-workers frame. The workers moving into the eligible-worker frame are typically young and predominately non-white, while the workers leaving the eligible workforce are typically older (60 plus years of age) and predominately white. One

Table 5: Flows into and out of Each Earnings Category

Year	Count $t-1$	Count t	Net Change	Stayers	Outflows	Inflows	Net Change
Exit from and Entry into Eligible Worker Status (Flows into and out of the Frame)							
2005	—	—	—	—	2,887,568	5,284,188	2,396,620
2006	—	—	—	—	2,940,260	5,501,749	2,561,489
2007	—	—	—	—	2,961,960	5,793,394	2,831,434
2008	—	—	—	—	3,048,753	5,850,756	2,802,003
2009	—	—	—	—	3,175,258	5,633,556	2,458,298
2010	—	—	—	—	3,189,382	4,680,774	1,491,392
2011	—	—	—	—	3,299,529	4,424,821	1,125,292
2012	—	—	—	—	3,296,122	4,350,437	1,054,315
2013	—	—	—	—	2,924,738	4,257,364	1,332,626
Eligible, but no Reported UI Earnings							
2005	83,200,954	83,819,319	618,365	71,931,565	11,269,389	11,887,754	618,365
2006	83,819,319	84,357,718	538,399	72,513,714	11,305,605	11,844,004	538,399
2007	84,357,718	85,518,594	1,160,876	73,295,146	11,062,572	12,223,448	1,160,876
2008	85,518,594	88,245,425	2,726,831	74,973,966	10,544,628	13,271,459	2,726,831
2009	88,245,425	94,864,949	6,619,524	78,698,529	9,546,896	16,166,420	6,619,524
2010	94,864,949	96,959,047	2,094,098	82,548,100	12,316,849	14,410,947	2,094,098
2011	96,959,047	96,619,700	-339,347	83,573,226	13,385,821	13,046,474	-339,347
2012	96,619,700	96,068,987	-550,713	83,628,961	12,990,739	12,440,026	-550,713
2013	96,068,987	96,151,327	82,340	83,990,110	12,078,877	12,161,217	82,340
Bottom 20% of the Overall UI Earnings Distribution							
2005	27,062,314	27,376,301	313,987	12,712,348	14,349,966	14,663,953	313,987
2006	27,376,301	27,598,826	222,525	12,919,731	14,456,570	14,679,095	222,525
2007	27,598,826	27,800,774	201,948	13,055,172	14,543,654	14,745,602	201,948
2008	27,800,774	28,120,283	319,509	13,270,031	14,530,743	14,850,252	319,509
2009	28,120,283	28,119,169	-1,114	13,215,490	14,904,793	14,903,679	-1,114
2010	28,119,169	28,154,014	34,845	13,057,840	15,061,329	15,096,174	34,845
2011	28,154,014	28,498,111	344,097	13,227,239	14,926,775	15,270,872	344,097
2012	28,498,111	28,269,636	-228,475	13,415,083	15,083,028	14,854,553	-228,475
2013	28,269,636	28,119,381	-150,255	13,437,328	14,832,308	14,682,053	-150,255
Middle 60% of the Overall UI Earnings Distribution							
2005	82,821,341	84,079,363	1,258,022	69,752,528	13,068,813	14,326,835	1,258,022
2006	84,079,363	84,946,369	867,006	70,696,052	13,383,311	14,250,317	867,006
2007	84,946,369	85,576,064	629,695	71,377,690	13,568,679	14,198,374	629,695
2008	85,576,064	85,548,690	-27,374	71,739,593	13,836,471	13,809,097	-27,374
2009	85,548,690	81,894,162	-3,654,528	69,594,276	15,954,414	12,299,886	-3,654,528
2010	81,894,162	81,314,722	-579,440	67,945,643	13,948,519	13,369,079	-579,440
2011	81,314,722	82,538,961	1,224,239	68,441,704	12,873,018	14,097,257	1,224,239
2012	82,538,961	83,930,862	1,391,901	69,837,520	12,701,441	14,093,342	1,391,901
2013	83,930,862	84,707,469	776,607	71,114,783	12,816,079	13,592,686	776,607
Top 20% of the Overall UI Earnings Distribution							
2005	26,678,860	26,885,106	206,246	22,942,722	3,736,138	3,942,384	206,246
2006	26,885,106	27,818,665	933,559	23,460,710	3,424,396	4,357,955	933,559
2007	27,818,665	28,657,580	838,915	24,260,307	3,558,358	4,397,273	838,915
2008	28,657,580	28,440,617	-216,963	24,629,289	4,028,291	3,811,328	-216,963
2009	28,440,617	27,935,033	-505,584	24,138,725	4,301,892	3,796,308	-505,584
2010	27,935,033	27,876,922	-58,111	24,278,404	3,656,629	3,598,518	-58,111
2011	27,876,922	27,773,225	-103,697	24,365,840	3,511,082	3,407,385	-103,697
2012	27,773,225	28,214,827	441,602	24,551,484	3,221,741	3,663,343	441,602
2013	28,214,827	28,838,761	623,934	25,057,962	3,156,865	3,780,799	623,934

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

Table 6: Demographic Characteristics of Workers by Transition Type

Transitions *Origin*	*Destination*	*Flow*	**Age** (*Avg*)	**Male** (*Share*)	**White** (*Share*)	**Born US** (*Share*)
Ineligibility	Inactivity:	0_1	21	0.521	0.221	0.710
	Bottom 20%:	0_2	19	0.481	0.336	0.886
	Middle 60%:	0_3	21	0.560	0.328	0.730
	Top 20%:	0_4	33	0.791	0.279	0.127
Inactivity	Ineligibility:	1_0	68	0.493	0.743	0.795
	Inactivity:	1_1	48	0.504	0.585	0.690
	Bottom 20%:	1_2	35	0.478	0.520	0.825
	Middle 60%:	1_3	38	0.580	0.582	0.765
	Top 20%:	1_4	44	0.730	0.701	0.737
Bottom 20%	Ineligibility:	2_0	64	0.525	0.784	0.888
	Inactivity:	2_1	39	0.484	0.576	0.831
	Bottom 20%:	2_2	34	0.421	0.566	0.891
	Middle 60%:	2_3	33	0.464	0.556	0.850
	Top 20%:	2_4	41	0.667	0.680	0.778
Middle 60%	Ineligibility:	3_0	66	0.560	0.785	0.858
	Inactivity:	3_1	45	0.548	0.659	0.806
	Bottom 20%:	3_2	39	0.468	0.604	0.846
	Middle 60%:	3_3	41	0.465	0.659	0.844
	Top 20%:	3_4	41	0.608	0.723	0.842
Top 20%	Ineligibility:	4_0	68	0.768	0.861	0.855
	Inactivity:	4_1	51	0.699	0.771	0.821
	Bottom 20%:	4_2	51	0.659	0.759	0.859
	Middle 60%:	4_3	46	0.626	0.751	0.861
	Top 20%:	4_4	47	0.667	0.776	0.855

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

interesting transition group is the 0_4 (workers not eligible in $t-1$ who transition to the top 20% of the earnings distribution in t). These workers are predominately older than other newly eligible workers, male, non-white, and overwhelmingly not born in the U.S. The remaining transitions have roughly similar characteristics, although older male workers are generally more prevalent in transitions associated with bin 4 (the top 20%), while female workers are more likely to be associated with transitions with bin 2 (the bottom 20%).

Next, we decompose the inflows and outflows further by using the transition matrix of counts. Figures 8 through 12 show the outflows and inflows for each earnings/inactivity category and the relevant transition counts by year. In particular, each year, an individual can be in one of the following five employment states:

0. *Ineligibility*: did not meet the requirements to be in the eligible-workers frame
1. *Inactivity*: part of the eligible-workers frame, but did not report positive UI earnings
2. *Bottom 20%*: annual UI earnings in the bottom 20% of the earnings distribution
3. *Middle 60%*: annual UI earnings in the middle 60% of the earnings distribution

4. *Top 20%*: annual UI earnings in the top 20% of the earnings distribution

We remind the reader that the cutoffs for the real earnings distribution bins are based on the pooled years 2004-2013, and do not change over time. It is possible for substantially more or less than 20% of current earners to be in the top 20% bin, for example, in any given year.

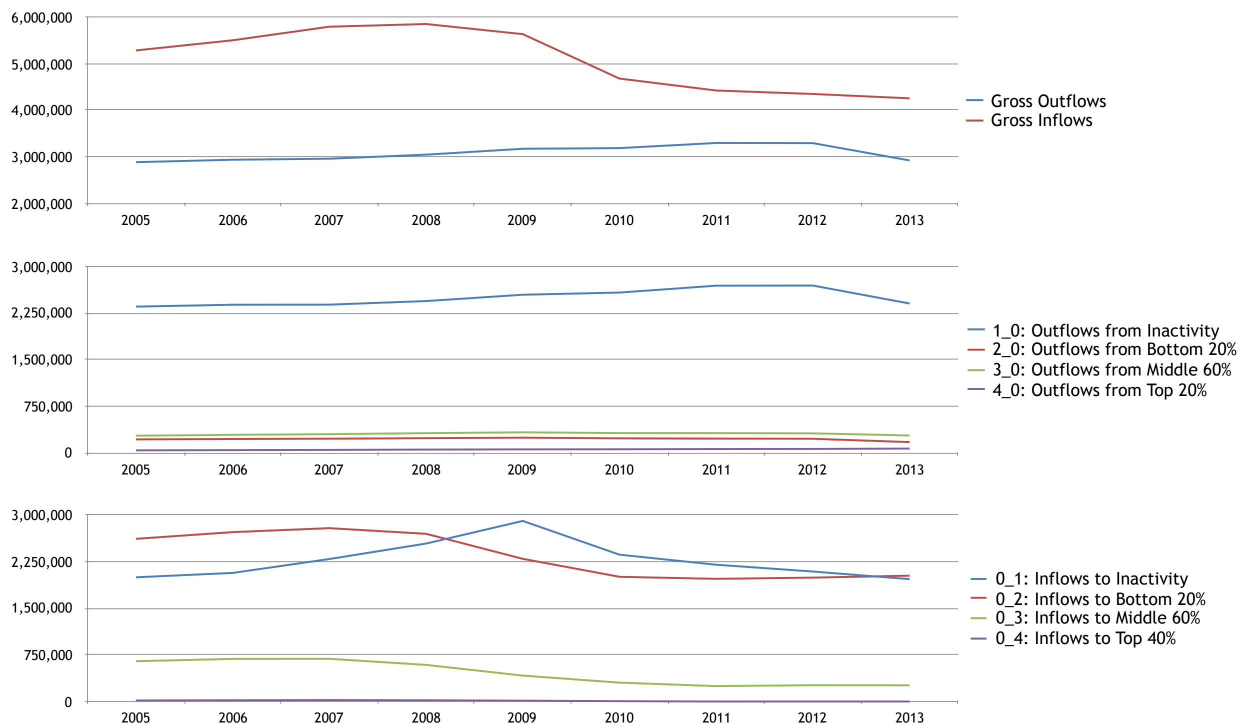
The first subplot in each figure shows the gross outflows (blue line) and the gross inflows (red line) into an earnings/inactivity category. The difference between the two lines is the net inflows presented in the last column of Table 5. The second subplot shows the transitions to other categories associated with the gross outflows line in the first graph. Note that the sum of each series in the second subplot is equal to the gross outflows line in the first subplot. The third subplot shows the transition counts associated with gross inflows. Similarly, the sum of each of the series in the third subplot is equal to the gross inflows line in the first subplot.

Figure 8 plots the counts of workers moving out of the eligible-workers frame and those moving into the frame. Notice that gross inflows into eligibility are always greater than gross outflows from eligibility. Thus, net inflows into eligibility are always positive, but decline and fail to recover after the Great Recession. This decline is primarily due to a decrease of inflows into eligibility, although there is a small increase in outflows as well. Notice that the outflows from eligibility come primarily from inactivity. Recall that inactive individuals are part of the eligible-workers frame but have no reported earnings. The first row of the second panel in Table 6 shows that inactive workers moving to ineligibility (1_0) tend to be older with an average age of 68. The inflows into eligibility tend to be young workers around the age of 20 moving into either inactivity or the bottom 20% of the earnings distribution. The exception are individuals moving into the top 20% of the earnings distribution (0_4), who tend to be older (average age is 33 years old), male (79.1% are male), and foreign born (12.7% are born in the U.S.). Prior to the Great Recession, on average, 38.3% of workers moving into eligibility went into inactivity and 49.1% went into the bottom of the earnings distribution. This flipped during the Great Recession, with 47.5% of newly eligible workers moving into inactivity and 43.4% moving into the bottom of the earnings distribution.

Figure 9 plots the flows into and out of inactivity. Net inflows are generally close to zero, except during the Great Recession when they became very positive. This spike in net inflows into inactivity was driven by both a large increase in inflows and a substantial decline in outflows. The increase in inflows is seen for every category except for the top of the distribution. The decrease in outflows during the Great Recession is primarily due to a reduction in workers moving to jobs in the bottom and middle of the earnings distribution, however these flows return to their pre-recession levels fairly quickly. The net result is a roughly symmetric increase in gross outflows and decrease in gross inflows. Without either a very large relative increase in gross outflows or a relatively large decrease in gross inflows, little progress can be made towards reducing the almost 13 million person increase in the number of eligible workers with no reported earnings during the Great Recession.

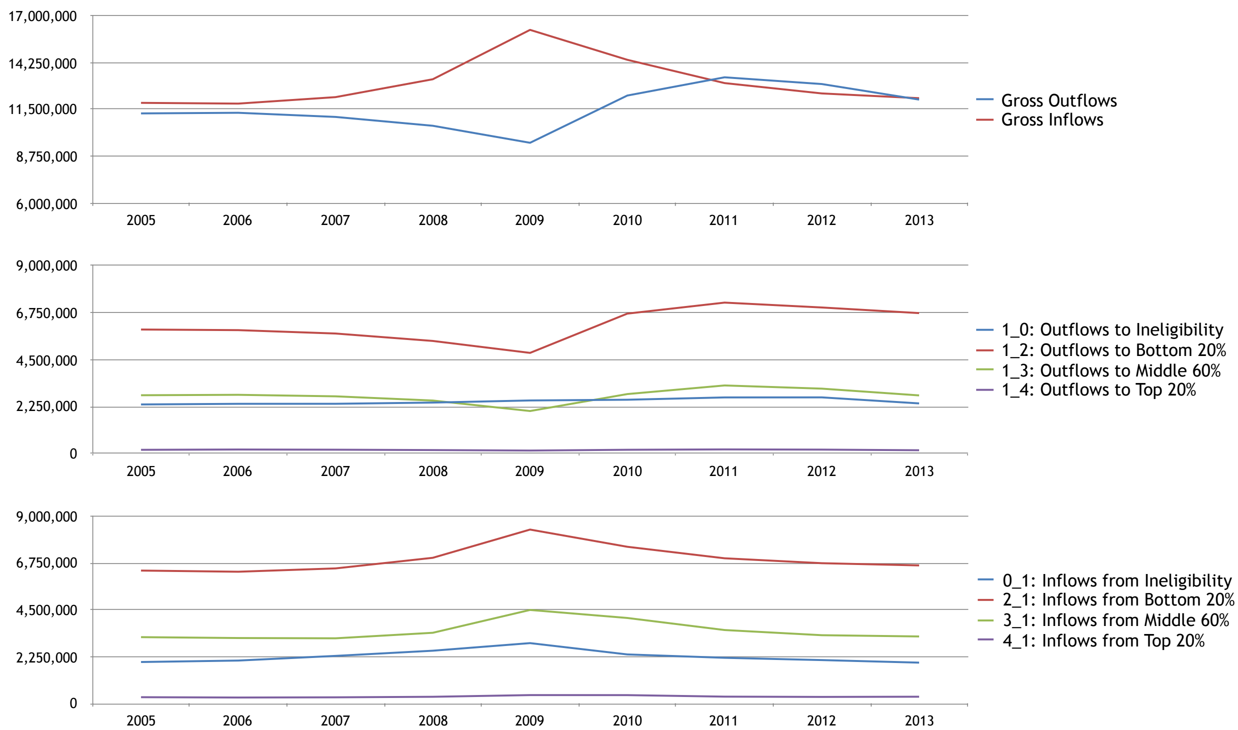
Figure 10 plots the flows into and out of the bottom 20% of the earnings distribution. Compared to some of the other categories, the counts for the bottom 20% are relatively stable.

Figure 8: Flows into and out of the Eligible-Workers Frame



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

Figure 9: Flows into and out of Inactivity



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

Figure 10: Flows into and out of the Bottom 20% of the Earnings Distribution

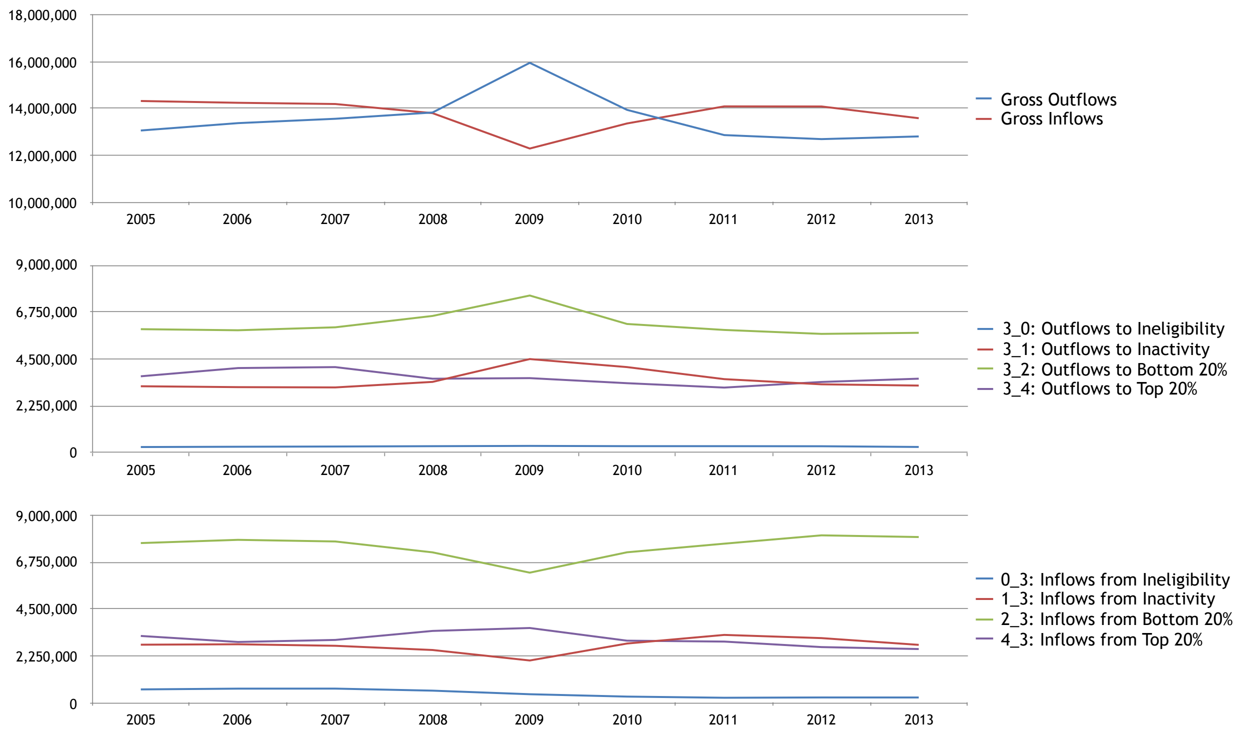
Notes: The estimates are based on the authors' calculations using transitions into and out of the eligible-workers frame used to construct the earnings distributions, including inactive workers, and transitions between the earnings categories.

The transitions show large changes occurring during the Great Recession. Outflows increased to nonemployment, while outflows to the middle of the earnings distribution fell (fewer workers moving up). At the same time inflows from nonemployment decreased and inflows from the middle of the distribution increased. Workers who moved to nonemployment are being replaced (not at the job level, but in terms of earnings) with workers from the middle of the distribution.

Figure 11 shows the flows into and out of the middle 60% of the earnings distribution. There is a large decrease in net inflows in 2009. This is largely due to an increase in the outflows of workers to the bottom 20% and nonemployment and a decrease in workers moving up from the bottom 20%. Although net inflows turn positive again after the Great Recession, these inflows are not large enough to halt the decreasing share of earnings accruing to workers in the middle 60%. There has also been a decline in workers moving from the middle to the top 20%, which peaked in 2007, implying a decrease in upward earnings mobility. A decrease in workers moving from the top to the bottom is also present, implying a decrease in downward earnings mobility. Post-recession, workers in the middle are more likely to stay in the middle and workers at the top are more likely to stay at the top.

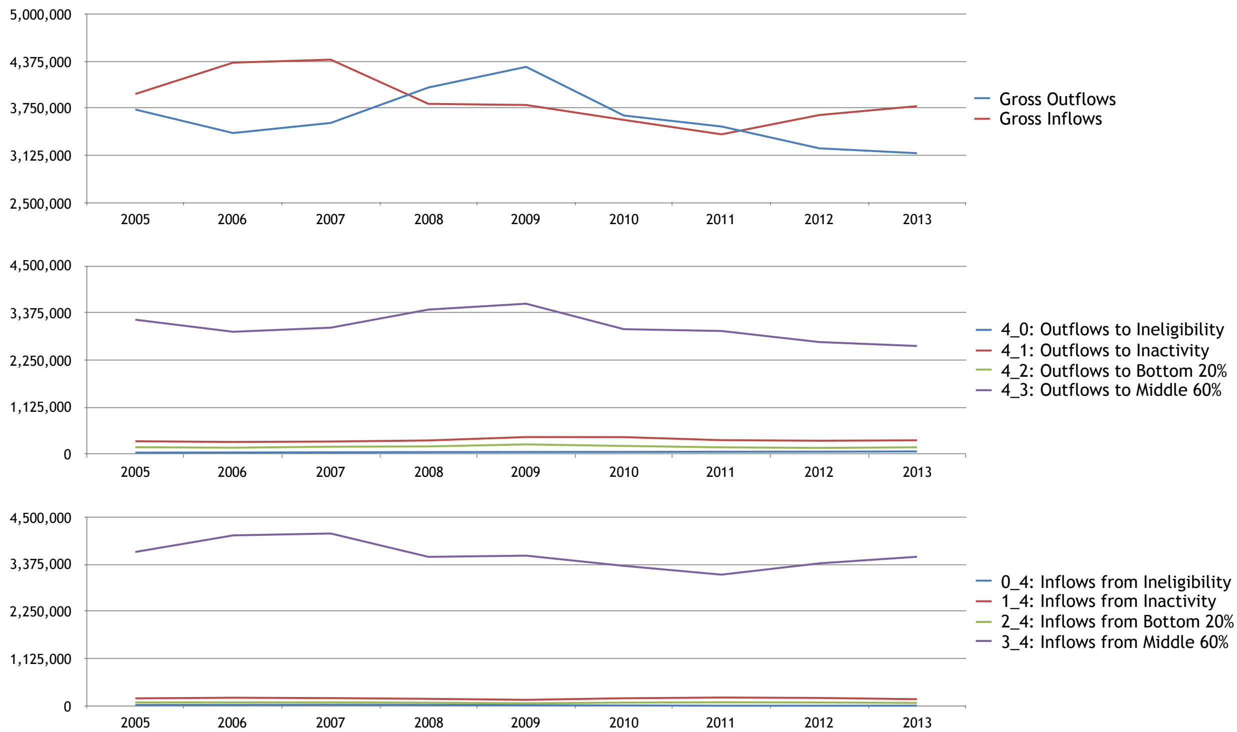
Figure 12 plots the flows into and out of the top 20% of the earnings distribution. Notice that there is a strong net inflow of workers to the top 20% prior to the Great Recession and a

Figure 11: Flows into and out of the Middle 60% of the Earnings Distribution



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

Figure 12: Flows into and out of the Top 20% of the Earnings Distribution



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

decrease during the recession. Although net inflows turn positive again in 2012, they do not return to the heights seen prior to the Great Recession. Workers at the top are relatively disconnected from the rest of the earnings distribution. The only substantial flows are to and from the middle, but the magnitudes of both of these flows appear to be declining post recession.[44]

Overall, mobility occurs most often between neighboring parts of the earnings/inactivity distribution. It is relatively rare to jump more than one earnings/inactivity category. For example, moving from the bottom to the top is a relatively rare event, while moving from the bottom to the middle is a common transition.[45]

[44] The disconnect is likely even greater than it may appear, especially for workers at the very top of the earnings distribution. Most of the workers moving from the bottom 20% to the top 20% and vice-versa have earnings near the minimum value of the top earnings bin, suggesting that most of the transitions may be associated with small earnings changes than one might infer from the average earnings in each bin.

[45] Appendix Figures E.1-E.3 repeat the analysis shown in Figures 10-12 using earnings changes instead of counts. The conclusions are essentially unchanged.

5 Firm Differences in Worker Earnings and Mobility

5.1 The Worker-Firm Earnings Decomposition

Having established that the eligible-workers frame is more likely to be representative of the entire U.S. labor market than uncorrected administrative record frames, and having established that it is complete and suitable for studying changes in the entire earnings distribution, including the movements into and out of activity, we now turn to the role of the employer as a source of earnings inequality. We use the statistical approach introduced to the labor economics literature in AKM.

To understand the role of firms in the rise in earnings inequality we estimate the following AKM model for the eligible-worker sample:

$$\ln y_{ijt} = x_{it}\beta + \theta_i + \psi_j + \varepsilon_{ijt}, \tag{3}$$

where y_{ijt} is real annual earnings for worker $i = 1$, ..., $200,665,944$, employed at firm $j = 1$, ..., $14,645,104$ in year $t = 2004$, ..., 2013.[46] On the right hand side we include θ_i, the fixed person effect; ψ_j, the fixed firm effect; and the vector x_{it}, which includes a constant, demographic characteristics interacted with actual labor-force experience, labor-force attachment variables, date-regime variables, and aggregate labor market conditions.

During the process of converting job-level firm-effect estimates to person-level firm-effect estimates, we move back and forth between dollars (levels) and logarithms as appropriate. We estimate equation (3) in semi-logarithmic form due to the approximately lognormal conditional distribution of the dependent variable. However, a semi-log specification returns estimated firm effects showing the log points of earnings attributable to the employer (approximately proportions for small effects). In order to combine the estimated job-level firm effects for workers who have multiple employers during the year into a single person-level firm effect the relative firm effects must be converted to dollar values. For example, suppose we have a worker who earns \$10,000 at job 1 and \$50,000 at job 2, with fixed firm effects of 0.2 and 0.1, respectively, in logs. Although the estimated firm effect is twice as large in job 1, the earnings are only 20% of the earnings in job 2. To account for the earnings differences across jobs, we convert the estimated firm effects to dollars, aggregate the dollar amounts, and then convert the dollar value of the person and firm components back to logs.

First, we isolate the real dollar value of the firm component of earnings for a given job in a given year. Specifically, for individual i, working at firm j in year t, the firm component of earnings y_{ijt} is defined as:

$$y_{ijt}^{\text{firm}} = y_{ijt} - \exp\left(\ln y_{ijt} - \widehat{\psi_j}\right). \tag{4}$$

[46] This is a slight abuse of notation. Two billion person-firm-year observations were used in the estimation from 826 million jobs held by 201 million persons at 14.7 million firms. Unlike the standard AKM approach of using only the employer with the highest earnings per worker per year, we use all jobs. The i, j, t subscripting is standing in for a more complicated notation to indicate multiple employers in a particular year. The model is fit to the 2004-2013 period, where the eligible-worker frame is complete for the entire U.S. labor market. See Appendix E.2 for further details.

We define the non-firm component of earnings y_{ijt} as the residual from the above equation:

$$y_{ijt}^{\text{non-firm}} = y_{ijt} - y_{ijt}^{\text{firm}}. \tag{5}$$

Aggregating across jobs for each worker gives a decomposition of total annual earnings into two person-level components:

$$\begin{aligned} \sum_j y_{ijt} &= \sum_j y_{ijt}^{\text{firm}} + \sum_j y_{ijt}^{\text{non-firm}} \\ y_{it}^{\text{total}} &= y_{it}^{\text{firm}} + y_{it}^{\text{non-firm}}. \end{aligned} \tag{6}$$

The person-level log firm component is recovered by taking the difference between the log of person-level earnings and the log of the non-firm component:

$$\ln y_{it}^{\text{firm}} = \ln y_{it}^{\text{total}} - \ln y_{it}^{\text{non-firm}}. \tag{7}$$

Continuing the example above, the total earnings for the hypothetical worker are \$60,000. After applying equations (4), (5), and (6) the estimated dollar value of the firm component across all jobs is \$6,571 and the estimated non-firm component is \$53,429. Applying equation (7) results in a log firm component of 0.116. Conceptually, the resulting person-level firm component is very similar to taking the earnings-weighted average of the log firm components.

We also extract a component of log earnings that we can associate with the worker's skill type. This component consists of the constant, which has been standardized to a year with average unemployment, and full-year full-time work, the estimated effects associated with labor force experience, and the estimated person effect. Thus the skill component can represent the log earnings associated with the annual wage rate for a worker with a given person effect and labor experience.[47]

We calculated both dollar and log estimates of the firm and non-firm earnings components. We calculated only the log component for estimating the skill type. We used the skill component, which is logarithmic, to classify workers by bottom 20%, middle 60% and top 20% of the skill distribution. When referencing the discrete distribution of the skill component, we refer to skill types. When referencing the value, we refer to the skill component of log earnings.

We used the log firm and log non-firm components to similarly classify firms and the non-firm contribution to earnings in bottom, middle and top bins. When referencing the discrete distribution of the firm component, we refer to firm types. We do not classify the cells of the discrete distribution of the non-firm component of earnings. In all cases, distributions were estimated using the pooled person-level observations over the 2004-2013 period. See Table 7 for the minimum, maximum, and average log values of all three components.

Our approach has two main benefits. First, all workers at a firm with only one job during

[47] We included all the effects labeled actual labor-force experience in Appendix Table E.4 using the coefficients in Appendix Table E.5.

Table 7: Statistics for Firm, Non-Firm and Skill Bins

	Firm-Type Bins		
Log Firm	**2**: *Bottom 20%*	**3**: *Middle 60%*	**4**: *Top 20%*
Minimum	-0.945	-0.374	0.556
Mean	-0.716	0.113	0.807
Maximum	-0.374	0.556	13.190
	Non-Firm-Type Bins		
Log Non-Firm	**2**: *Bottom 20%*	**3**: *Middle 60%*	**4**: *Top 20%*
Minimum	-4.499	8.922	10.480
Mean	7.708	9.803	10.980
Maximum	8.922	10.480	19.590
	Skill-Type Bins		
Log Skill	**2**: *Bottom 20%*	**3**: *Middle 60%*	**4**: *Top 20%*
Minimum	-16.960	6.255	7.203
Mean	5.898	6.695	7.679
Maximum	6.255	7.203	17.460

Notes: The estimates are based on the authors' calculations using the decomposition of person-level log earnings into firm non-firm components that sum to log total earnings. Firm-type and non-firm-type bins are formed for each component separately using the logarithmic scale. The skill component, also on a logarithmic scale, uses only the constant, person effect and actual labor-force experience effects as the basis for the skill-type bins. Distributions are based on the pooled person-level observations for the eligible-workers frame from 2004-2013. Statistics are rounded to four significant digits.

the year are placed in the same firm-type bin. Second, the total earnings of the worker do not affect the firm-type bin assignment. Third, classifying the workers by their skill bin, rather than the non-firm bin, controls for the state of the labor market and labor force attachment as well as eliminating the influence of the AKM residual. Using the log values of each component also allows us to study all possible mixes of worker skill types. If we had used the dollar-value bin assignments, the highest-paid workers would have dominated the top and bottom categories for each estimated component. For example, a worker with a very small log firm effect, but high earnings, would likely dominate a low earning worker with a large log firm effect.

Comparing the earnings bins in Table 2 with the firm and non-firm component bins in Table 7, notice that, except for a relatively small number of extreme values, the distribution of log earnings and log non-firm earnings component are similar within bin. This result is due to the small relative magnitude of the log firm component.[48] In spite of their relatively small magnitude in logs, the firm components can have a substantial effect, conditional on the size of the non-firm component of earnings. The log firm component is about -0.716 for the typical firm in the bottom of the firm compensation distribution, 0.113 in the middle and 0.807 at the top. All statistics are worker duration-weighted averages, implying that a worker at the middle-type firm receives about 11% more than would be expected given the worker's characteristics. In comparison, the

[48]The non-firm component here includes the constant, person effect θ_i, index $(x_{it}\beta)$, and ε_{ijt}. The estimated constant is equal to 6.01, and the average value of the person effect and residual, across all observations in the estimation sample, are both zero. The average experience component is 0.57, the labor force attachment component is 2.24, and the aggregate labor market conditions component is -0.045.

difference between the average log earnings value and the average log non-firm earnings component for workers in the middle is 0.135 ($= 9.938 - 9.803$), which is very similar to the 0.113 estimated for the worker duration-weighted average log firm component. Although this comparison is not a true worker-level comparison, it should be similar given that most workers are in the same non-firm-type and overall-earnings bins.

There is some evidence of worker sorting. Taking the difference between the average log earnings for each overall-earnings bin (Table 2) and the average log non-firm components for the same non-firm bin (Table 2) shows that workers at the bottom of the overall-earnings distribution and the bottom of the non-firm component distribution tend to work in lower-paying firms ($7.473 - 7.708 = -0.235$) compared with workers at the top ($11.236 - 10.976 = 0.260$).

Consider the magnitude of the average firm effect for each firm type and its potential effect on workers in different parts of the non-firm component distribution. For example, the typical worker at the bottom of the non-firm component distribution has average log earnings of 7.708. If this worker were employed at a firm in the middle of the firm component distribution, his log earnings would be greater by about 0.113, which is not enough to move him to the middle of the overall-earnings distribution ($7.708 + 0.113 = 7.821 < 9.938$). Even if this worker were able to transition to a firm at the top of the firm component distribution, ceteris paribus, his log earnings would be greater by 0.694 ($= 0.807 - 0.113$) log points. The resulting log earnings of 8.515 ($= 7.708 + 0.807$) would still not be large enough to move the worker to the middle overall-earnings bin. In comparison, for a worker in the top of the non-firm component distribution, moving from a middle to a top firm, results in an earnings increase large enough for the worker to transition from the middle to the top of the overall-earnings distribution ($10.976 + 0.807 = 11.783 > 11.236$). Although the relative effect is the same, the dollar value of the effect of working at a high paying firm increases the greater is a worker's non-firm component of earnings.

Table 6 shows characteristics of workers associated with each of the 24 possible earnings and inactivity transitions. Given that the overwhelming majority of workers are in the same bin in the overall earning and non-firm component distributions, the characteristics of the workers in each corresponding non-firm component bin will largely be the same. In Table 8 we show the characteristics of firms across each of the three firm-type categories. There are clear differences in the industry distribution by where the firm lies in the firm component distribution. Low-paying firms are highly concentrated in "Trade, Transportation, & Utilities" and "Leisure and Hospitality," with over 50% of workers at low paying firms in these two industries. Firms in the middle of the pay distribution are not nearly as concentrated by industry, but nevertheless workers in these firms are prevalent in "Trade, Transportation, & Utilities," "Education and Health," and "Manufacturing." Somewhat surprisingly, except for "Leisure and Hospitality" and "Finance," the distribution of workers in high-paying firms across industries is relatively diffuse. Most industries have a substantial number of workers in high-paying firms, implying that, except for "Leisure and Hospitality," it isn't necessary to change industries to work at a high-paying firm. As found in other studies, low-paying firms tend to be both younger and smaller than high-paying firms (Haltiwanger

Table 8: Firm Characteristics by Position in the Firm Component Distribution

Characteristic	Bottom	Middle	Top
Industry Distribution (Percent)			
Natural Resources/Mining	0.012	0.011	0.023
Construction	0.012	0.060	0.075
Manufacturing	0.014	0.104	0.178
Trade, Trans, & Utilities	0.246	0.211	0.140
Information	0.017	0.016	0.050
Finance	0.017	0.057	0.119
Prof. and Bus. Services	0.118	0.123	0.212
Education and Health	0.209	0.276	0.074
Leisure and Hospitality	0.290	0.070	0.006
Other Services	0.048	0.031	0.021
State/Local Government	0.017	0.041	0.062
Federal Government	0.000	0.001	0.041
Firm Age			
Mean	20.988	22.700	24.760
Standard Deviation	9.670	9.419	9.882
P25	10	13	17
Median	24	27	29
P75	32	32	33
Firm Size			
P10	6	8	11
P25	26	40	118
Median	330	502	2,359
P75	9,433	9,088	20,991
P90	73,330	68,535	64,448

Notes: All statistics are calculated at the worker-year-job level with the value for each job weighted by (y_{ijt}/e_{it}) when forming the averages. $N = 2,014,000,000$. A firm is defined by the state-level unemployment insurance account number, called an SEIN in LEHD data. Firm age (measured in years) and firm size are based on the national firm definitions used in other LEHD data products like the Quarterly Workforce Indicators. (See Haltiwanger et al. (2012)).

et al., 2012, Figure 7).

5.2 Earnings and Mobility by Person and Firm Type

In this section we use the AKM decomposition to explore how the three types of workers (bottom, middle and top of the skill-type distribution) sort into the three types of firms (bottom, middle and top of the firm-type distribution). The results for each worker type are presented separately. Tables 9, 10 and 11 present outcomes for workers in the bottom, middle and top bins of the skill-type distribution, respectively.

The tables were created as follows. Bin types are based on the previous year's data; i.e., year $t-1$ classifications. 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 skill type of the worker for that year, which can vary over time as workers accumulate experience. Within each skill 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.[49] Within each of the skill-type × 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.[50]

To fix ideas, we will take a detailed look at two rows in Table 9. To be recorded in this table, the person must have been in the bottom (lowest) bin of the skill-type 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-type distribution in $t-1$. Persons in this row are also in the bottom bin of the overall-earnings distribution in year $t-1$, and 0.0782 is the share of such persons relative to those in the middle or top of the overall-earnings distribution. The flow labeled "2_0" is the movement from the bottom of the overall-earnings distribution (bin 2) to the ineligible (bin 0); that is, this is the flow out of the frame for persons at the bottom of the overall-earnings distribution. There were, on average, 39,565 such persons each previous year ($t-1$). They represent 0.7% of the flows from bin 2 of the overall-earnings distribution for low-skill workers in bottom-paying firms. Average earnings in $t-1$ were \$1,921 of which $-\$2,285$ are attributed to the firm component of our decomposition and \$4,207 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 skill

[49] The estimated AKM firm effects do not vary during the period, the value of the firm effect used in the firm-type bins depends upon all employers during the year, and actual earnings. Hence, these effects do change values even when workers do not change employers. Of course, the AKM firm effect changes when an individual changes employers as well.

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

component distribution in year $t-1$. Of all such persons, 60% are employed by a firm in the middle of the firm-type distribution. Of all persons at the bottom of the skill-type distribution and in the middle of the firm-type distribution in year $t-1$, the proportion 0.608 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, 8,507,780 in the 9 pairs of years for which the table was constructed. Those who stayed in the middle of the overall-earnings distribution represented 82.4% of all persons who were in the middle of the overall-earnings distribution, in the low-skill bin, and in a middle-paying firm in year $t-1$, on average. In year $t-1$, their earnings averaged $17,361 of which $2,337 are attributed to the firm component of our decomposition and $15,024 are attributed to the non-firm component of our decomposition. In the current year, year t, average earnings were $18,013 of which $2,619 are associated with the firm component and $15,394 are associated with the non-firm component.

We use these tables to investigate worker sorting directly by looking at the interaction of the skill 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 skill-type distribution an individual occupied, given his place in the overall-earnings distribution. This hypothesis is clearly not supported by the data, and forms the basis of our major conclusion that the influence of the firm operates through channels that are, at least in part, different from the channels that intermediate the skill-type effect. For example, again using Table 9 showing the bottom of the skill-type distribution, about 27% of the earnings observations are in firms at the bottom of the firm-type distribution, 60% are in firms of the middle type, and only 13% are in top firms. By comparison, Tables 10 and 11 show that persons in the middle and top of the skill-type distributions are much less likely to be employed at firms in the bottom type (17% and 21%, respectively), and much more likely to be employed at top firms (21% and 25%, respectively).

Focusing on each skill type, we start with the earnings observations for low-skill types in Table 9. For workers at the bottom of the skill-type distribution, working at a higher-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 skill-type and overall-earnings distributions has a probability of moving to the middle of the overall-earnings distribution of 16.8% at a low-paying firm, 24.2% at a middle-paying firm, and 23.3% at a high paying firm. Prior to the 2_3 transition, the average low-skill worker at a low, middle, and high-paying firm earns $3,398, $3,726, and $3,749, respectively.[51] After the transition, the average low-skill worker at a low, middle, and high-paying firm earns $10,487, $11,739, and $15,047, 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.

[51]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, weeks worked perhaps, the impact of working at a higher-paying firm would be much greater if the person component of earnings were the same across firm types.

Table 9: Earnings Associated with Flows by Firm Bin for Low-Skill Persons

					Previous Year			**Current Year**		
All Earnings	**Share**	**Flow**	**Average Count**	**Percent**	*Total*	*Firm*	*Non-Firm*	*Total*	*Firm*	*Non-Firm*
				Bottom Firm (27%)						
		2_0	39,565	0.7%	1,921	-2,285	4,207	—	—	—
		2_1	1,102,440	18.9%	1,014	-1,039	2,053	—	—	—
Bottom	0.782	2_2	3,713,490	63.6%	2,282	-2,492	4,774	2,562	-2,262	4,824
		2_3	978,706	16.8%	3,398	-3,236	6,634	10,487	-3,913	14,400
		2_4	1,326	0.0%	1,977	-4,694	6,672	80,726	25,927	54,800
		3_0	5,655	0.3%	10,686	-10,132	20,818	—	—	—
		3_1	32,834	2.0%	10,298	-9,484	19,782	—	—	—
Middle	0.217	3_2	380,336	23.5%	9,359	-8,313	17,673	3,771	-3,133	6,904
		3_3	1,194,533	73.9%	12,342	-10,417	22,760	13,660	-8,628	22,289
		3_4	2,705	0.2%	31,037	-32,962	63,999	58,165	-34,136	92,301
		4_0	20	0.3%	100,000	-130,000	230,000	—	—	—
		4_1	160	2.1%	66,408	-85,919	152,326	—	—	—
Top	0.001	4_2	362	4.6%	180,230	-163,472	343,702	2,486	-6,377	8,863
		4_3	2,151	27.6%	59,430	-67,891	127,321	33,563	-33,499	67,062
		4_4	5,106	65.5%	66,745	-74,157	140,902	67,457	-64,741	132,199
				Middle Firm (60%)						
		2_0	65,494	1.0%	2,447	2	2,445	—	—	—
		2_1	1,428,036	22.1%	1,749	1	1,748	—	—	—
Bottom	0.381	2_2	3,411,867	52.7%	2,781	-209	2,990	2,719	-505	3,225
		2_3	1,566,180	24.2%	3,726	-167	3,893	11,739	438	11,302
		2_4	2,684	0.0%	2,981	252	2,729	69,808	27,963	41,845
		3_0	46,878	0.5%	12,749	1,501	11,249	—	—	—
		3_1	310,560	3.0%	11,957	1,722	10,235	—	—	—
Middle	0.608	3_2	1,396,266	13.5%	11,989	725	11,263	3,367	-386	3,752
		3_3	8,507,780	82.4%	17,361	2,337	15,024	18,013	2,619	15,394
		3_4	60,049	0.6%	35,329	7,277	28,052	55,148	13,567	41,580
		4_0	289	0.2%	73,365	14,053	59,313	—	—	—
		4_1	1,586	0.9%	72,382	12,255	60,128	—	—	—
Top	0.011	4_2	5,942	3.2%	67,356	7,642	59,714	1,985	-454	2,439
		4_3	47,624	25.7%	55,476	10,714	44,761	34,993	6,711	28,282
		4_4	130,216	70.1%	62,989	12,797	50,193	63,549	13,399	50,150
				Top Firm (13%)						
		2_0	8,510	1.7%	2,456	1,367	1,089	—	—	—
		2_1	176,522	34.3%	2,131	1,203	928	—	—	—
Bottom	0.143	2_2	207,481	40.4%	2,979	1,642	1,336	2,711	920	1,791
		2_3	119,756	23.3%	3,749	2,039	1,710	15,047	6,912	8,135
		2_4	1,650	0.3%	3,051	1,820	1,232	109,852	63,556	46,296
		3_0	12,848	0.5%	18,789	10,749	8,040	—	—	—
		3_1	128,415	4.6%	16,324	9,505	6,820	—	—	—
Middle	0.771	3_2	209,939	7.6%	17,926	9,809	8,117	2,911	914	1,997
		3_3	2,351,400	84.6%	24,984	13,686	11,297	25,168	13,437	11,731
		3_4	76,365	2.7%	36,786	21,709	15,077	56,997	33,932	23,065
		4_0	677	0.2%	88,542	65,161	23,381	—	—	—
		4_1	4,682	1.5%	93,519	62,575	30,944	—	—	—
Top	0.087	4_2	11,856	3.8%	82,512	49,460	33,051	1,976	377	1,599
		4_3	63,883	20.4%	59,192	36,042	23,151	33,313	19,085	14,228
		4_4	232,083	74.1%	72,919	47,838	25,081	73,785	48,162	25,624

Notes: 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 10 supports a similar conclusion. Middle-skilled workers in the bottom of the overall-earnings distribution also have a greater chance of moving to the middle of the earnings distribution the higher the firm-type for which they work. When they transition, their current-year earnings will also be greater the higher is the firm-type for which they work. The vast majority (62%) of workers in the middle of the skill-type distribution are employed at middle-paying firms. The next most prevalent outcomes for such workers are employment at top- and bottom-paying firms, 21% and 17% respectively. 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-skill worker in the middle (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 \$15,820, a middle-skill worker in a middle-type firm has $t-1$ earnings of \$24,165, and a middle-skill worker at a top firm has earnings of \$31,965. Most of the difference is due to a larger firm effect, although the non-firm component declines as a middle-skill person is found in increasing firm types, giving back some of the gains. Another benefit of finding employment at a high-paying firm is a greater probability of moving to the top of the earnings distribution. For middle-skill workers in the middle bin of the overall-earnings distribution in the previous year, the relevant comparisons are as follows. The estimated probability of a 3_4 transition for workers in a low-paying firm is 0.6%. The estimated probability of the same transition for workers in a middle-paying firm is 2.6%. Finally, the estimated probability of a transition to the top of the overall-earnings distribution for middle-skill workers in a top-paying firm is 11.8%.

Table 10: Earnings Associated with Flows by Firm Bin for Middle-Skill Persons

					Previous Year			Current Year		
All Earnings	**Share**	**Flow**	**Average Count**	**Percent**	*Total*	*Firm*	*Non-Firm*	*Total*	*Firm*	*Non-Firm*
				Bottom Firm (17%)						
		2_0	36,376	0.6%	2,766	-4,188	6,953	—	—	—
		2_1	1,413,148	21.9%	1,828	-1,999	3,827	—	—	—
Bottom	0.444	2_2	3,101,345	48.1%	2,774	-3,688	6,462	3,056	-3,388	6,444
		2_3	1,894,531	29.4%	3,540	-3,808	7,348	11,709	-6,015	17,725
		2_4	6,629	0.1%	2,514	-4,166	6,680	71,419	15,598	55,821
		3_0	31,137	0.4%	12,442	-11,801	24,242	—	—	—
		3_1	212,893	2.7%	11,397	-10,634	22,031	—	—	—
Middle	0.547	3_2	983,862	12.4%	10,788	-10,662	21,450	3,785	-3,651	7,436
		3_3	6,669,365	84.0%	15,820	-14,127	29,946	16,803	-12,701	29,505
		3_4	45,779	0.6%	33,527	-32,569	66,096	57,183	-33,647	90,830
		4_0	195	0.2%	69,394	-89,283	158,677	—	—	—
Top	0.008	4_1	1,370	1.1%	74,112	-139,818	213,929	—	—	—
		4_2	1,950	1.6%	68,735	-88,619	157,355	2,844	-5,364	8,209
		4_3	30,542	24.9%	56,942	-59,886	116,828	35,086	-31,411	66,497
		4_4	88,637	72.2%	63,707	-64,986	128,693	64,695	-60,302	124,997
				Middle Firm (62%)						
		2_0	44,834	0.7%	3,037	27	3,010	—	—	—
		2_1	1,917,160	28.5%	2,653	10	2,642	—	—	—
Bottom	0.130	2_2	2,264,104	33.6%	3,077	-231	3,308	2,992	-803	3,795
		2_3	2,485,434	36.9%	3,706	-149	3,855	14,551	398	14,152
		2_4	19,469	0.3%	3,366	427	2,939	63,885	21,915	41,969
		3_0	124,637	0.3%	19,293	1,958	17,335	—	—	—
		3_1	1,502,717	3.6%	15,071	2,193	12,877	—	—	—
Middle	0.802	3_2	2,535,882	6.1%	15,344	1,061	14,283	3,442	-487	3,929
		3_3	36,185,391	87.4%	24,165	3,268	20,897	24,688	3,412	21,276
		3_4	1,075,334	2.6%	37,532	8,695	28,837	54,207	14,489	39,719
		4_0	4,246	0.1%	64,346	14,661	49,686	—	—	—
		4_1	20,143	0.6%	69,334	12,984	56,350	—	—	—
Top	0.068	4_2	32,524	0.9%	64,867	13,923	50,943	2,927	-123	3,050
		4_3	812,972	23.2%	54,592	12,472	42,120	36,813	8,389	28,424
		4_4	2,636,922	75.2%	61,625	14,644	46,981	62,511	15,362	47,149
				Top Firm (21%)						
		2_0	7,309	1.2%	3,258	1,763	1,495	—	—	—
		2_1	243,227	38.9%	3,027	1,642	1,385	—	—	—
Bottom	0.036	2_2	156,614	25.1%	3,382	1,834	1,548	3,114	625	2,489
		2_3	201,902	32.3%	3,825	2,020	1,804	19,512	7,872	11,640
		2_4	15,593	2.5%	3,681	2,053	1,628	64,011	35,748	28,263
		3_0	27,268	0.3%	25,229	13,304	11,925	—	—	—
		3_1	520,030	5.5%	20,967	11,452	9,514	—	—	—
Middle	0.547	3_2	350,961	3.7%	22,339	11,765	10,573	3,224	715	2,509
		3_3	7,391,254	78.7%	31,965	16,295	15,670	32,082	15,736	16,346
		3_4	1,107,423	11.8%	38,199	20,579	17,620	55,738	30,184	25,553
		4_0	9,556	0.1%	70,452	41,322	29,130	—	—	—
		4_1	70,437	1.0%	70,076	42,343	27,733	—	—	—
Top	0.417	4_2	61,125	0.9%	72,097	41,507	30,590	2,881	939	1,942
		4_3	899,132	12.6%	57,986	31,890	26,095	35,291	18,151	17,140
		4_4	6,115,576	85.5%	67,457	38,800	28,656	68,906	39,504	29,402

Notes: 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 11: Earnings Associated with Flows by Firm Bin for High-Skill Persons

					Previous Year			**Current Year**		
All Earnings	**Share**	**Flow**	**Average Count**	**Percent**	*Total*	*Firm*	*Non-Firm*	*Total*	*Firm*	*Non-Firm*
					Bottom Firm (21%)					
		2_0	10,534	1.8%	2,773	-8,174	10,948	—	—	—
		2_1	280,788	46.7%	2,249	-3,701	5,950	—	—	—
Bottom	0.104	2_2	176,404	29.3%	2,907	-11,164	14,071	2,982	-10,511	13,493
		2_3	128,775	21.4%	3,591	-7,554	11,145	15,617	-20,775	36,391
		2_4	5247	0.9%	2,815	-7,587	10,402	94,017	-43,216	137,233
		3_0	19,797	0.5%	20,605	-33,443	54,048	—	—	—
		3_1	160,196	4.2%	18,005	-23,558	41,563	—	—	—
Middle	0.655	3_2	129,284	3.4%	17,280	-29,586	46,865	3,576	-7,467	11,043
		3_3	3,287,634	87.0%	27,830	-35,576	63,405	28,061	-34,972	63,033
		3_4	181,857	4.8%	38,233	-46,714	84,946	57,343	-64,656	121,999
		4_0	4,283	0.3%	94,022	-135,507	229,529	—	—	—
		4_1	13,507	1.0%	95,299	-176,061	271,360	—	—	—
Top	0.241	4_2	6,970	0.5%	78,069	-144,049	222,118	3,081	-6,682	9,763
		4_3	159,557	11.5%	59,330	-78,709	138,039	36,446	-43,579	80,024
		4_4	1,205,661	86.7%	80,717	-112,912	193,629	81,938	-110,686	192,624
					Middle Firm (54%)					
		2_0	9,555	1.7%	3,485	58	3,427	—	—	—
		2_1	324,938	57.2%	3,172	40	3,132	—	—	—
Bottom	0.038	2_2	101,070	17.8%	3,231	-99	3,330	3,150	-770	3,921
		2_3	117,605	20.7%	3,745	-40	3,785	19,859	-1,244	21,103
		2_4	14,650	2.6%	3,638	441	3,198	76,591	14,464	62,127
		3_0	30,029	0.5%	25,036	572	24,463	—	—	—
		3_1	463,428	8.2%	20,864	1,649	19,215	—	—	—
Middle	0.378	3_2	150,873	2.7%	22,473	308	22,165	3,521	-658	4,179
		3_3	4,210,520	74.1%	33,479	-1,789	35,268	33,621	-2,337	35,958
		3_4	826,287	14.5%	37,799	3,338	34,461	59,210	6,969	52,240
		4_0	23,563	0.3%	113,825	18,140	95,686	—	—	—
		4_1	99,260	1.1%	94,122	17,584	76,539	—	—	—
Top	0.585	4_2	35,671	0.4%	79,444	11,857	67,587	3,374	-192	3,566
		4_3	769,334	8.7%	62,012	7,288	54,724	35,502	2,761	32,741
		4_4	7,870,823	89.5%	92,507	15,835	76,673	93,962	16,259	77,703
					Top Firm (25%)					
		2_0	1,600	1.8%	3,762	2,079	1,683	—	—	—
		2_1	49,575	56.4%	3,674	2,003	1,671	—	—	—
Bottom	0.012	2_2	13,208	15.0%	3,763	2,175	1,588	3,462	1,162	2,300
		2_3	16,128	18.4%	3,986	2,174	1,812	21,346	6,827	14,519
		2_4	7,364	8.4%	3,888	2,123	1,765	94,733	48,125	46,608
		3_0	6,924	1.0%	23,484	12,755	10,729	—	—	—
		3_1	172,259	25.1%	22,641	12,565	10,076	—	—	—
Middle	0.097	3_2	26,117	3.8%	23,274	12,578	10,696	3,412	633	2,779
		3_3	257,654	37.5%	28,736	15,256	13,480	28,673	12,434	16,239
		3_4	224,480	32.7%	32,774	17,590	15,184	76,747	39,976	36,771
		4_0	12,289	0.2%	171,799	97,735	74,064	—	—	—
		4_1	129,640	2.0%	145,870	87,655	58,215	—	—	—
Top	0.891	4_2	24,092	0.4%	130,394	74,667	55,727	3,317	928	2,389
		4_3	260,123	4.1%	90,264	50,200	40,064	30,378	14,237	16,141
		4_4	5,902,247	93.3%	143,296	82,412	60,884	147,043	83,785	63,257

Notes: 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 11 shows that most of the high-skill workers are also in the top of the overall-earnings distribution, since the top bin of the overall-earnings distribution shows shares of 0.241, 0.585 and 0.891 for low, middle and top-type firms, respectively. There is also a substantial minority in the middle of the overall-earnings distribution. Since transitions to the top of the overall-earnings distribution (3_4) are more likely at top-paying firms (32.7%) as compared to low- or middle-paying firms (4.8% and 14/5%, respectively), we note that once again working at such a firm offers an advantage distinct from the worker's skill-type. In the high-skill category, the earnings differences between working at a middle- compared to a bottom-type firm after making a 3_4 transition are relatively small, but the earnings gains from working at a top-paying firm are very large. While working at a top-paying firm is clearly preferred and the gains are large, a typical worker in any part of the skill distribution would also have a strong preference for working at a middle-paying rather than a bottom-paying firm. Although the dollar gains may be relatively small, the difference in the earnings for bottom- and middle-paying firms is significant. For example, 78% of the low-skill persons employed at a bottom-paying firm are at the bottom of the overall-earnings distribution, while only 38% of the the low-skill persons employed at a middle-paying firm are at the bottom of the overall-earnings distribution. Overall earnings within the bottom bin are not dramatically different in this case, but workers in the middle bin of the overall-earnings distribution have noticeably higher earnings at a middle-paying firm ($17,361) vs. a bottom-paying firm ($12,342). Somewhat surprisingly, there are a relatively large number of top-skill workers at bottom- and middle-type firms. On average, these workers, especially in the middle, are employed at worse-paying firms than middle skill-type workers.[52]

Table 12 analyzes the earnings of individuals who do not move in the earnings distribution between consecutive years by firm and skill type. This table allows us to analyze the potential effects of redistributing the skill types across firm types or redistributing the firm types across skill types. For the bottom and middle skill types, there is almost no advantage to being employed in a higher paying firm type, given their place in the earnings distribution. For example, a bottom-skill person in the top 20% of the earnings distribution earns $66,745 in a bottom-paying firm and $72,919 in a top-paying firm. But for a top-skill person, there is a big advantage to employment in a top-paying firm, $80,717 as compared to $143,736 when in the top 20% of the earnings distribution. The skill pay premium for all firms is approximately the same when considering a bottom-skill as opposed to a middle-skill worker. For example, $12,342 vs. $15,820 for bottom firm and $17,361 vs. $24,165 for a middle firm. This is also the case for considering a middle-skill vs. top-skill worker, except for the top-paying firms, where that premium is much greater than for middle or low-paying firms–$67,457 vs. $143,296 for the top-paying firm as compared to $61,625 vs. $92,507 for the middle-paying firm.

[52] See Appendix Section E.3 for an analysis of the changes in earnings inequality using only the firm-type and non-firm-type distributions. Some anomalies appear in that analysis that do not appear when we use only the skill-type to characterize worker differences. We thank a referee for suggesting this modification.

Table 12: Within Firm-Type and Skill-Type Inequality

		Bottom Skill		**Middle Skill**		**Top Skill**	
	Flow	*Earnings*	*Ratio*	*Earnings*	*Ratio*	*Earnings*	*Ratio*
Bottom Firm	2_2	2,282	0.185	2,774	0.175	2,907	0.104
	3_3	12,342	—	15,820	—	27,830	—
	4_4	66,745	5.408	63,707	4.027	80,717	2.900
Middle Firm	2_2	2,781	0.160	3,077	0.127	3,231	0.097
	3_3	17,361	—	24,165	—	33,479	—
	4_4	62,989	3.628	61,625	2.550	92,507	2.763
Top Firm	2_2	2,979	0.119	3,382	0.106	3,763	0.131
	3_3	24,984	—	31,965	—	28,736	—
	4_4	72,919	2.919	67,457	2.110	143,296	4.987

Notes: Estimates are based on the authors' calculations using only individuals who do not change cells in the overall-earnings distribution between years $t-1$ and t. The column labeled "Flow: indicates the bin in the overall-earnings distribution that the individual occupied. For example, 2_2 means the individual was in the bottom 20% of the income distribution in both years. The column labeled "Earnings" is the average earnings for the indicated firm type, skill type and flow cell. The column labeled "Ratio" is the ratio of the earnings in the indicated row to the earnings in the 3_3 cell of the same firm and skill type.

6 Conclusion

We use administrative earnings data from the LEHD infrastructure files to analyze the role of the employer in explaining the rise in earnings inequality in the U.S. from 2004-2013. In order to demonstrate the importance of carefully selecting the frame and defining the earnings universe under study, we supplement these earnings data with information from a variety of sources, which we analyze to establish the validity of our final analysis of change in the earnings distribution.

We use SSA-supplied data to identify both invalid SSNs and the fraudulent use of valid SSNs. This allows us to transform the found jobs data in the all-workers frame into the designed eligible-workers frame that references a consistent population over time. When comparing the evolution of the ratios of top-to-bottom percentiles of the earnings distribution between the two worker frames, we find that while both frames show a decrease in earnings inequality in the late-1990s, their patterns diverge starting in 2000. The found frame of all workers shows little to no change in the earnings inequality since 2000. On the other hand, upon removing the immigrant candidates, the designed frame of eligible workers shows a rise in inequality starting in 2000 that is robust across several measures of earnings inequality. This difference highlights the need to be mindful of the sample of workers used when interpreting results from studies of earnings inequality. Furthermore, we compare these inequality results to ones from CPS/ACS. We find that the trends in earnings inequality observed among the eligible workers in the LEHD data are very similar to those observed among the workers expected to be covered under UI in CPS/ACS.

Our results also suggest that, unlike in 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 that recession. While previous research focused primarily on employed persons or persons in the labor force, 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. Using our eligible-workers frame, we have shown that such persons are attached to the labor force, as evidenced by their dynamic employment histories, but the exclusion of their inactive periods from earnings inequality measures understates the degradation at the bottom of the distribution.

Using our designed frame, we assess the role firms play in the rise in earnings inequality. We decomposed earnings in to firm and non-firm component. Using a part of the non-firm component that relates only to measured and unmeasured individual characteristics and controls for differences in labor-force attachment and macroeconomic conditions, we characterize the individuals as low-, medium- or high-skill types. Using the firm component, we characterize the firms as low-, medium- or high-paying firm types. Using the model for changes in the earnings distribution that we constructed for the eligible-workers frame, we analyzed the role played by the position of the worker in the skill-type and firm-type distributions. We show that a typical worker of any skill type would benefit from working at a middle-paying firm relative to a low-paying firm, but it is the workers of any skill type employed at high-paying firms who benefit the most. These individuals not only make higher earnings, they also experience an increase in the probability of moving up the earnings distribution in the following year. While we make no structural claim for this relation between the firm type and placement in the overall-earnings distribution, it is clear that the influence of firms works through channels that are not the same as those through which the effects of individual differences operate.

References

Abowd, John M., Robert Creecy, and Francis Kramarz (2002) "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," LEHD Technical Report TP-2002-06, U.S. Census Bureau.

Abowd, John M., Francis Kramarz, and David N. Margolis (1999) "High Wage Workers and High Wage Firms," *Econometrica*, Vol. 67, pp. 251–333.

Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock (2009) "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators," *Producer Dynamics: New Evidence from Micro Data: University of Chicago Press*, pp. 149–230, URL: http://www.nber.org/chapters/c0485.

Abowd, John M. and Martha H. Stinson (2013) "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data," *Review of Economics and Statistics*, Vol. 95, pp. 1451–1467.

Abowd, John M. and Lars Vilhuber (2011) "National Estimates of Gross Employment and Job Flows from the Quarterly Workforce Indicators with Demographic and Industry Detail," *Journal of Econometrics*, Vol. 161, pp. 82–99.

Acemoglu, Daron and David H. Autor (2011) "Skills, Tasks and Technologies: Implications for Employment and Earnings," *Handbook of Labor Economics: eds. O. Ashenfelter and D. Card*, Vol. 4B, pp. 1043–1171.

Altonji, Joseph G., Anthony A. Smith, Jr., and Ivan Vidangos (2013) "Modeling Earnings Dynamics," *Econometrica*, Vol. 81, pp. 1395–1454, URL: http://dx.doi.org/10.3982/ECTA8415.

Autor, David (2014) "Skills, Education, and the Rise of Earnings Inequality among the 'Other 99 Percent'," *Science*, Vol. 344, pp. 843–851.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney (2008) "Trends in U.S. Wage Inequality: Revising the Revisionists," *The Review of Economics and Statistics*, Vol. 90, pp. 300–323.

Baker, Michael and Gary Solon (2003) "Earnings Dynamics and Inequality among Canadian Men, 1976-1992: Evidence from Longitudinal Income Tax Records," *Journal of Labor Economics*, Vol. 21, pp. 289–321.

Barlevy, Gadi and Daniel Tsiddon (2006) "Earnings Inequality and the Business Cycle," *European Economic Review*, Vol. 50, pp. 55–89.

Blackburn, McKinley L. and David E. Bloom (1987) "Earnings and Income Inequality in the United States," *Population and Development Review*, Vol. 12, pp. 575–609.

Brown, J. David, Julie L. Hotchkiss, and Myriam Quispe-Agnoli (2013) "Does Employing Undocumented Workers Give Firms a Competitive Advantage?" *Journal of Regional Science*, Vol. 53, pp. 158–170.

Card, David, Ana Rute Cardoso, Jörg Heining, and Patrick Kline (forthcoming) "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*.

Card, David, Joerg Heining, and Patrick Kline (2013) "Workplace Heterogeneity and the Rise of West German Wage Inequality," *The Quarterly Journal of Economics*, Vol. 128, pp. 967–1015.

Castañeda, Ana, Javier Díaz-Giménez, and José-Victor Ríos-Rull (1998) "Exploring the Income Distribution Business Cycle Dynamics," *Journal of Monetary Economics*, Vol. 42, pp. 93–130.

Freeman, Richard B., Erling Barth, and James Davis (forthcoming) "Augmenting the Human Capital Earnings Equation with Measures of Where People Work," *Journal of Labor Economics*.

Groves, Robert (2010) "Designed Data and Organic Data," Director's Blog, May, URL: http://directorsblog.blogs.census.gov/2011/05/31/designed-data-and-organic-data/.

Guvenen, Fatih, Serdar Ozkan, and Jae Song (2014) "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, Vol. 122, pp. 621–660.

Hall, Robert (2005) "Job Loss, Job Finding, and Unemployment in the U.S. Economy over the Past Fifty Years," *NBER Macroeconomics Annual*, pp. 101–137.

Haltiwanger, John, Henry Hyatt, and Erika McEntarfer (forthcoming) "Who Moves Up the Firm Productivity Job Ladder?" *Journal of Labor Economics*.

Haltiwanger, John, Henry Hyatt, Erika McEntarfer, and Liliana Sousa (2012) "Business Dynamic Statistics Briefing: Job Creation, Worker Churning, and Wages at Young Businessess,"Technical report, U.S. Census Bureau.

Hirsch, Barry T. and Muhammad M. Husain (2013) "Multiple Job Holding, Local Labor Markets, and the Business Cycle," *mimeo*.

Imbens, Guido W. and Donald B. Rubin (2015) *Causal Inference for Statistics, Social, and Biomedical Sciences*, New York: Cambridge University Press.

Jacobson, Louis S., Robert J. Lalonde, and Daniel G. Sullivan (1993) "Earnings Losses of Displaced Workers," *The American Economic Review*, Vol. 83, pp. 685–709.

Juhn, Chinhui, Kristin McCue, Holly Monti, and Brooks Pierce (forthcoming) "Firm Performance and the Volatility of Worker Earnings," *Journal of Labor Economics*.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce (1993) "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy*, Vol. 101, pp. 410–442.

Katz, Lawrence F. and David H. Autor (1999) "Changes in the Wage Structure and Earnings Inequality," *Handbook of Labor Economics: eds. O. Ashenfelter and D. Card*, Vol. 3, pp. 1463–1555.

Katz, Lawrence F. and Kevin M. Murphy (1992) "Changes in Relative Wages, 1963-1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, Vol. 107, pp. 35–78.

Kopczuk, Wojciech, Emmanuel Saez, and Jae Song (2010) "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data Since 1937," *The Quarterly Journal of Economics*, Vol. 125, pp. 91–128.

Levy, Frank and Richard J. Murnane (1992) "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations," *Journal of Economic Literature*, Vol. 30, pp. 1333–1381.

Piketty, Thomas and Emmanuel Saez (2003) "Income Inequality in the United States, 1913-1998," *Quarterly Journal of Economics*, Vol. 118, pp. 1–41.

——— (2013) "Top Incomes and the Great Recession," *IMF Economic Review*, Vol. 61, pp. 456–478, URL: http://dx.doi.org/10.1057/imfer.2013.14.

Rubin, Donald B. (1987) *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley & Sons, Inc.

Saez, Emmanuel (2015) "Striking It Richer: The Evolution of Top Incomes in the United States (Updated with 2013 preliminary estimates)," *Manuscript, University of California Berkeley*, URL: http://eml.berkeley.edu/~saez/saez-UStopincomes-2014.pdf.

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter (2015) "Firming up Inequality," *mimeo*.

Spletzer, James R. (2014) "Inequality Statistics from the LEHD," *mimeo*.

Stevens, Ann Huff (1997) "Persistent Effects of Job Displacement: The Importance of Multiple Job Losses," *Journal of Labor Economics*, Vol. 15, pp. 165–188.

Van Reenen, John, David Card, Nicholas Bloom, Stefan Bender, and Stefanie Wolter (forthcoming) "Management Practices, Workforce Selection, and Productivity," *Journal of Labor Economics*.

Whither Biometrics Committee (2010) *Biometric Recognition: Challenges and Opportunities*: National Academies Press, URL: http://nap.edu/12720.

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:

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

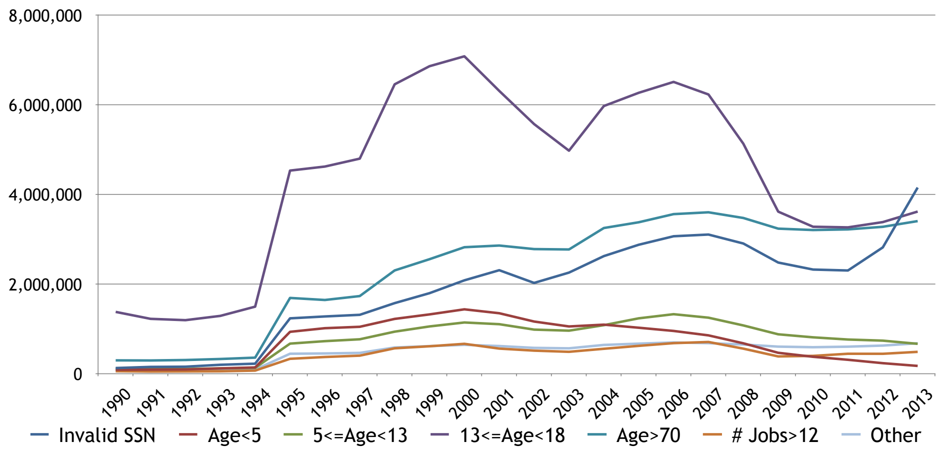
where AW_t is the set of workers in the all-workers frame in year t.

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.

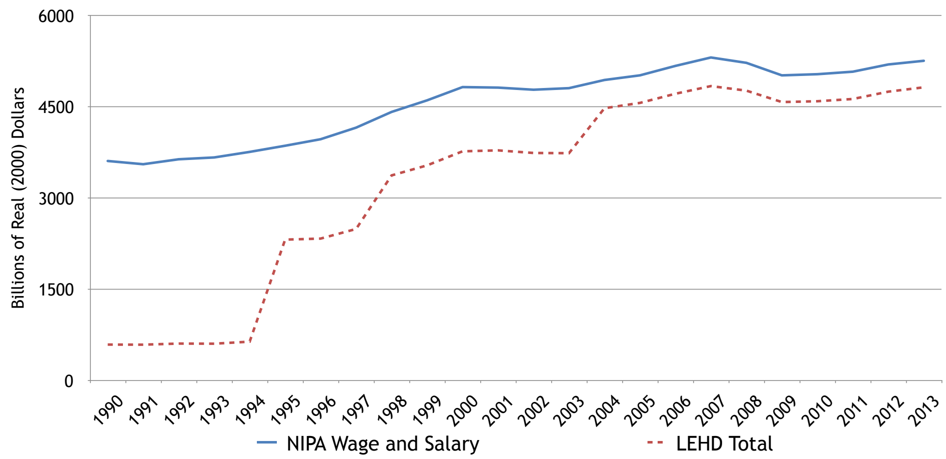
The coverage of both the all-workers and eligible-workers frames is very low relative to the NIPA estimates in the early 1990s but increases dramatically in 1995 once the historical data for the more populous states (CA, FL, NY, and TX) have entered the LEHD infrastructure files. When the frame is complete (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*	Workers *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.

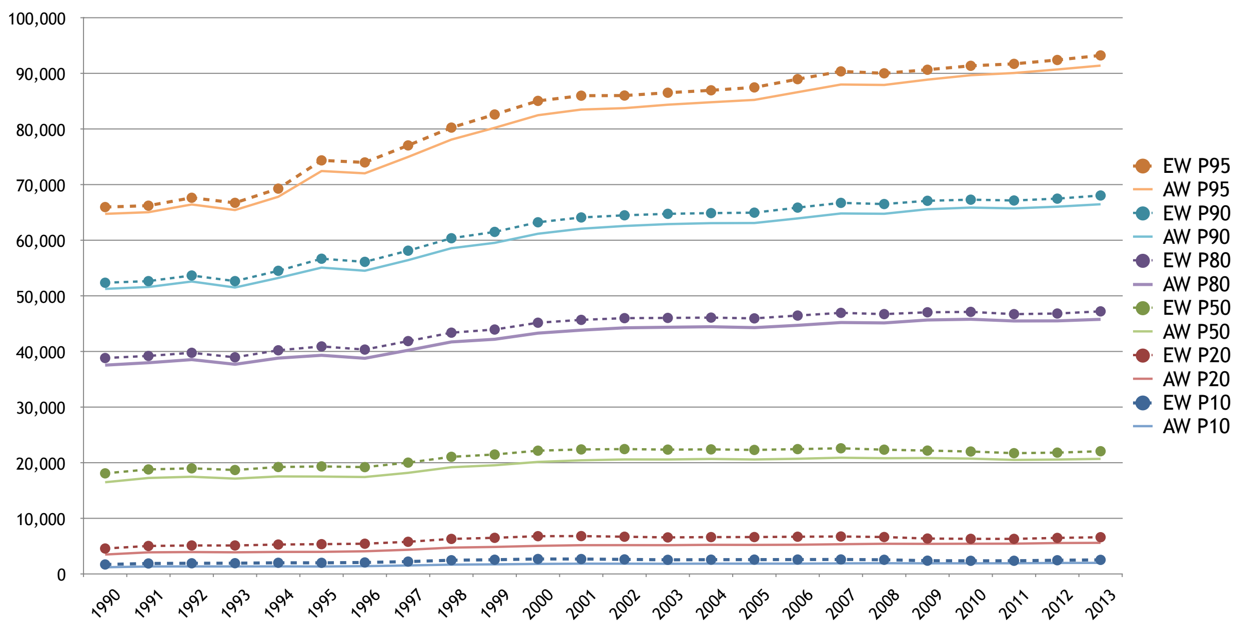
B Inequality Trends in the LEHD All-Workers Frame (1990-2013)

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.

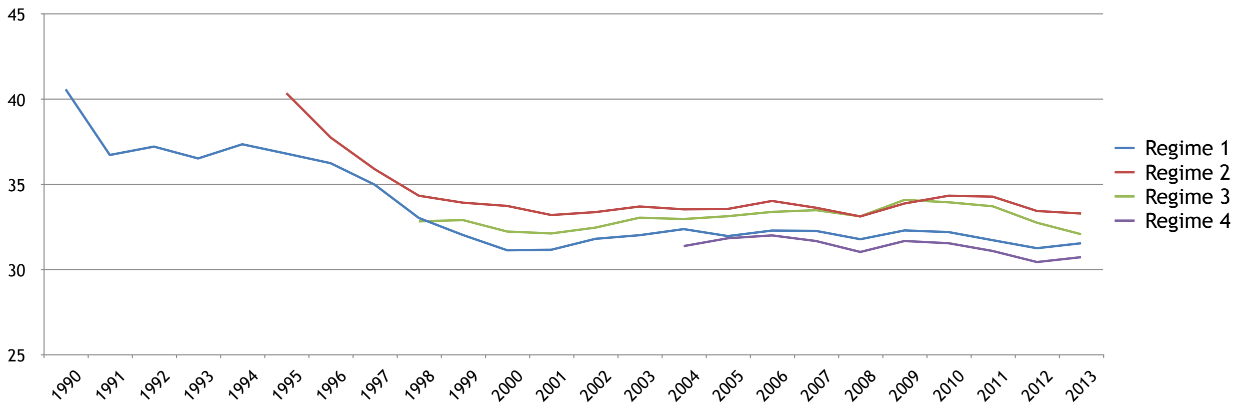
Figure B.1: Percentiles of the Earnings Distribution by Worker Frame



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.

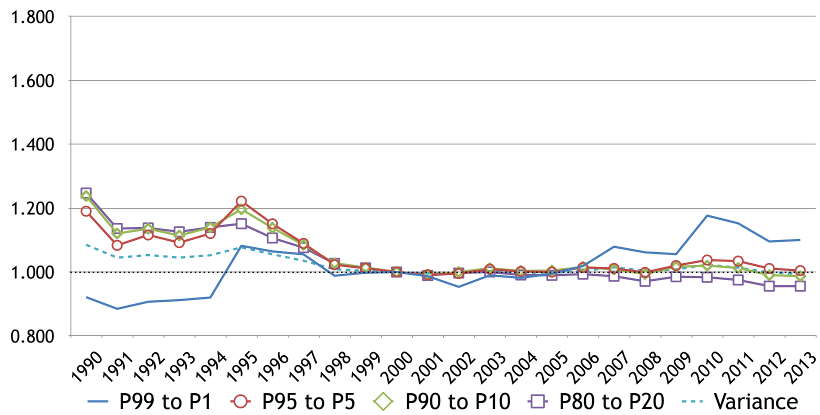
Table B.1: Average Percentiles of the Earnings Distribution by Worker Frame (1995-2013)

	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)



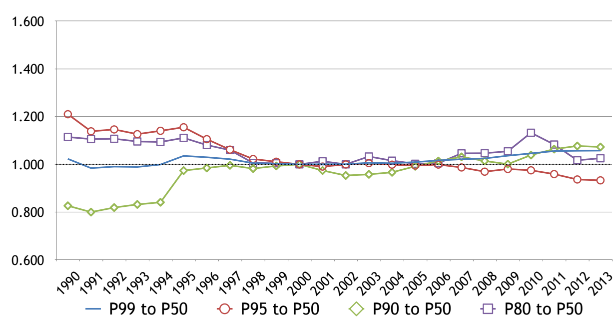
Notes: Subplot (a) presents measures of earnings inequality for *all* workers in all states relative to 2000 from 1990-2013. Subplot (b) presents measures of earnings inequality for *eligible* workers in all states relative to 2000 from 1990-2013. The measures of earnings inequality considered are (i) *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

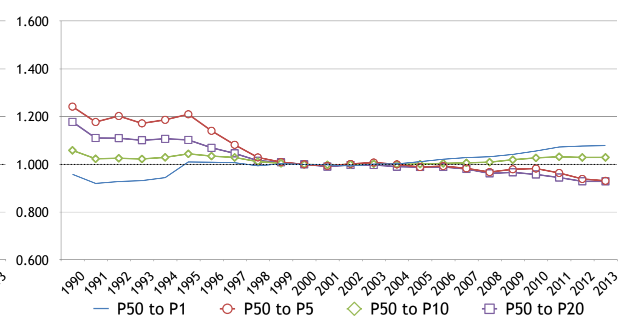
	Inequality Measures				
	99th/1st	*95th/5th*	*90th/10th*	*80th/20th*	*Variance*
			All Workers		
Pre-2000	1.038	1.099	1.092	1.075	1.036
Post-2000	1.050	1.010	1.004	0.983	1.004
Pre-GR	1.001	1.003	1.005	0.992	1.001
GR	1.059	1.009	1.005	0.978	1.006
Post-GR	1.131	1.022	1.002	0.968	1.007
			Eligible Workers		
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)



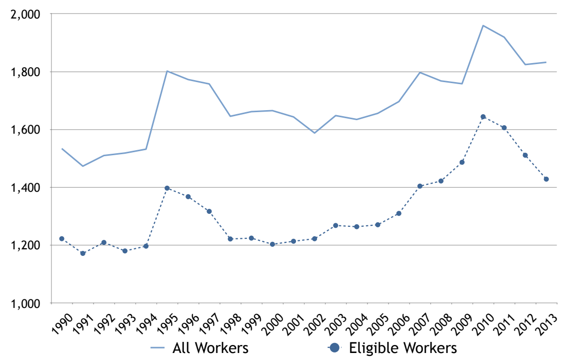
(a) All Workers - Top



(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

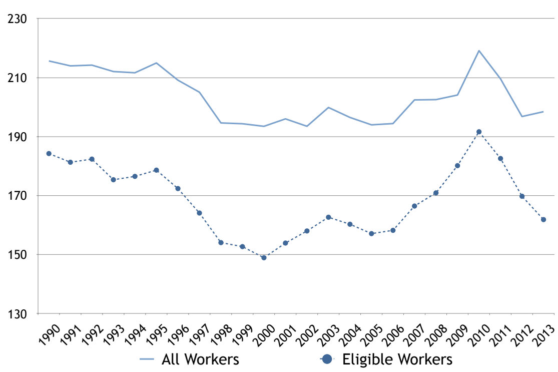


(a) Ratio of 99th to 1st

(b) Ratio of 95th to 5th

(c) Ratio of 99th to 50th

(d) Ratio of 95th to 50th

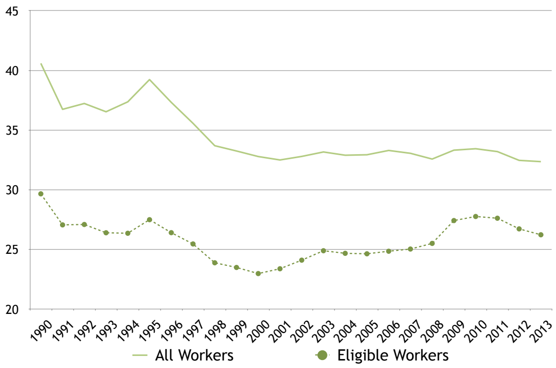


(e) Ratio of 50th to 1st

(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

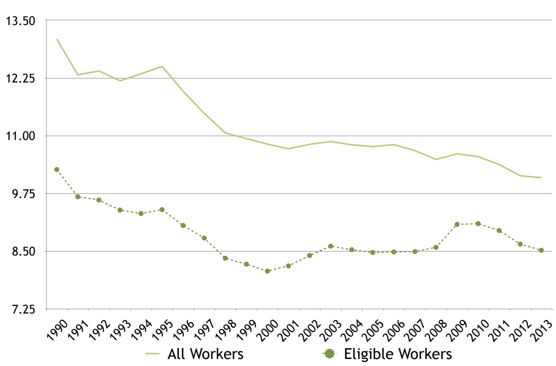


(a) Ratio of 90th to 10th

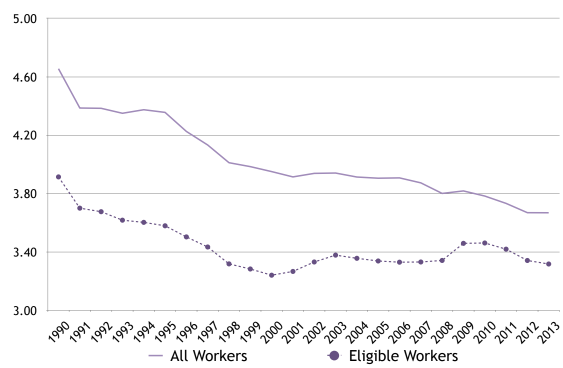
(b) Ratio of 80th to 20th

(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.

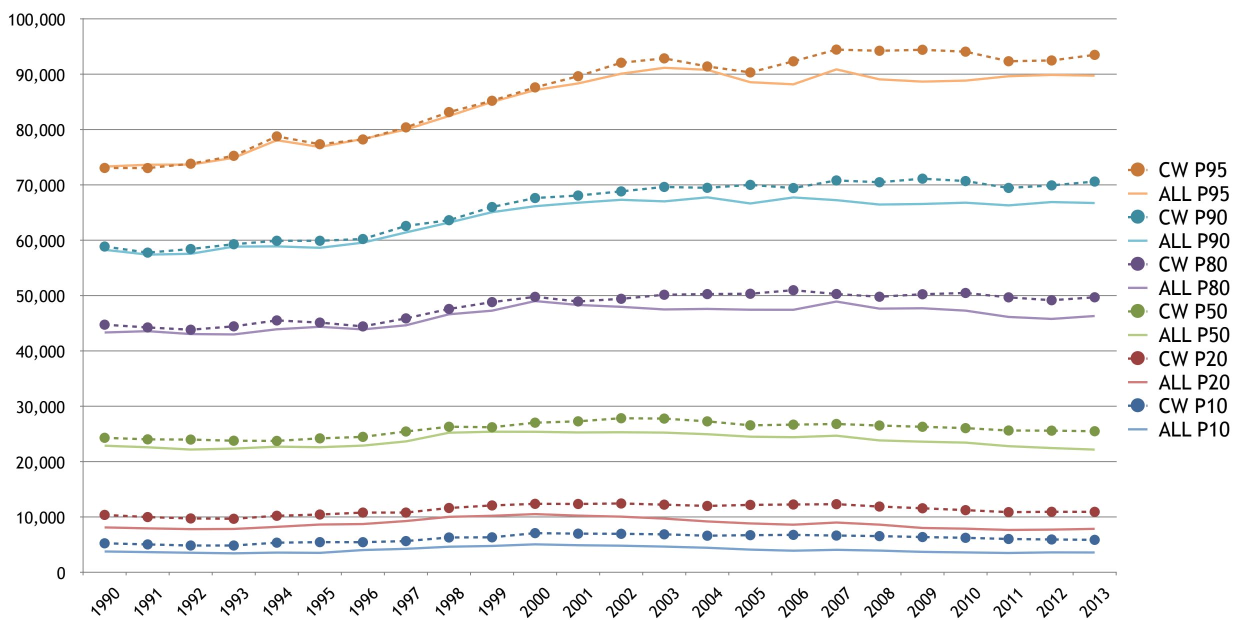
Finally, in most of the results to follow, we do not report results separately for CPS-ASEC and ACS individuals. Instead, in the overlapping years (2000-2003), we interpolate estimates computed from the CPS-ASEC and the ACS to create a single time series using the method in Abowd and Vilhuber (2011).

[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



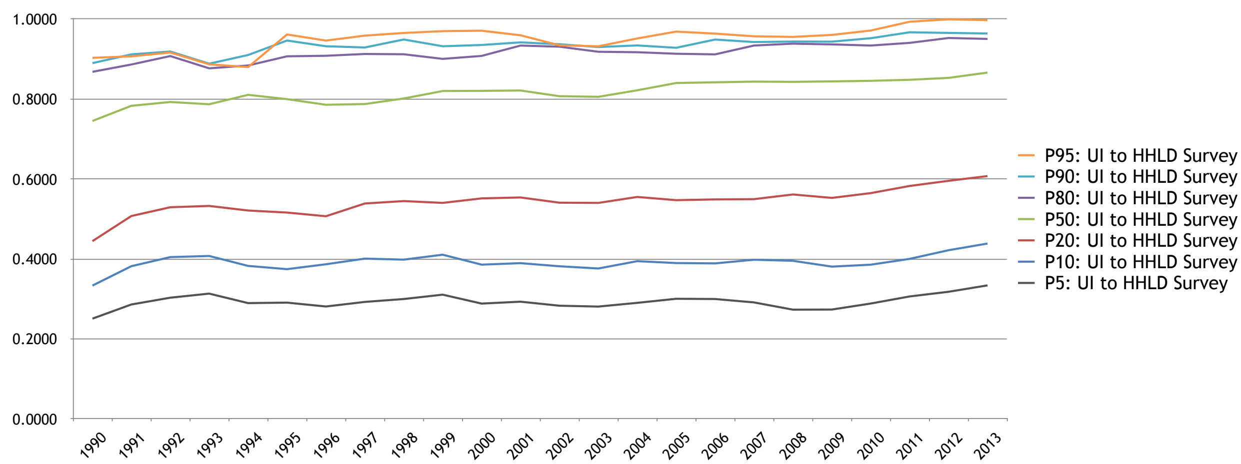
Notes: This figure plots the 10^{th}, 20^{th}, 50^{th}, 80^{th}, 90^{th}, and 95^{th} percentiles of the earnings distribution for all workers (ALL) and for covered workers (CW) in the CPS-ASEC (1990-2003) and the ACS (2000-2013) surveys by year. Covered workers include respondents whose employment relation 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.

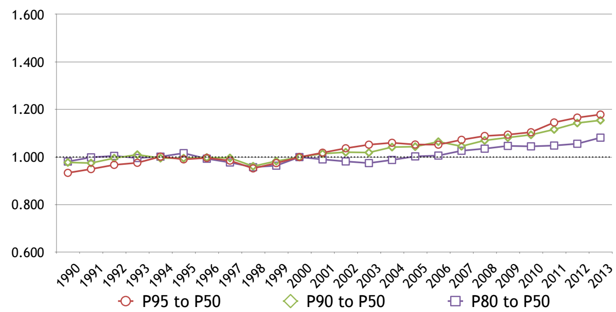
Table C.1: Inequality Measures Relative to 2000 for Workers in Household Surveys

	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

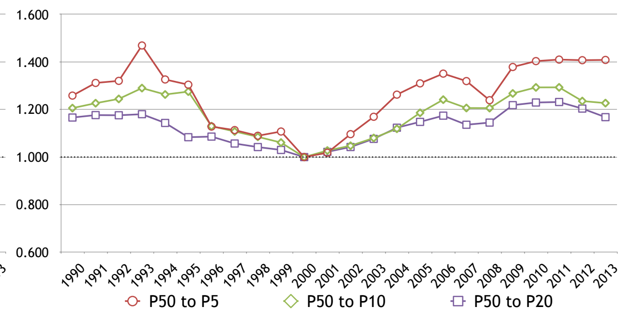
Notes: The first panel presents measures of earnings inequality for *all* workers in CPS/ACS, while the second panel presents the same measures for *covered* workers. The measures of earnings inequality considered are (i) *$95^{th}/5^{th}$*: the ratio of the 95^{th} to the 5^{th} percentile; (ii) *$90^{th}/10^{th}$*: 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.

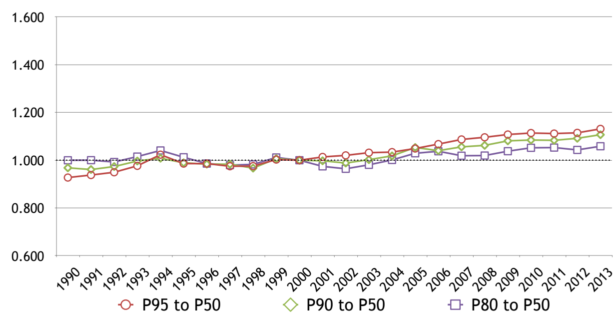
Figure C.3: Selected Inequality Measures 1990-2013 for the Top and Bottom of the Earnings Distribution, Relative to 2000 (CPS/ACS)



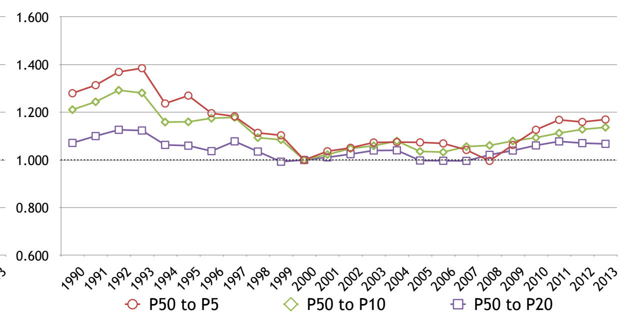
(a) All Workers - Top



(b) All Workers - Bottom



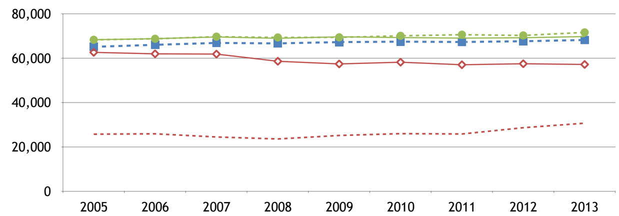
(c) Covered Workers - Top



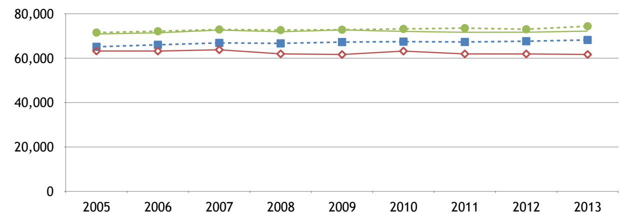
(d) Covered Workers - Bottom

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

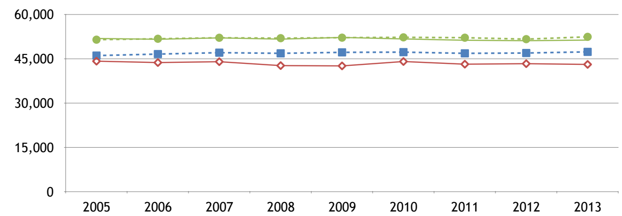


(a) 90th Percentile (All Workers)

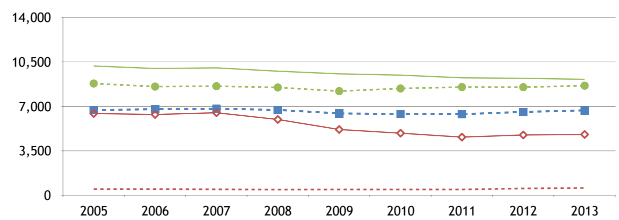


(b) 90th Percentile (Covered Workers)

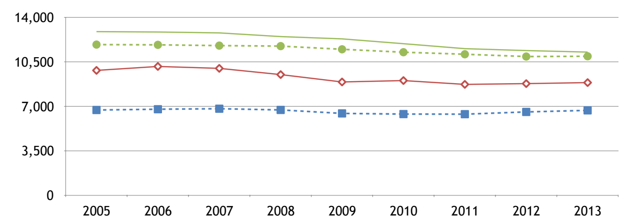
(c) 80th Percentile (All Workers)



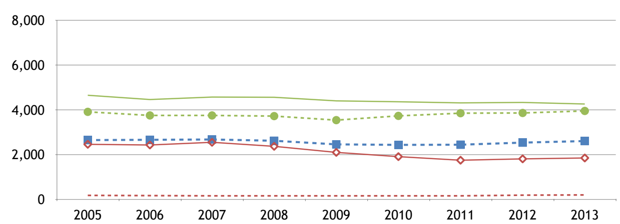
(d) 80th Percentile (Covered Workers)



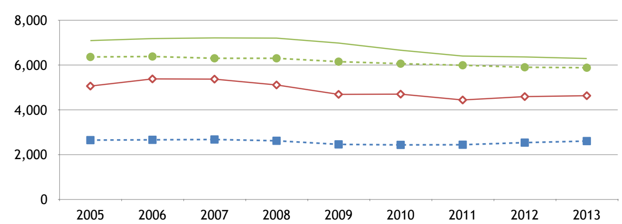
(e) 20th Percentile (All Workers)



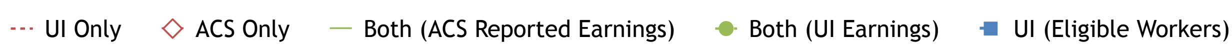
(f) 20th Percentile (Covered Workers)



(g) 10th Percentile (All Workers)



(h) 10th Percentile (Covered Workers)



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]

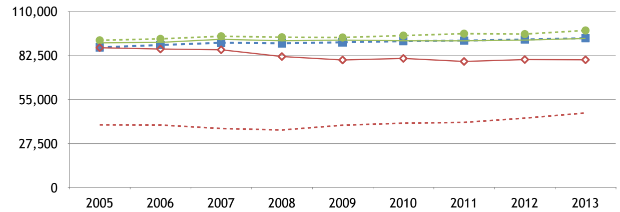
For workers whose earnings are both reported in the ACS and found in LEHD (matched workers), the percentiles computed using the ACS earnings are nearly identical to those computed using UI earnings. Specifically, notice that in Figure C.5 the solid green line and the dotted

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

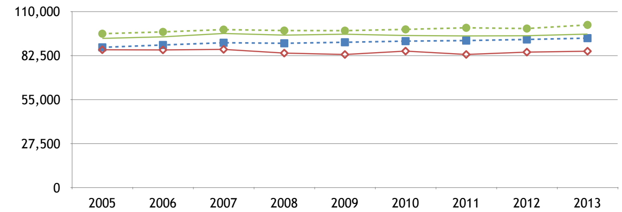
green line with circles are very close to each other in all subplots, and especially at and above the median. The differences in the CPS/ACS percentiles and the LEHD percentiles in Figure C.2 are, therefore, very unlikely to be due to misreporting in household surveys. Instead, they must be due to differences 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.

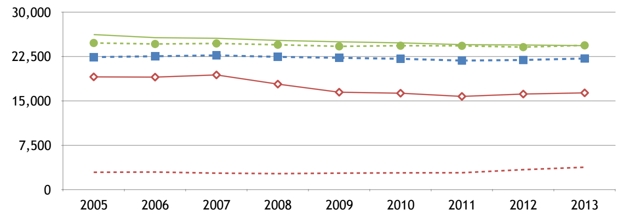
Figure C.5: Comparison of Percentiles in the ACS and LEHD



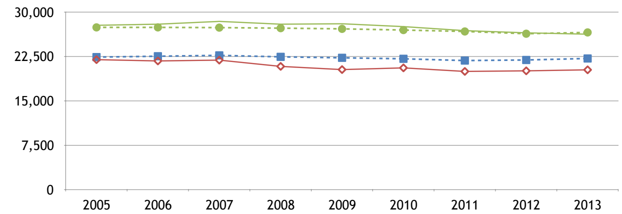
(a) 95th Percentile (All Workers)



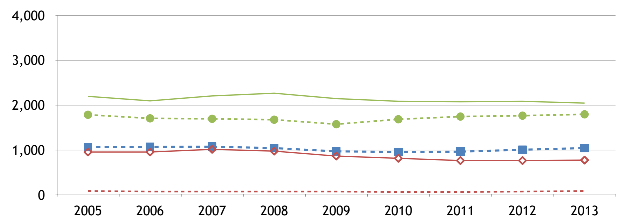
(b) 95th Percentile (Covered Workers)



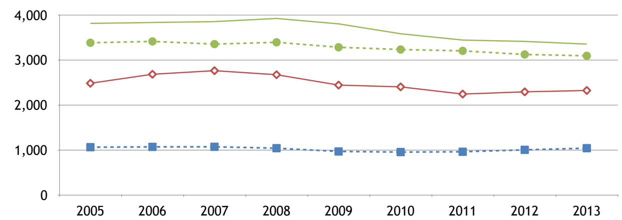
(c) 50th Percentile (All Workers)



(d) 50th Percentile (Covered Workers)



(e) 5th Percentile (All Workers)



(f) 5th Percentile (Covered Workers)

UI Only ACS Only Both (ACS Reported Earnings) Both (UI Earnings) UI (Eligible 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.

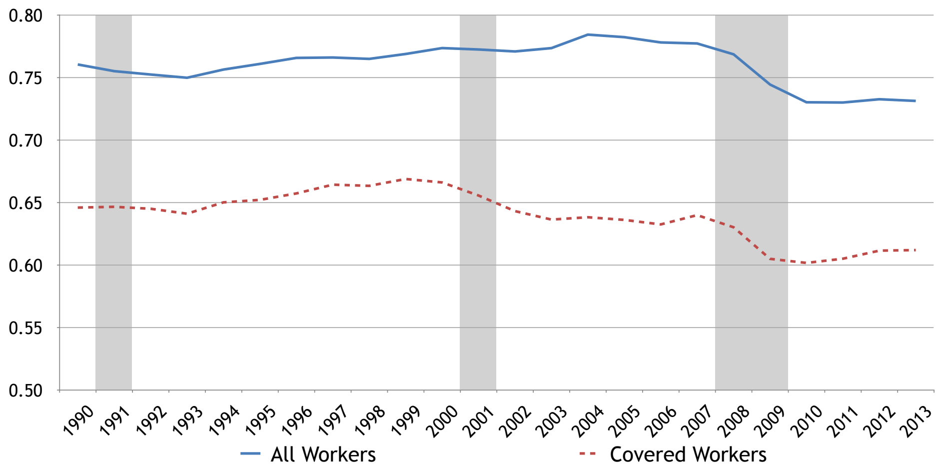
In Appendix B, we excluded inactive workers from the analysis so that we could focus on trends in the ratios of top and bottom percentiles over time. While some inactive workers, given the wages and employment terms on offer, choose to be nonparticipants, others are involuntarily excluded from the labor 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).

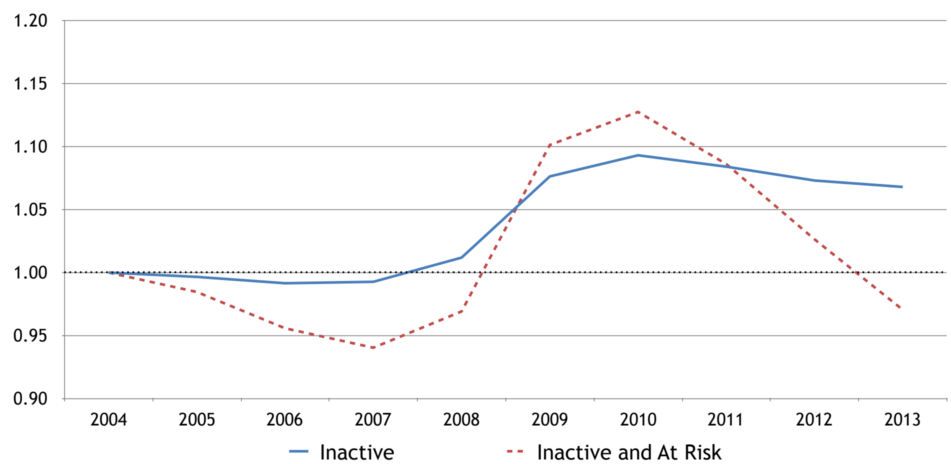
Table D.1: Inequality Measures with and without Inactive Workers

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

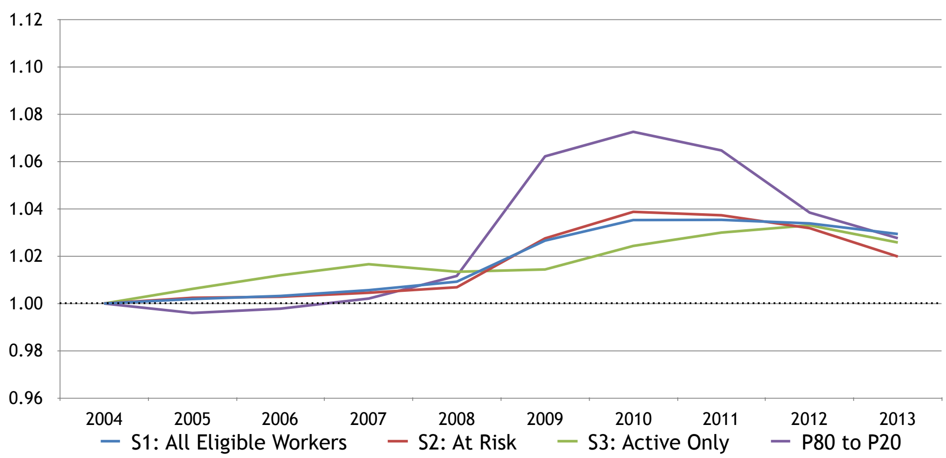


Notes: This figure shows the share of eligible workers who are inactive (solid blue line) and the share who are "at risk" (dashed red line) relative to the base year 2004. In a given year, a person is *inactive* if that person did not make positive earnings that year. In a given year, a person is "*at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

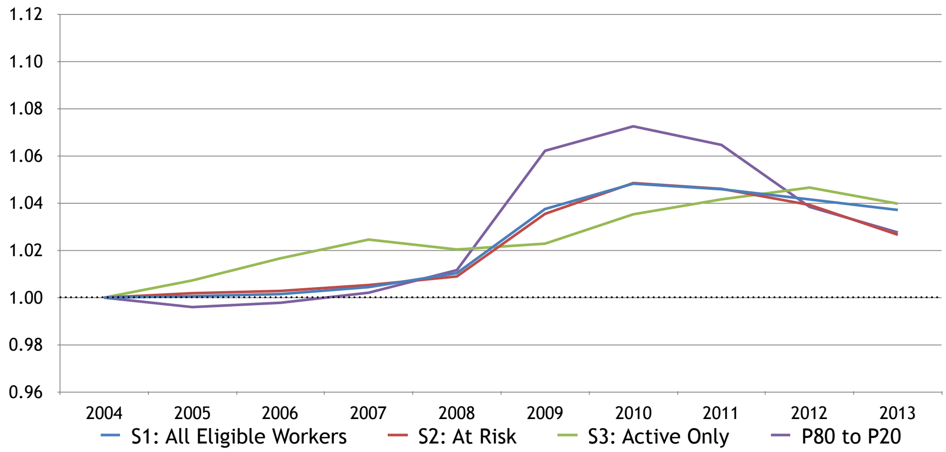
Figure D.3: Inequality Measures – Gini Coefficient



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.

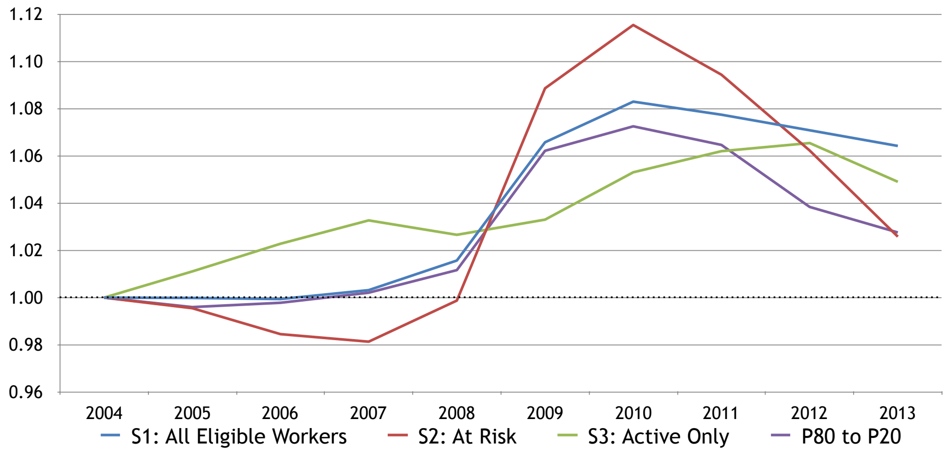
Figure D.4: Inequality Measures – Hoover Index



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:

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

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:

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

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 Workers	12,950,373 14.4%	1,057,067 3.8%	1,886,128 2.3%	2,159,901 7.8%	18,053,469 7.9%
Total Earnings	— —	2,671 3.4%	10,752 0.5%	305,357 12.2%	318,780 7.0%
Average Earnings	— —	-11 -0.4%	-400 -1.7%	3,986 4.4%	-175 -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:

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

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. ($\bar{e}^*_2 = \bar{e}^B_{02} = \bar{e}^A_{20} = \bar{e}^B_{12} = \bar{e}^A_{21} = \bar{e}^B_{22} = \bar{e}^A_{22} = \bar{e}^B_{32} = \bar{e}^A_{23} = \bar{e}^B_{42} = \bar{e}^A_{24}$).

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

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.

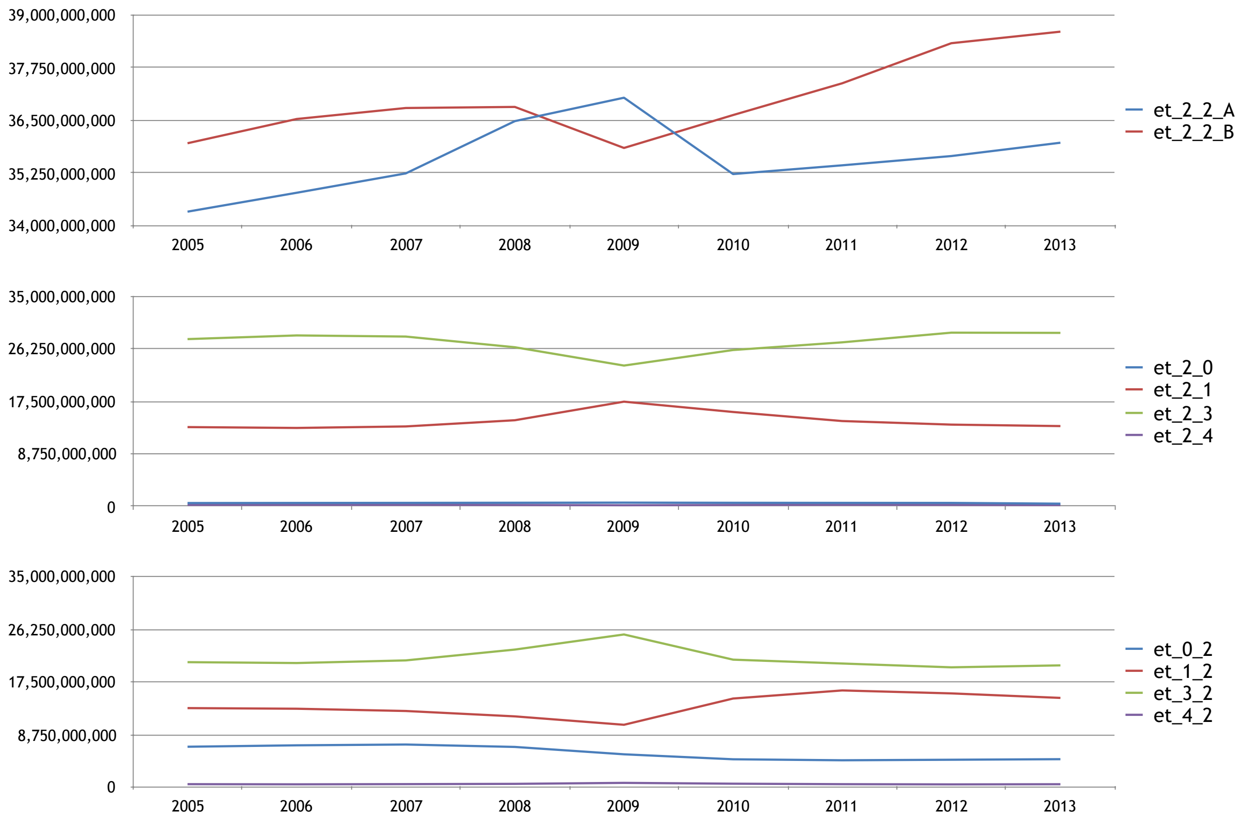
Table E.3: Average Earnings and Variability by Transition Type

	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.

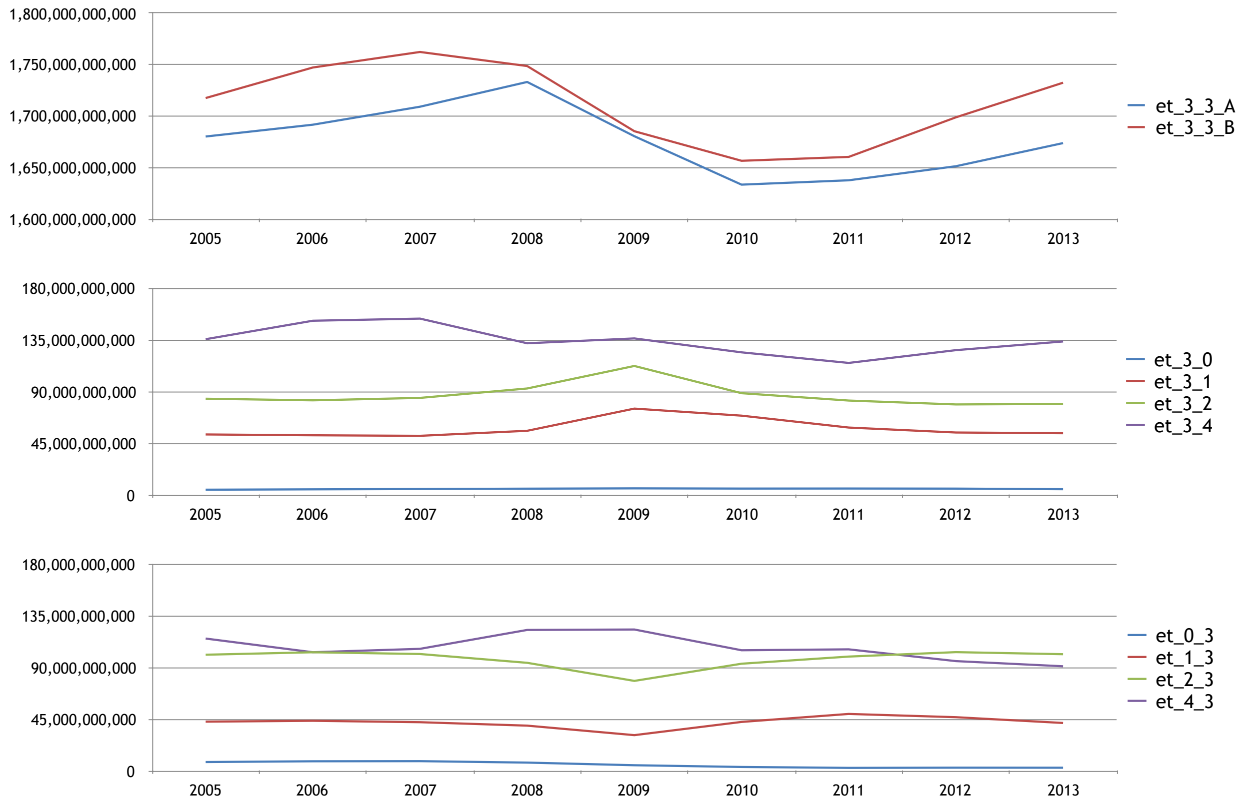
Appendix Figures E.1-fig:TE4 repeat the analysis shown in the main text in Figures 10-12.

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



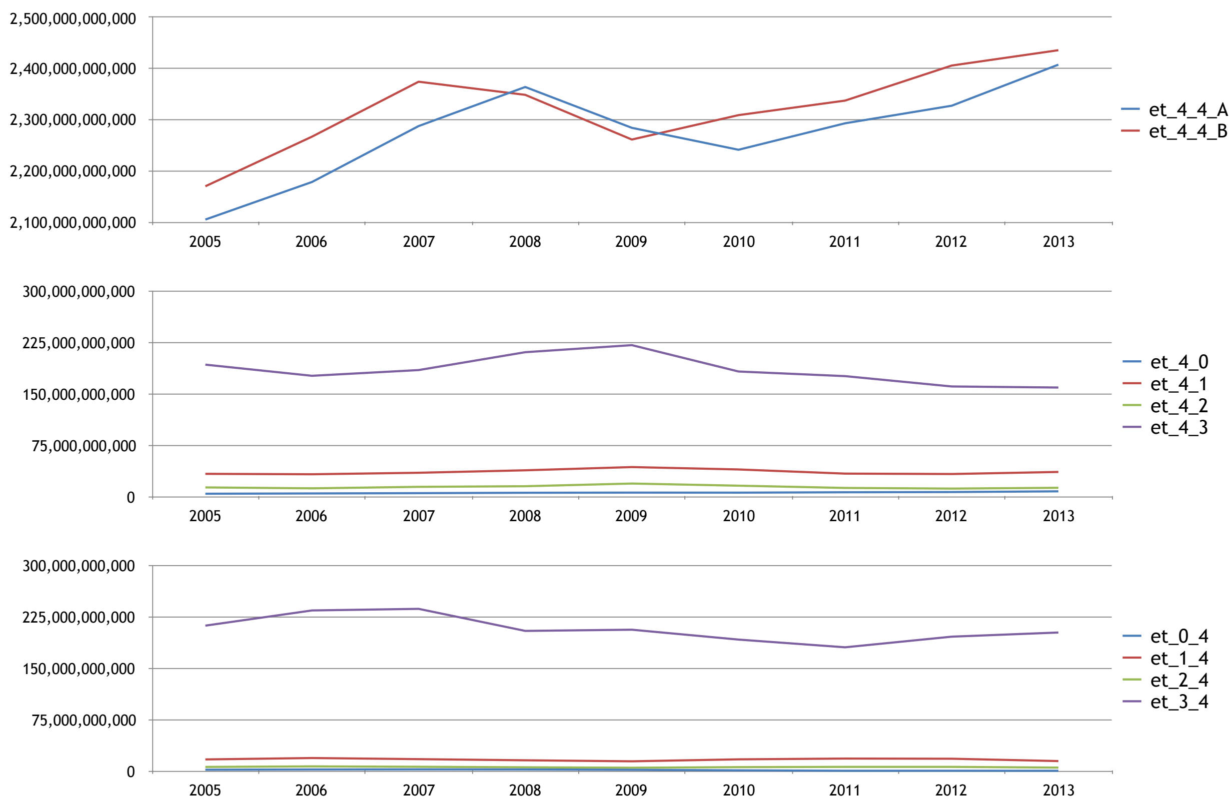
Notes: The estimates are based on the authors' calculations using transitions into and out of categories of the earnings distribution, including inactivity.

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)

					Previous Year			**Current Year**		
All Earnings	**Share**	**Flow**	**Average Count**	**Pct**	*Total*	*Firm*	*Non-Firm*	*Total*	*Firm*	*Non-Firm*
					Bottom Firm (33%)					
		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	—	—	—
Bottom	1.000	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
		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	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 (56%)					
		2_0	116,724	0.9%	2,660	72	2,588	—	—	—
		2_1	3,613,606	27.3%	2,289	42	2,247	—	—	—
Bottom	0.841	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
		3_0	21,191	0.8%	8,381	2,189	6,193	—	—	—
		3_1	428,729	17.2%	8,384	2,249	6,135	—	—	—
Middle	0.159	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
		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%	—	—	—	—	—	—
					Top Firm (11%)					
		2_0	17,420	1.4%	2,913	1,598	1,314	—	—	—
		2_1	469,324	38.3%	2,758	1,515	1,243	—	—	—
Bottom	0.396	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
		3_0	16,910	0.9%	12,121	7,299	4,822	—	—	—
		3_1	375,155	20.3%	12,082	7,280	4,802	—	—	—
Middle	0.596	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
		4_0	172	0.7%	97,790	93,786	4,005	—	—	—
		4_1	1,924	7.6%	96,408	92,599	3,809	—	—	—
Top	0.008	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

					Previous Year			Current Year		
All Earnings	**Share**	**Flow**	**Average Count**	**Pct**	*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

					Previous Year			**Current Year**		
All Earnings	**Share**	**Flow**	**Average Count**	**Pct**	*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.