

IZA DP No. 6182

Decomposing the Sources of Earnings Inequality: Assessing the Role of Reallocation

Fredrik Andersson
Elizabeth E. Davis
Matthew L. Freedman

Julia I. Lane
Brian P. McCall
L. Kristin Sandusky

November 2011

Decomposing the Sources of Earnings Inequality: Assessing the Role of Reallocation

Fredrik Andersson

Office of the Comptroller of the Currency

Julia I. Lane

National Science Foundation and IZA

Elizabeth E. Davis

University of Minnesota

Brian P. McCall

University of Michigan

Matthew L. Freedman

Cornell University

L. Kristin Sandusky

U.S. Census Bureau

Discussion Paper No. 6182
November 2011

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Decomposing the Sources of Earnings Inequality: Assessing the Role of Reallocation*

This paper exploits longitudinal employer-employee matched data from the U.S. Census Bureau to investigate the contribution of worker and firm reallocation to changes in earnings inequality within and across industries between 1992 and 2003. We find that factors that cannot be measured using standard cross-sectional data, including the entry and exit of firms and the sorting of workers across firms, are important sources of changes in earnings distributions over time. Our results also suggest that the dynamics driving changes in earnings inequality are heterogeneous across industries.

JEL Classification: J0, J30

Keywords: inequality, linked employer-employee data, sorting

Corresponding author:

Julia Lane
National Science Foundation
4201 Wilson Boulevard
Arlington, VA 22230
USA
E-mail: jlane@nsf.gov

* The authors would like to thank Charlie Brown, David Card, Erica Groshen, participants of the labor lunch at the University of California-Berkeley, and two anonymous *Industrial Relations* referees for their helpful comments. This document reports the results of research and analysis undertaken by the U.S. Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This research is a part of the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) Program, which is partially supported by the National Science Foundation Grants SES-9978093 and SES-0427889 to Cornell University (Cornell Institute for Social and Economic Research), the National Institute on Aging Grant R01-AG018854-02, and the Alfred P. Sloan Foundation. The views expressed on statistical, methodological, and technical issues are those of the authors and not necessarily those of the Office of the Comptroller of the Currency or of the U.S. Census Bureau, its program sponsors, or its data providers. Some of the data used in this research are confidential data from the LEHD Program. The U.S. Census Bureau supports external researchers' use of these data through the Research Data Centers (www.ces.census.gov). For questions regarding the data, please contact the U.S. Census Bureau, LEHD Program, Room 6H136C, 4600 Silver Hill Road, Suitland, Maryland 20746, USA (lehd.did.census.gov).

Introduction

There has long been interest in disentangling the sources of changes in earnings inequality. However, while there is a large body of research on the evolution of the wage distribution, few studies have focused on the role of firm turnover in driving changes in earnings inequality, and fewer still have considered the importance of sorting among workers and firms in affecting the distribution of wages. This paper takes advantage of linked employer-employee data to examine whether the very high levels of job and worker reallocation in the U.S. economy, which have important implications for productivity, also affect the distribution of earnings within and across industries.

We develop a decomposition methodology that exploits our longitudinal data and permits us to quantify factors contributing to changes in earnings inequality over time. First, we re-examine the impact of changes in the composition of the workforce, which has received considerable attention in past work using worker-based surveys. Second, we evaluate the impact of changes in firm composition that occur as employers enter and exit the market. Finally, we consider the impact on the earnings distribution of the way in which workers are allocated across firms. In addition to an examination of the overall economy, we control for the effect of changing industry structure by separately examining each of the nine major industries that comprise the private sector.

We find that there is no single factor – workforce composition, firm composition, or the match between firms and workers – that can fully explain changes in the earnings distribution in the broad economy or in any given industry. Further, even when the direction of change in earnings inequality is similar across industries, the underlying forces contributing to changes can be very different.

That being said, some common patterns are evident. Between 1992 and 2003, the entry and exit of firms acted to reduce wage inequality in most industries, primarily by raising wages at the lower end of the wage distribution. Meanwhile, entering workers tended to be similar to exiting workers within sectors, and hence changes in worker composition can only explain a small fraction of the overall changes in earnings inequality during the period. That is not to say that the extensive amount of worker turnover in the economy did not have important implications for the distribution of earnings. Indeed, increasing assortative matching between workers and firms

during the 1990s and early 2000s contributed to greater earnings inequality within and between industries. Coupled with likely changes in the returns to worker characteristics, the trend toward greater assortative matching is consistent with skill-biased technological change as an explanation for recent increases in earnings inequality.

Importantly, while our decomposition methodology allows us to quantify the contribution of various factors to changes in earnings distributions over time, we do not estimate causal parameters. As with other decomposition techniques (e.g., DiNardo et al., 1996; Foster et al., 2001; Machado and Mata, 2005; Firpo et al., 2007), the methodology we develop in this paper is descriptive in nature. However, our approach sheds new light on how factors that cannot be measured using cross-sectional data, and in particular the extensive amount of ongoing worker and firm reallocation in the economy, have qualitatively and quantitatively important effects on the earnings distribution.

The paper is organized as follows. After a brief review of the literature in the next section, we develop an econometric method for decomposing the sources of change in the earnings distribution that takes advantage of longitudinal employer-employee matched data. Next, we discuss our data and present some basic empirical facts about recent changes in earnings distributions in each of the nine major industries that make up the private sector. We then describe the results of performing the decomposition, first for the private sector as a whole and then for each major industry. Finally, we offer concluding thoughts on how our results improve our understanding of the importance of reallocation in driving changes in the distribution of earnings within and across industries.

Background

Despite a vast literature that attempts to explain the increase in earnings inequality that the U.S. has witnessed in recent decades, there is still no consensus on its primary causes (Davidson and Reich, 1988; Levy and Murnane, 1992; Juhn et al., 1993; Danziger and Gottschalk, 1995; Katz and Autor, 1999; Firpo et al., 2007). A large body of research suggests that the increase in inequality was driven by skill-biased technological change interacting in complex ways with changes in unionization, management structures, and international trade (Acemoglu, 2002). However, there is some disagreement about the relative importance of labor market versus institutional factors. For example, some researchers point to changes in the composition of the

workforce as an important contributor to growing earnings inequality (Lemieux, 2006). Others, such as DiNardo et al. (1996), Lee (1999), and Card and DiNardo (2002), identify structural changes, such as the fall in the real value of the minimum wage and declines in unionization, as key drivers behind recent increases in inequality. Fortin and Lemieux (1997) also suggest that deregulation in transportation, communication, and banking industries in the 1980s may have played a role.

In more recent work, Autor et al. (2008) present evidence that since the late 1980s, there has been a divergence in patterns of inequality between the upper and lower halves of the wage distribution, with the lower half (as measured by the 50-10 difference in log wages) either being compressed or not changing and the upper half (as measured by the 90-50 difference) exhibiting increasing dispersion. They point to technological change as a possible explanation, highlighting how computerization may have reduced demand for workers in the middle of the wage distribution.

Nearly all of the literature on wage inequality to date, including recent work using more sophisticated decomposition techniques (e.g., Firpo et al., 2007), is based on surveys of workers, most notably the Current Population Survey (CPS). However, there is reason to believe that changes in the distribution of earnings may be due at least in part to changes on the firm side of the labor market. It is well established that different firms pay observationally similar workers different wages (Mortensen, 2003). Therefore, even holding worker characteristics constant, firm entry and exit and the reallocation of resources across different sectors of the economy may contribute to changes in earnings inequality. Consistent with this, Bernard and Jensen (2000) find that changes in wage inequality across U.S. states are highly correlated with shifts in industrial composition.

Meanwhile, the sheer magnitude of reallocation of workers across firms over time suggests that it could potentially have a large impact on the earnings distribution. Davis et al. (1996) document the large and persistent rates of job creation and destruction in the U.S. economy and highlight the dominance of idiosyncratic factors in driving ongoing reallocation of jobs across firms. Even within firms, the amount of worker turnover is large. Burgess et al. (2000) point out that after nine years, only about one third of private-sector workers are still employed by the same employer. At the same time, a burgeoning literature suggests that the way in which workers are matched to different types of firms is not random; in particular, there is some evidence to

suggest that high wage workers tend to be matched with high wage firms and low wage workers with low wage firms (Abowd et al., 2002; Woodcock, 2008; Abowd et al., 2009a). Changes in that allocation can change earnings distributions over time (Lane, 2009).

Taking advantage of longitudinal employee-employer matched data, we focus in this study on the impact of not only changes in the types of workers and the types of firms in different industries, but also changes in the allocation of workers across firms within industries on changes in earnings distributions over time. Our decomposition approach complements other recently developed techniques to study the extent to which certain variables explain changes in an outcome variable at different points in the outcome variable's distribution.¹ For example, building on work by DiNardo et al. (1996), Machado and Mata (2005) develop a decomposition technique that uses quantile regressions to partition observed changes in the distribution of earnings into several factors contributing to those changes and to quantify each factor's effect on overall wage inequality through simulations. Meanwhile, Firpo et al. (2009) use recentered influence function regressions to generalize the Oaxaca-Blinder decomposition to allow for the examination of the contribution of covariates to an outcome at points other than the mean.

Our approach to decomposing changes in earnings inequality over time bears some resemblance to decomposition methods adopted to determine the sources of changes in aggregate productivity (e.g., Baily et al., 1992; Griliches and Regev, 1995; Foster et al., 2001), but also draws on recent innovations in exploiting linked employee-employer data to disentangle the contributions of workers, firms, and worker-firm matches to earnings (Abowd et al., 1999; Woodcock, 2008). While our approach has several important limitations, which we describe in detail below, our results both complement and expand upon recent empirical work on the sources of change in earnings inequality in the U.S. In particular, in line with Firpo et al. (2007), we show that changes in the industrial structure of the economy have played little role in driving overall changes in earnings inequality. However, we also demonstrate the critical importance of the entry and exit of firms and the sorting of workers across firms in shaping the distribution of earnings, and further explore heterogeneity across industries in the role of various factors in driving changes in inequality over time. We turn to a discussion of how we decompose changes in earnings inequality using panel data on workers and firms in the next section.

¹ See Fortin et al. (2010) for a recent review of decomposition methods.

Decomposition Methodology

In this section, we develop an approach to decomposing changes in earnings distributions that exploits employer-employee matched panel data. The aim is to decompose the change in earnings inequality observed between two time periods into portions attributable to changes in the composition of workers, changes in the composition of firms, and changes in the allocation of workers across firms. As previously mentioned, the decomposition methodology we outline allows for a descriptive analysis of the contributions of various factors to changes in earnings inequality and is not aimed at providing causal estimates.

Our empirical model relies on the human capital estimates in the Longitudinal Employer-Household Dynamics data, which are described in detail in the next section. These estimates are based on a model that follows that of Abowd et al. (1999) in assuming that the earnings of individual i at time t , denoted y_{it} , are a linear function of observed time-varying characteristics, \mathbf{x}_{it} , a time-constant individual fixed effect, θ_i , a time-constant fixed effect for the firm j at which i is employed at time t , $\psi_{j(i,t)}$, and an error term, ε_{it} :

$$y_{it} = \mathbf{x}_{it}\beta + \theta_i + \psi_{j(i,t)} + \varepsilon_{it}. \quad (1)$$

In this model, no assumptions are made about the relationship between the variables in \mathbf{x}_{it} and the worker and firm effects. The random error component ε_{it} , however, is assumed to be uncorrelated with \mathbf{x}_{it} , θ_i , and $\psi_{j(i,t)}$.

The individual fixed effect θ_i captures the portable component of an individual's earnings, or that component that belongs to an individual as he or she moves from job to job and that is separate from the type of firm for which he or she works. This person effect, which represents all time invariant observable and unobservable individual heterogeneity, is our measure of human capital. In interpreting the person effect, several remarks should be made. First, the human capital measure is not simply a ranking of the earnings of the worker, precisely because earnings include both person and firm effects. Second, the person effect reflects the influence of any time-invariant personal characteristics. Thus, for each individual, it will reflect factors including educational attainment, other observable accumulated skill correlates, and unobserved dimensions of ability. At the same time, it abstracts from firm-specific factors that may be present in measures based upon observable characteristics.

The firm effect $\psi_{j(i,t)}$, meanwhile, captures the extent to which the firm at which a worker is employed pays above or below average earnings (after controlling for workforce characteristics). The firm effect may reflect many factors, including capital intensity, rent sharing, firm-specific human capital, compensating differentials, or unionization effects (Abowd et al., 2008). Changes in the distribution of firms that pay relatively high and low wages over time and across sectors due to entry and exit could contribute to changes in earnings inequality. Moreover, changes in the joint distribution of worker and firm effects, or changes in the extent to which high wage workers match with high wage firms, could affect the distribution of earnings (Abowd et al., 1999; Lane, 2009).

For the decomposition, we discretize values of $\mathbf{x}\beta$ and allow for different joint distributions of the worker effect, the firm effect, and the error term for each distinct value of $\mathbf{x}\beta$.² Let q_t^r be the proportion of workers with the r^{th} distinct value of $\mathbf{x}\beta$, $\mathbf{x}\beta^r$, at time t , and let $g_t^r(\theta, \psi, \varepsilon)$ be the joint distribution of $(\theta, \psi, \varepsilon)$ for the r^{th} value of $\mathbf{x}\beta$ at time t , where $t=1, 2$ and $r=1, 2, \dots, R$.³

In order to introduce the decomposition technique, we first consider the simple case of decomposing the sources of change in average earnings between two periods.⁴ Mean earnings in period 1 are

$$\bar{y}_1 = \bar{\mathbf{x}}_1\beta + \bar{\theta}_1 + \bar{\psi}_1 + \bar{\varepsilon}_1, \quad (2)$$

or, recalling that q_t^r is the fraction of workers with the r^{th} distinct value of $\mathbf{x}\beta$,

$$\bar{y}_1 = \sum_{r=1}^R q_1^r [\mathbf{x}\beta^r + \bar{\theta}_1^r + \bar{\psi}_1^r + \bar{\varepsilon}_1^r]. \quad (3)$$

Similarly, average earnings in period 2 are

² Although we discretize values of $\mathbf{x}\beta$ for the decomposition, the time-varying characteristics used to estimate equation (1) include continuous variables and are described in detail in the data section.

³ Note that, given the discretization, $\mathbf{x}\beta^r$ represents the mean $\mathbf{x}\beta$ within each group $r=1, 2, \dots, R$.

⁴ Results from the general class of decomposition techniques into which our approach falls can be sensitive to the order of variables. In our application, several different orderings yielded similar results.

$$\bar{y}_2 = \bar{\mathbf{x}}_2\beta + \bar{\theta}_2 + \bar{\psi}_2 + \bar{\varepsilon}_2, \quad (4)$$

or

$$\bar{y}_2 = \sum_{r=1}^R q_2^r [\mathbf{x}\beta^r + \bar{\theta}_2^r + \bar{\psi}_2^r + \bar{\varepsilon}_2^r]. \quad (5)$$

To decompose the change in average earnings, we first consider what average earnings would have been in period 2 had the distribution of $\mathbf{x}\beta$ remained the same as in period 1. This entails substituting q_1^r for q_2^r in (5):

$$\bar{y}_2^x = \sum_{r=1}^R q_1^r [\mathbf{x}\beta^r + \bar{\theta}_2^r + \bar{\psi}_2^r + \bar{\varepsilon}_2^r]. \quad (6)$$

Next, we examine the counterfactual of what average earnings would be in period 2 if the distribution of $\mathbf{x}\beta$ were the same as in period 1 *and* if no workers entered or left the market. We denote market “entrants” by n , “stayers” by s , and “leavers” by l . We also let $p_k(w, f)$ be the proportion of worker-firm matches in period k for each type of worker (w) and each type of firm (f). For example, $p_2(s, s)$ equals the proportion of worker-firm matches in period 2 among firm and worker stayers (i.e., among firms and workers who were present in both periods 1 and 2). Similarly, $p_2(n, s)$, $p_2(s, n)$, and $p_2(n, n)$ denote the fraction of worker-firm matches in period 2 among new entrant workers and stayer firms, stayer workers and new entrant firms, and new entrant workers and new entrant firms, respectively. Analogously, worker-firm matches in period 1 are comprised of workers and firms who both stay until period 2, $p_1(s, s)$, workers who stay until period 2 and firms who leave by period 2, $p_1(s, l)$, workers who exit by period 2 and firms who stay until period 2, $p_1(l, s)$, and firms and workers who both exit by period 2, $p_1(l, l)$.

Using this notation, we can write the joint density of θ , ψ , and ε among workers with $\mathbf{x}\beta^r$ in period 2 as

$$g_2^r(\theta, \psi, \varepsilon) = p_2(s, s)g_2^{rs}(\theta, \psi, \varepsilon) + p_2(n, s)g_2^{ms}(\theta, \psi, \varepsilon) + p_2(s, n)g_2^{rn}(\theta, \psi, \varepsilon) + p_2(n, n)g_2^{mn}(\theta, \psi, \varepsilon), \quad (7)$$

or

$$g_2^r(\theta, \psi, \varepsilon) = p_2(s, s)g_2^{rs}(\psi, \varepsilon | \theta)g_2^{rs}(\theta) + p_2(n, s)g_2^{ms}(\psi, \varepsilon | \theta)g_2^{ms}(\theta) + p_2(s, n)g_2^{rn}(\psi, \varepsilon | \theta)g_2^{rn}(\theta) + p_2(n, n)g_2^{mn}(\psi, \varepsilon | \theta)g_2^{mn}(\theta). \quad (8)$$

The counterfactual of no worker entry or exit between periods 1 and 2 converts $g_2^r(\theta, \psi, \varepsilon)$ to

$$g_2^{rw}(\theta, \psi, \varepsilon) = p_2(s, s)g_2^{rs}(\psi, \varepsilon | \theta)g_2^{rs}(\theta) + p_2(s, n)g_2^{rn}(\psi, \varepsilon | \theta)g_2^{rn}(\theta) + R \times [p_1(l, s)g_1^{ls}(\theta) + p_1(l, l)g_1^{ll}(\theta)]g_2^{rs}(\psi, \varepsilon | \theta), \quad (9)$$

where $R = W_2/W_1$ and W_t equals the number of workers in period $t = 1, 2$. Notably, we have assumed here that worker leavers would have matched with firms in the same manner as worker stayers matched with firms that remained in the market. Therefore, the worker entry and exit component in the decomposition implicitly captures possible selection effects owing to, for example, a higher degree of negative assortative matching among those workers who leave relative to those who stay.

Given the counterfactual distribution for each r , we can calculate $\bar{\theta}_2^{rw}$, $\bar{\psi}_2^{rw}$, and $\bar{\varepsilon}_2^{rw}$.⁵ Then we can determine average earnings in period 2 had there been no change in time-varying observable characteristics and no change in workers' market participation:

$$\bar{y}_2^{xw} = \sum_{r=1}^R q_1^r [\mathbf{x}\beta^r + \bar{\theta}_2^{rw} + \bar{\psi}_2^{rw} + \bar{\varepsilon}_2^{rw}]. \quad (10)$$

Next, assuming no firm entry or exit produces the following counterfactual distribution:

⁵ For example, $\bar{\theta}_2^{rw} = \sum_{\theta} \sum_{\psi} \sum_{\varepsilon} \theta_{i2} g_2^{rw}(\theta, \psi, \varepsilon)$.

$$g_2^{rwf}(\theta, \psi, \varepsilon) = \left[\begin{array}{l} p_1(l, s)g_1^{ns}(\theta) + p_1(l, l)g_1^{nl}(\theta) \\ + p_1(s, s)g_1^{rs}(\theta) + p_1(s, l)g_1^{rl}(\theta) \end{array} \right] g_2^{rss}(\psi, \varepsilon | \theta), \quad (11)$$

where we have assumed that for those firms that actually left the market, had they not left, the distribution of the firm fixed effect ψ and error term ε conditional on the worker fixed effect θ would have been the same as that for stayers. Similar to the worker entry and exit component, the firm entry and exit component in the decomposition captures possible selection effects with respect to firm exit along these margins. Using equation (11), we can derive average earnings in period 2 conditional on there being no change in time-varying observable characteristics, no worker turnover, and no firm turnover:

$$\bar{y}_2^{xwf} = \sum_{r=1}^R q_1^r [\mathbf{x}\beta^r + \bar{\theta}_2^{rwf} + \bar{\psi}_2^{rwf} + \bar{\varepsilon}_2^{rwf}]. \quad (12)$$

How firms and workers match could also have changed between periods 1 and 2. That is, $g_2^{rss}(\psi, \varepsilon | \theta)$ and $g_1^{rss}(\psi, \varepsilon | \theta)$ could differ. Thus, in the next step of the decomposition, we assume that the matching mechanism between workers and firms in period 2 is the same as in period 1. Now,

$$g_2^{rss}(\psi, \varepsilon | \theta) = g_2^{rss}(\psi | \theta, \varepsilon) g_2^{rss}(\varepsilon | \theta), \quad (13)$$

and we can write the counterfactual distribution as

$$g_2^{rwfm}(\theta, \psi, \varepsilon) = \left[\begin{array}{l} p_1(s, s)g_1^{rs}(\psi | \theta, \varepsilon)g_1^{rs}(\theta) + p_1(l, s)g_1^{ns}(\psi | \theta, \varepsilon)g_1^{ns}(\theta) \\ + p_1(s, l)g_1^{rl}(\psi | \theta, \varepsilon)g_1^{rl}(\theta) + p_1(l, l)g_1^{nl}(\psi | \theta, \varepsilon)g_1^{nl}(\theta) \end{array} \right] g_2^{rss}(\varepsilon | \theta). \quad (14)$$

Using this distribution, average earnings in period 2 conditional on there being no change in observable time-varying characteristics, no change in the market participation of workers or firms, and no change in the manner in which workers are allocated across firms is

$$\bar{y}_2^{xwfm} = \sum_{r=1}^R q_1^r [\mathbf{x}\beta^r + \bar{\theta}_2^{rwfm} + \bar{\psi}_2^{rwfm} + \bar{\varepsilon}_2^{rwfm}]. \quad (15)$$

Finally, if we assume that the conditional distribution of ε is the same in period 2 as in period 1, this brings us back to equation (2). Therefore, the decomposition of the change in mean earnings is

$$\bar{y}_2 - \bar{y}_1 = (\bar{y}_2 - \bar{y}_2^x) + (\bar{y}_2^x - \bar{y}_2^{xw}) + (\bar{y}_2^{xw} - \bar{y}_2^{xwf}) + (\bar{y}_2^{xwf} - \bar{y}_2^{xwfm}) + (\bar{y}_2^{xwfm} - \bar{y}_1), \quad (16)$$

where the first difference in parentheses on the right-hand side equals the change in average earnings due to changes in the distribution of the variables in \mathbf{x} , the second difference is the change due to worker entry and exit, the third difference is the change due to firm entry and exit, the fourth difference is the change due to differences in matching between firms and workers, and the fifth difference is the difference due to changes in the distribution of residuals.

The residual component may capture several effects. Perhaps most importantly, it will reflect any systematic changes in the returns to observable or unobservable characteristics between periods 1 and 2. Our decomposition is designed to quantify the importance of turnover and sorting in bringing about changes in earnings distributions over time, and in effect assumes that the returns to characteristics are constant. To the extent that those returns do change systematically over the time period we consider in our application, it will be captured in the residual. Also, changes in the “quality” of worker-firm matches that are not due to merely the reallocation of different workers across different firms will show up in the residual component.

While here we present our approach in the specific context of decomposing changes in average earnings, an advantage of this decomposition methodology is that it can be performed for any statistic. In the analysis that follows, we consider changes in a variety of percentiles of the earnings distribution in an attempt to paint a more complete picture of what might be driving changes in inequality at different points in the distribution. Our unique data, which we describe in detail in the next section, permits us to isolate the importance of certain factors that have received relatively less attention to date, including in particular firm entry and exit and the matching of workers and firms.

Data

To decompose the sources of change in earnings inequality across and within industries, we take advantage of a database created by the Longitudinal Employer-Household Dynamics (LEHD) Program at the U.S. Census Bureau. These confidential data enable us to match workers with past and present employers, together with employer and worker characteristics (Abowd et al., 2009b). This database consists of quarterly records of the employment and earnings of almost all individuals from the unemployment insurance (UI) systems of a number of U.S. states in the 1990s and 2000s.

These data have been extensively described elsewhere (Haltiwanger et al., 2006; Abowd et al., 2009b), but it is worth noting several advantages of the LEHD data. Since the scope of the data is almost the universe of employers and workers in the private sector, the dataset is extremely large, and it is possible to follow workers as they move between employers and along the earnings distribution.⁶ The UI records have also been matched to internal administrative and survey data containing some limited demographic information on individuals.

Of particular importance given the focus of this study is the highly accurate reporting of both earnings and industry in the LEHD data. Hirsch and Schumacher (2004) point out that up to 30% of respondents to the CPS, the major source of information on earnings inequality in the literature, do not respond to income questions. As a result, their answers are imputed. In the LEHD data, earnings are quite accurately and universally reported by firms due to financial penalties for misreporting or failing to report. In addition, research comparing earnings and employer characteristics as reported by survey respondents to those recorded in administrative files suggests that workers not only often misreport earnings, but also frequently fail to identify their industry correctly even at the major industry level (Roemer, 2002; Stinson, 2002; Bound et al., 2001).

The LEHD data have several limitations. First, the data are not available for all states, and the amount of historical data varies by state. As such, we isolate attention to four large states (California, Illinois, Maryland, and North Carolina) for which we have data for the period 1992-2003.⁷ Based on County Business Patterns data, these four states accounted for approximately

⁶ See Stevens (2002) for a description of UI data coverage.

⁷ The choice of the years 1992 and 2003 was determined largely by practical considerations and data availability. However, both years are at a similar point in the economic cycle and provide a comparison to Autor et al. (2008)'s results for a similar time period using CPS data.

one-fifth of total U.S. employment in 2003 and, taken together, are similar in industrial composition as the nation as a whole.⁸

While the states in our sample are broadly representative, including only a subset of the U.S. poses an additional issue in our particular application. Some workers whom we classify as new entrants to an industry (as a result of not being observed in our sample in the initial period) may have in fact transitioned within industry but originated in a state outside our sample. Similarly, some workers whom we label as exits from an industry may in fact transition within industry, only to a state outside our sample. This may lead us to overstate industry turnover and could affect our estimates of the contribution of changes in worker composition to changes in earnings inequality, especially if interstate migrants are systemically different than non-migrants. This potential selection issue should be borne in mind in the interpretation of the results.

Another limitation of the LEHD data is that they lack information on hours worked, which makes it impossible to calculate an hourly wage rate or determine full-time or part-time status. Therefore, our measure of earnings is real (2003 dollars) annualized earnings, which is the full-year equivalent of the hourly real wage. To calculate this measure, we use only earnings from each worker's dominant employer, or that employer that contributes the most to the worker's total earnings in each year. To eliminate workers with minimal attachment to the labor market and those employed only part of a quarter (and hence whose reported earnings represent compensation for an indeterminate amount of time), we also use only workers who have real earnings of at least \$250 in at least one quarter of the year and who are full-quarter employed, where being full-quarter employed in quarter t is defined as having an employment history with positive earnings for quarters $t - 1$, t , and $t + 1$. Abowd et al. (2009b) contains further details on constructing samples based on LEHD data.

The LEHD data include the unique measures of human capital and firm pay policies discussed in the previous section. In particular, based on equation (1), LEHD staff have estimated individual and firm effects, θ and ψ , controlling for observable time-varying characteristics \mathbf{x} . These time-varying characteristics include a quartic in labor force experience, a set of work history dummies, and annual national estimates of per capita earnings and

⁸ For example, manufacturing represented 12.5% of total employment in our sample of four states in 2003, compared to 12.7% nationally. Meanwhile, retail trade represented 12.4% of total employment in our sample of states in 2003, only slightly less than the 13.3% of total employment retail trade represented nationwide.

unemployment rates (to capture macroeconomic trends), all interacted with gender.⁹ While we focus on only two years (1992 and 2003) and four states with sufficient historical data in the decomposition, estimates of β , θ , and ψ are derived from running equation (1) using annual data for an unbalanced sample of 22 states. Parameter estimates as well as goodness-of-fit measures from this regression appear in Appendix Table A1.

In accordance with the methodology laid out in the previous section (and for computational reasons), we discretize the values of $\mathbf{x}\hat{\beta}$ prior to performing the decomposition.¹⁰ Incorporating detailed information on earnings with observable time-varying characteristics as well as estimates of θ and ψ , the decomposition allows us to quantify the importance of changes in workforce composition (both due to changes in observable characteristics among stayers and due to worker entry and exit), changes firm composition, and changes in the allocation of different workers across different firms to changes in the overall distribution of earnings. The estimated residual is also informative regarding the importance of other potential factors that might drive changes in earnings inequality, such as changes in the returns to skills.

While previous studies have made use of observable measures of human capital such as education and experience, rarely have researchers been able to control adequately for unobservable productive characteristics such as ability or interpersonal skills. Changes in the unobservable characteristics of workforces across industries as well as changes in how workers with different levels of unobservable skills are allocated across different types of firms within industries could help to explain changes in earnings inequality. Similarly, controlling for unobservable firm characteristics that might affect compensation structures has been difficult in

⁹ In principle, we could also separately include observed time-varying firm characteristics, such as firm size and age. While an interesting potential extension to our analysis, it is unlikely that changes in such characteristics made a large contribution to changes in overall earnings inequality within or across industries during our sample period. Cross-sectional distributions of firm size and age tend not to vary substantially over time (Cabral and Mata, 2003; Angelini and Generale, 2008).

¹⁰ Note that while the description of the decomposition in the previous section presumed our having the true values of β , θ , ψ , and ε , we use estimates of each in the actual decomposition. We discretized $\mathbf{x}\hat{\beta}$ by breaking the range into 100 mutually exclusive intervals and assigning the midpoint $\hat{\beta}$ value to each observation that falls within the interval. This method is applied for all intervals except the lowest and highest intervals (which are unbounded). For the highest (lowest) interval, we assign a value that equals the average of the lower (higher) boundary value and the highest (lowest) observed value in the (industry) sample. As described in the methodology section, we estimate the continuous distribution of $(\hat{\theta}, \hat{\psi}, \hat{\varepsilon})$ for each category of $\mathbf{x}\hat{\beta}$. However, for ease of exposition, we refer to these parameters as if we have the true values in subsequent sections of the paper.

the past given data limitations. The LEHD dataset allows us to quantify the contributions of changes in such unobservable characteristics to changes in the distribution of earnings.

For the purposes of our analysis, we restrict attention to all private establishments and use the Standard Industrial Classification (SIC) to identify major sectors of the economy. In particular, we focus on agriculture, forestry, and fishing (which we henceforth refer to simply as agriculture) (SICs 01-09); mining (SICs 10-14); construction (SICs 15-17); manufacturing (SICs 20-39); transportation, communication, and utilities (TCU) (SICs 40-49); wholesale trade (SICs 50-51); retail trade (SICs 52-59); finance, insurance, and real estate (FIRE) (SICs 60-67); and services (SICs 70-89). We exclude public administration due to incomplete data coverage in that sector.

Basic Facts

In this section, we present basic descriptive statistics regarding overall wage inequality, workforce composition, firm turnover, and the allocation of workers across firms within and between industries in our sample. In the next section, we turn to our decomposition results, which shed light on the relative importance of different factors in explaining changes in earnings distributions over time.

Changes in Inequality

We first consider basic characteristics of the earnings distribution in each of the nine major industries and in all industries combined by calculating levels and log differences in real annualized earnings at different percentiles. Table 1 shows the 90th, 50th (median), and 10th percentile of earnings in 2003 and the 90-10, 90-50, and 50-10 log earnings differences by sector and for all sectors together.

An examination of the first three columns of Table 1 reveals that there are substantial differences in earnings across industries. For example, median earnings are over twice as high in mining as in the agricultural sector, and similar differences hold at both the 90th and 10th percentiles. Earnings at the high and low end of the distribution also vary greatly across sectors. The highest 90th percentile earnings are found in FIRE (\$114,428), while the lowest 10th percentile earnings are in retail and agriculture (both under \$10,000). The distribution of earnings also varies across sectors; the differences are starkest for the 90-10 and 90-50 log

earnings differences. The 90-10 log earnings gap is highest in services, followed by FIRE, wholesale trade, retail trade, and manufacturing. These same five industries also had the highest 90-50 log earnings differences in 2003. In contrast, inequality at the lower end of the earnings distribution does not vary as much across industries, though services had the largest 50-10 log earnings difference.

As the last row of Table 1 shows, the 90-10 earnings gap increased by 0.06 log points between 1992 and 2003 across all sectors in the sample. Looking across industries, though, earnings inequality as measured by the 90-10 log earnings difference declined in four industries (agriculture, mining, construction, and retail trade) and increased in the remaining five industries (manufacturing, TCU, wholesale trade, FIRE, and services). The latter five industries accounted for about three-fourths of total employment in the sample each year, with services alone accounting for roughly one-third of total employment. The most marked increases in inequality, though, were in FIRE, manufacturing, and TCU.

In order to compare changes in earnings inequality in the upper and lower tails of the earnings distribution, Table 1 also breaks out changes in the 90-50 and 50-10 log earnings gaps between 1992 and 2003 by industry. There are clear differences in trends in inequality across sectors. In three of the four industries in which overall inequality (the 90-10 log earnings difference) declined, much if not all of the decrease was in the lower half of the earnings distribution (the 50-10 log earnings gap). Only in mining was there much of a decline in the upper half (the 90-50 log difference). In contrast, earnings inequality in the lower end of the distribution did not increase in the five industries in which overall inequality increased. The increase in the 90-10 log earnings difference in manufacturing, TCU, wholesale trade, FIRE, and services was driven almost entirely by an increase in the spread between the 90th and the 50th percentiles of earnings.

Overall, the results reported in Table 1 are consistent with Autor et al. (2008), who find using the CPS that economy-wide, the 90-50 earnings gap grew during the 1990s while the 50-10 difference leveled off after about 1987. However, given the marked differences in changes in upper and lower tail inequality across sectors evident in Table 1, looking only at economy-wide trends may miss an important part of the story. The heterogeneity across sectors suggests that different labor market or institutional factors may have affected different industries in different

ways. In what follows, we discuss some of these factors and attempt to identify the relative importance of each.

Workforce Composition

One possible reason for changes in the earnings distribution in any given industry is that the composition of the workforce has changed over time. Table 2 suggests that there is ample potential for such changes to occur. In manufacturing, for example, of the more than five million workers who were employed in 1992, 2003, or both years, 44% were only in the industry in 1992, 35% were only in the industry in 2003, and 21% were there (but not necessarily in the same firm) in both years. That implies that only about 32% of all workers employed in the industry in 1992 were still there 11 years later. As one might expect, turnover in the workforce is even more substantial in the wholesale and retail trade industries, where less than a quarter of the workers in each industry in 1992 were still in the same industry in 2003.¹¹

The considerable turnover within industries was accompanied by dramatic changes in the distribution of workers across industries, as Table 3 shows. While the mining and manufacturing industries experienced double-digit drops in employment in percentage terms between 1992 and 2003, construction and services industries witnessed over 40% increases in employment over the same period in our sample of states. However, these structural changes did not coincide with enormous shifts in the age, gender, and human capital distributions of workers within industries. In other words, even as some industries shrank and others expanded, the workforce characteristics of each changed little. Mining and manufacturing, for example, remained over two-thirds male and skewed toward older workers. In contrast, industries such as FIRE and services continued to employ more females and younger workers. Similarly, although the average human capital level of the workforce increased in all industries, the swings are not substantial.

Firm Turnover

Another possible reason for changes in earnings inequality is changes in the types of firms that are hiring workers. We examine this possibility in Table 4, which can be read the same way

¹¹ As previously discussed, these figures may overstate the extent of turnover in each industry given that some individuals that we classify as entrants may in fact have moved within industry from a state outside our sample. Similarly, some individuals we classify as exits may have moved within industry to a state outside our sample.

as Table 2. For example, in manufacturing, of the more than 100,000 firms that employed individuals in 1992, 2003, or both years, about 36% were only in the industry in 1992, 37% were only in the industry in 2003, and only 27% were there in both years. Therefore, only about 43% of manufacturing firms in existence in 1992 survived until 2003. Survival rates are even lower in retail trade, where only about one-third of firms operating in 1992 were still alive in 2003. Such high levels of firm turnover are not surprising given that close to two-thirds of all new establishments fail within ten years.¹² To the extent that entering, exiting, and continuing firms vary in their characteristics and compensation policies, firm turnover could lead to changes in earnings inequality over time.

Sorting

Another potential source of change in earnings inequality is changes in the joint distribution of worker and firm characteristics. In other words, changes in the allocation of different workers across different firms could affect the distribution of earnings.

In terms of the two-way fixed effects model previously described, changes in the distribution of earnings may be due to changes not only in θ and ψ independently, but also in the joint distribution of θ and ψ . For example, over time it may be the case that high θ individuals are more likely to work at high ψ firms and that low θ individuals are more likely to work at low ψ firms. All else being equal, this trend would tend to increase earnings inequality over time.¹³

Recent work suggests that job tenure and long-term employment relationships have been on the decline in recent decades (Farber, 2010). Higher rates of job mobility potentially open the door to more assortative matching. Using estimated values of θ and ψ from the model, Figure 1 plots their joint distributions for 1992 and 2003. As the mass points in the left and right corners of each figure suggest, there is a tendency for low θ workers to be employed with low ψ firms and for high θ workers to be employed with high ψ firms. While this pattern held in both 1992 and 2003, a closer look at the extremes of the joint distribution of individual and firm effects reveal an interesting trend suggestive of greater assortative matching. The expected average θ (deviated from year means) more than tripled from 0.75 to 2.50 between 1992 and 2003 for the

¹² See the Business Employment Dynamics statistics at the Bureau of Labor Statistics (<http://www.bls.gov/bdm>).

¹³ While some have found little evidence of any cross-sectional correlation between θ and ψ (Abowd et al., 1999), others have found some evidence of a positive correlation (Abowd et al., 2002; Woodcock, 2008). More recent work suggests that there is assortative matching in the labor market, but that the effects of sorting are attenuated due to a lack of heterogeneity among workers and firms in the market (Abowd et al., 2009a).

top 5% of firms as ranked by their ψ s. Meanwhile, the expected average θ remained roughly constant for the bottom 5% of firms at -0.75. Thus, individuals with very high skill levels were more likely in 2003 than 1992 to be paired with firms with high pay policies, while those with low skill levels were not. Whether this finding is a result of entry and exit of different types of firms and workers or is due to a reshuffling of workers across different firms, however, requires further investigation. In the next section, we apply our decomposition approach to examine different possible sources of changes in inequality and to quantify their relative importance.

Decomposition Results

Results for All Industries

Using the LEHD data, we first decompose the change in the earnings distribution for all nine sectors combined between 1992 and 2003. In decomposing changes in the earnings distribution for all sectors, we must account for the change in the employment distribution across sectors over time; the contribution of such changes is considered first in the decomposition. That is followed by estimates of the contributions of worker entry and exit, changes in observable time-varying characteristics, firm entry and exit, and changes in the distribution of worker unobserved attributes (θ) for a given firm pay policy (ψ) – i.e., sorting. That leaves us with a residual component that could reflect factors not considered explicitly in the decomposition, including changes in the returns to characteristics.

The results of the decomposition appear in Table 5. In the table, we decompose the sources of change in earnings at the 90th, 50th, and 10th percentiles of the earnings distribution as well as the sources of change in earnings inequality as measured by the 90-10 log earnings difference, the 90-50 log difference, and the 50-10 log difference. Panel (a) shows the decomposition in levels, while panel (b) provides the implied relative contribution of each component to the change in the statistic. In other words, panel (b) reports the change in the statistic when the factor is held at its 1992 value (i.e., assuming that there was no change in the sectoral distribution of employment, the market participation of workers, time-varying characteristics, the market participation of firms, and the conditional distribution of worker-firm matches).

The net changes in earnings levels and inequality reported in column (8) of Table 5, which echo those reported in the last row of Table 1, mask considerable flux in the earnings distribution due to changes in underlying factors, which are spelled out in the intervening columns. First, an

examination of column (2) reveals that sectoral changes in the distribution of employment contributed to small increases in both the 90-50 and 50-10 log earnings differences, due largely to a reallocation of employment into industries with relatively low earnings (the 10th and 50th percentiles of earnings declined relative to the 90th). For instance, had there been no change in the sectoral distribution of employment, the 90-50 log earnings difference would have been 0.975 instead of the actual 0.981 (see panel (a)). Thus, changes in the sectoral distribution of employment contributed to an increase in the 90-50 log earnings difference of 0.006 (see panel (b)). Such small impacts of sectoral changes in the distribution of employment on inequality are consistent with the findings of Firpo et al. (2007). Meanwhile, as is evident in column (3), the marginal contribution of worker entry and exit conditional on changes in the sectoral distribution of employment had virtually no impact on the 50-10 log earnings difference and resulted in a slight decrease in the 90-50 log earnings difference.

More important were changes in observable time-varying characteristics, which as column (4) of Table 5 shows, led to decreases in both the 50-10 and 90-50 log earnings differences. These decreases came about as changes in observable characteristics boosted earnings at the 10th percentile more so than at the 50th and 90th. The entry and exit of firms (column (5)) reinforced the effect of changes in time-varying characteristics, leading to sizable decreases in the 50-10 and 90-50 log earnings differences. Firm turnover had its largest effect at the lower end of the earnings distribution, propping up earnings at the 10th percentile.

However, these effects were largely offset by the impact of sorting among workers and firms and the residual component, as columns (6) and (7) reveal. The reallocation of workers across firms clearly played a large role in determining changes in the shape of the overall earnings distribution, with changes in the joint distribution of worker skill and firm pay policies tending to work against individuals at the lower end of the earnings distribution and favoring those at the upper end. The apparent increase in assortative matching is consistent with skill-biased technological change to the extent that such change might raise the relative returns for capital-intensive firms to hire highly skilled workers.

The contribution of the residual component to changes in inequality is also consistent with skill-biased technological change. To the extent that the residual component reflects changes in the returns to skills, its positive contribution to earnings inequality is in line with the results of Autor et al. (2008) and others who have suggested that skill-biased technological change has

avored those at the upper end of the earnings distribution and acted to increase the 90-50 log earnings difference more so than the 50-10 difference in recent decades.

Results by Industry

In Tables 6 and 7, we break out the results by industry. For the sake of brevity, we present only the implied changes in earnings levels and inequality driven by each factor. Table 6 presents results for the four industries in which inequality (as measured by the 90-10 log earnings difference) decreased between 1992 and 2003 (agriculture, mining, construction, and retail trade). Table 7 presents results for the five industries in which inequality increased (manufacturing, TCU, wholesale trade, FIRE, and services).

Looking first at the four industries in which overall inequality declined, there are several striking similarities and differences in the factors driving changes in the distributions of earnings over time. Despite the high levels of worker churning across all industries, column (2) of Table 6 suggests that entering and exiting workers were of roughly the same average skill level (θ), resulting in basically no change in inequality. That this is true in every industry suggests that, by and large, workforce quality within each industry is quite persistent, which is consistent with work by Haltiwanger et al. (2006). An analysis of the third column of Table 6 reveals that, holding θ constant, changes in time-varying observable characteristics acted to decrease earnings inequality in three of the four industries that experienced declines in inequality between 1992 and 2003. More specifically, changes in time-varying characteristics generally led to higher earnings at both ends of the distribution, but with a larger impact at the 10th percentile than at the 90th (except in retail).

The entry and exit of firms had an enormous impact on the earnings distribution, as column (4) of Table 6 shows. Holding workforce composition constant, if no firm entry or exit had occurred in the mining industry between 1992 and 2003, the 90-10 log earnings gap would have swung by 121 log points, most of which occurred between the 50th and 10th percentiles. Notably, across all industries, firm entry and exit typically acted to increase earnings at the bottom end of the distribution more than at the top, resulting in a decline in the 90-10 log earnings difference in each industry.

The effect of the sorting of workers and firms is evident in column (5) of Table 6. Changes in the extent of assortative matching had a negative impact on earnings for workers at the lower end

of the distribution in each the four industries with declining overall inequality. Meanwhile, it acted to raise 90th percentile earnings in three of the four industries. Still, for these four industries, the sorting effect was largely overshadowed by the effect of firm entry and exit.

The residual component was relatively small in each of the industries that experienced declining inequality, as column (6) of Table 6 shows. Holding workforce and firm composition as well as the allocation of workers across firms constant, other factors not included in the decomposition, which could include changes in the returns to observable and unobservable characteristics, worked to increase earnings inequality. It did so largely by depressing earnings at the lower end of the distribution relative to the upper end. Yet the effects are quite small, which is perhaps not surprising given that none of the four industries in Table 6 were likely subject to substantial skill-biased technological change.

Table 7 reports the decompositions for the five industries in which overall earnings inequality (again measured as the 90-10 log earnings gap) increased. As was the case for the declining-inequality industries, changes in the distribution of θ due to the entry and exit of workers had little effect on the earnings distributions in these sectors, as can be seen in column (2) of Table 7. Meanwhile, column (3) shows that changing time-varying observable characteristics lowered inequality by raising earnings more at the bottom than at the top of the earnings distribution in each of the industries. Interestingly, however, the magnitude of the impact of changing observable characteristics on earnings is quite different across industries. For example, changes in such characteristics affected the 10th and 50th earnings percentiles in manufacturing by about 20-21 log points each, compared with 11-13 log points in services.

As column (4) of Table 7 reveals, the effect of firm entry and exit was substantial in these five industries. In general, turnover among businesses led to a decrease in earnings inequality by bolstering earnings at the bottom more than at the top end of the distribution. In wholesale trade, the 90th percentile of earnings dropped considerably due to firm entry and exit.

Column (5) of Table 7 indicates that sorting of workers among firms generally led to an increase in inequality in the five industries that experienced an increase in overall inequality. Earnings at the bottom of the distribution were much lower due to sorting, leading to a rise in all three inequality measures in all five industries. This effect was especially large in manufacturing and services. The contributions of the residual component (column (6)) followed a similar pattern, also tending to boost overall inequality in each of the five industries in Table 7.

However, perhaps reflecting skill-biased technological change that benefited workers at the upper end of the distribution more than at the lower end, the residual component tended to increase the 90-50 log earnings difference more so than the 50-10 difference.

Taken together, the decomposition results in Table 6 and Table 7 show that, while trends in overall inequality (as measured by the 90-10 log earnings difference) diverged across industries, similar factors were often at work beneath the surface. Worker entry and exit had little effect on earnings inequality measures despite high levels of worker churning in the economy. Changes in time-varying observable characteristics within each industry acted to increase earnings at all levels, but tended to have a larger impact at the lower end of the distribution. Firm entry and exit and the sorting of workers across firms had larger effects on earnings distributions across industries, with the former acting to decrease inequality and the latter acting to increase it in most industries. The residual component, which captures changes in returns to characteristics and other factors not accounted for in the decomposition, reinforced the effect of sorting, acting to further increase inequality. Nonetheless, despite the similarities in underlying factors, the size of these effects differed considerably across industries. Perhaps even more strikingly, even in industries in which there were small net changes in earnings distributions, there were often very large, offsetting effects from the underlying forces driving changes in inequality over time.

Minimum Wage Legislation

One of our central findings is the very large and offsetting effects of firm turnover and sorting among workers and firms. An examination of Tables 5-7 reveals that the largest manifestation of this phenomenon occurs in the 10th percentile of the distribution. This result raises the possibility that the adjustment is due to the substantial increase of the minimum wage in California. All four states in the sample had minimum wages of \$4.25 per hour in 1992 and experienced incremental federally-mandated increases to \$5.15 per hour by September 1997. However, California increased its minimum wage further thereafter, raising it to \$5.75 by March 1998, \$6.25 by January 2001, and \$6.75 by January 2002. In the meantime, the minimum wage in Illinois, Maryland, and North Carolina remained at \$5.15 per hour between 1997 and 2003. Given this, we would expect that if the minimum wage does have an effect on the earnings distribution, it would be apparent in the contribution of firm entry and exit on the earnings distribution for California in particular. The 1992 California firms included in the counterfactual

would have been at a substantially lower minimum wage, and thus have lower estimated firm fixed effects, than the 2003 California firms.

We examine the possible influence of the minimum wage legislation in California in Table 8, where we present results of the decomposition for retail trade for a sample that excludes California. Of all nine major industries, the retail trade industry employs the largest fraction of individuals working at or below the federal minimum wage (approximately 8% according to CPS data for 2002).¹⁴ In the retail trade industry, holding worker market participation and time-varying observable characteristics at their 1992 levels, the counterfactual had there been no entry and exit of firms is that log earnings at the 10th percentile of the distribution would increase 0.48 if California is included (see Table 6). Without California, the same counterfactual is that the log earnings would rise by 0.91 (see Table 8). At the 50th percentile, the effect of firm entry is 0.13 whether California is included in the sample or not. This pattern runs counter to our expectation that the minimum wage hikes would have tended to increase the ψ s of entering firms at the low end of the earnings distribution in California. Instead, it seems that firm entry actually had a larger positive effect on earnings at the low end of the distribution in Illinois, Maryland, and North Carolina, where there were no state-mandated changes in the minimum wage, than in California, where there was. While not a direct test of the impact of the minimum wage on earnings inequality over the sample period, these results suggest that we can rule out minimum wage legislation as the main explanation for our results with respect to the key role of firm entry and exit in driving changes in the earnings distribution.

Conclusion

In this paper, we use linked employer-employee data from the Longitudinal Employer-Household Dynamics Program at the U.S. Census Bureau to explore changes in earnings distributions across sectors of the economy. We investigate how changes in workforce composition, firm entry and exit, and the matching of workers and firms affect economy-wide and industry-specific earnings distributions.

¹⁴ Under SIC codes, retail trade includes eating and drinking places, which employ a large fraction of those individuals working at or below the minimum wage. About 62% of all workers earning the federal minimum wage or less were employed in the retail trade industry in 2002.

Our decomposition results suggest that even in industries in which there was very little change in the aggregate earnings distribution between 1992 and 2003, there were enormous, albeit offsetting, changes in the factors contributing to changes in that distribution. Similar factors were at work in industries with declining inequality as well as those with increasing inequality. The magnitudes of these effects, however, varied considerably.

We find that worker entry and exit had very little impact on changes in earnings distributions between 1992 and 2003 for the industries examined. In other words, although worker turnover rates were high across industries, the average characteristics of industry workforces remained, by and large, very similar. Meanwhile, changes in time-varying observable characteristics tended to shift the earnings distributions of all industries to the right. In every industry but one, such changes also worked to decrease earnings inequality, in each case primarily by increasing earnings at the bottom of the earnings distribution relative to the top end.

The net impact of firm entry and exit was to reduce the dispersion of earnings. In nearly all industries, firm turnover acted to increase earnings at the bottom end of the distribution more than at the top. This effect persisted even after excluding from the sample one state that experienced a sizable increase in its minimum wage during the sample period, which might otherwise be expected to account for some of the large increases in earnings attributed to firm turnover at the lower end of the distribution in certain industries. While our results do not imply that changes in the minimum wage have no effect on the earnings distribution, they do suggest that their effects do not manifest themselves through changes in the composition of firms in affected industries.

Meanwhile, sorting of workers and firms over time tended to increase the dispersion of industry earnings distributions between 1992 and 2003, with high wage workers increasingly finding their way to high wage firms. This trend has worked to increase earnings inequality. Though the mechanism driving this sorting is unclear, it is consistent with skill-biased technological change to the extent that such change might increase the relative returns for capital-intensive firms to hiring highly skilled workers. Also consistent with skill-biased change is the substantive role of the residual component in the decomposition, which could reflect changes in the returns to observable and unobservable characteristics and act to increase inequality more at the upper end of the earnings distribution than the lower end.

Overall, we find that underlying even very small changes in earnings distributions over time are potentially very large but offsetting effects of firm turnover and the sorting of workers across firms. The extensive amounts of worker and firm reallocation in the U.S. economy, which Haltiwanger et al. (2006) and others have shown to have important implications for productivity, also clearly play a key role in shaping the distributions of earnings within and across industries. In particular, the entry and exit of firms and the sorting of workers across firms based on underlying worker skills are important determinants of changes in earnings inequality over time.

REFERENCES

- Abowd, John, Francis Kramarz, and David Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67(2): 251-333.
- Abowd, John, Francis Kramarz, Sébastien Pérez-Duarte, and Ian Schmutte. 2009a. "A Formal Test of Assortative Matching in the Labor Market." NBER Working Paper No. 15546.
- Abowd, John, Julia Lane, and John Haltiwanger. 2008. "Wage Structure and Labor Mobility in the United States." In *The Structure of Wages: An International Comparison*, edited by Edward Lazear and Kathryn Shaw, pp. 81-100. Chicago: University of Chicago Press.
- Abowd, John, Paul Lengeremann, and Kevin McKinney. 2002. "The Measurement of Human Capital in the U.S. Economy". Technical Paper No. TP-2002-09, Longitudinal Employer-Household Dynamics Program, U.S. Census Bureau.
- Abowd, John, Bryce Stephens, Lars Vilhuber, Fredrik Andersson, Kevin McKinney, Marc Roemer, and Simon Woodcock. 2009b. "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators." In *Producer Dynamics: New Evidence from Micro Data*, edited by Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts, pp. 149-230. Chicago: University of Chicago Press.
- Acemoglu, Daron. 2002. "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature* 40(1): 7-72.
- Angelini, Paolo, and Andrea Generale. 2008. "On the Evolution of Firm Size Distributions." *American Economic Review* 98(1): 426-438.
- Autor, David, Lawrence Katz, and Melissa Kearney. 2008. "Trends in U.S. Wage Inequality: Reassessing the Revisionists." *Review of Economics and Statistics* 90(2): 300-23.
- Baily, Martin, Charles Hulten, and David Campbell. 1992. "Productivity Dynamics in Manufacturing Plants." *Brookings Papers on Economic Activity, Macroeconomics*: 187-267.
- Bernard, Andrew, and J. Bradford Jensen, 2000. "Understanding Increasing and Decreasing Wage Inequality." In *The Impact of International Trade on Wages*, edited by Robert C. Feenstra, pp. 227-268. Chicago: University of Chicago Press.
- Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement Error in Survey Data." In *Handbook of Econometrics* 5, edited by James Heckman and Edward Leamer, pp. 3705-3843. Amsterdam: North-Holland.
- Burgess, Simon, Julia Lane, and David Stevens. 2000. "Job Flows and Worker Flows in the Life Cycle of the Firm." *Oxford Bulletin of Economics and Statistics* 62(0): 885-908.
- Cabral, Luís, and José Mata. 2003. "On the Evolution of the Firm Size Distribution: Facts and Theory." *American Economic Review* 93(4): 1075-1090.
- Card, David, and John DiNardo. 2002. "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles." *Journal of Labor Economics* 20(4): 733-783.
- Danziger, Sheldon, and Peter Gottschalk. 1995. *America Unequal*. New York: Russell Sage Foundation.
- Davidson, Carlos, and Michael Reich. 1988. "Income Inequality: An Inter-Industry Analysis." *Industrial Relations* 27(3): 263-286.
- Davis, Steven, John Haltiwanger, and Scott Schuh. 1996. *Job Creation and Destruction*. Cambridge, MA: MIT Press.
- DiNardo, John, Nicole Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica* 64(5): 1001-44.

- Farber, Henry S. 2010. "Job Loss and the Decline in Job Security in the U.S." In *Labor in the New Economy*, edited by Katherine Abraham, James Spletzer, and Michael Harper, pp. 223-266. Chicago: University of Chicago Press.
- Firpo, Sergio, Nicole Fortin, and Thomas Lemieux. 2007. "Decomposing Wage Distributions Using Recentered Influence Function Regressions." Mimeo. University of British Columbia.
- _____. 2009. "Unconditional Quantile Regression." *Econometrica* 77(3): 953-973.
- Fortin, Nicole, and Thomas Lemieux. 1997. "Institutional Changes and Rising Wage Inequality: Is There a Linkage?" *Journal of Economic Perspectives* 11(2): 75-96.
- Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2010. "Decomposition Methods in Economics." NBER Working Paper No. 16045.
- Foster, Lucia, John Haltiwanger, and C.J. Krizan. 2001. "Aggregate Productivity Growth: Lessons from Microeconomic Evidence." In *New Developments in Productivity Analysis*, edited by Charles R. Hulten, Edwin R. Dean, and Michael J. Harper, pp. 303-372. Chicago: University of Chicago Press.
- Griliches, Zvi, and Haim Regev. 1995. "Firm Productivity in Israeli Industry, 1979-1988." *Journal of Econometrics* 65(1): 175-203.
- Haltiwanger, John, Julia Lane, and James Spletzer. 2006. "Wages, Productivity and the Dynamic Interaction of businesses and Workers." *Labour Economics* 14(3): 575-602.
- Hirsch, Barry, and Edward Schumacher. 2004. "Match Bias in Wage Gap Estimates Due to Earnings Imputation." *Journal of Labor Economics* 22(3): 689-722.
- Juhn, Chinhui, Kevin Murphy, and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy* 101(3): 410-442.
- Katz, Lawrence, and David Autor. 1999. "Changes in the Wage Structure and Earnings Inequality." In *Handbook of Labor Economics*, Volume 3, Part 1, edited by Orley Ashenfelter and David Card, pp. 1463-1555. Amsterdam: North-Holland.
- Lane, Julia. 2009. "Inequality and the Labour Market – Employers." In *The Oxford Handbook on Economic Inequality*, edited by Wiemer Salverda, Brian Nolan, and Timothy Smeeding, pp. 204-229. Oxford: Oxford University Press.
- Lee, David. 1999. "Wage Inequality in the United States During the 1980s: Rising Dispersion or Falling Minimum Wage?" *Quarterly Journal of Economics* 114(3): 977-1023.
- Lemieux, Thomas. 2006. "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?" *American Economic Review* 96(3): 461-98.
- 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* 30(3): 1333-81.
- Machado, José A. F., and José Mata. 2005. "Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression." *Journal of Applied Econometrics* 20(4): 445-465.
- Mortensen, Dale. 2003. *Wage Dispersion: Why Are Similar Workers Paid Differently?* Cambridge, MA: MIT Press.
- Roemer, Marc. 2002. "Using Administrative Earnings Records to Assess Wage Data Quality in the March Current Population Survey and the Survey of Income and Program Participation." Technical Paper No. TP-2002-22, Longitudinal Employer-Household Dynamics Program, U.S. Census Bureau.
- Stevens, David. 2002. "Employment that is not Covered by State Unemployment Insurance Laws." Technical Paper No. TP-2002-16, Longitudinal Employer-Household Dynamics Program, U.S. Census Bureau.

- Stinson, Martha. 2002. "Estimating Measurement Error in SIPP Annual Job Earnings: A Comparison of Census Survey and SSA Administrative Data." Technical Paper No. TP-2002-24, Longitudinal Employer-Household Dynamics Program, U.S. Census Bureau.
- Woodcock, Simon. 2008. "Wage Differentials in the Presence of Unobserved Worker, Firm, and Match Heterogeneity." *Labour Economics* 15(4): 771-793.

TABLE 1
EARNINGS LEVELS, DIFFERENCES, AND CHANGES BY SECTOR, 1992-2003

	2003 Earnings Percentiles			90-10 Log Wage Difference		90-50 Log Wage Difference		50-10 Log Wage Difference	
	90 th	50 th	10 th	2003	Change from 1992	2003	Change from 1992	2003	Change from 1992
Agriculture	\$44,149	\$19,234	\$9,126	1.58	-0.14	0.83	-0.02	0.75	-0.12
Mining	\$82,705	\$45,879	\$22,427	1.30	-0.09	0.59	-0.06	0.72	-0.03
Construction	\$73,174	\$34,181	\$14,831	1.60	-0.12	0.76	-0.01	0.83	-0.11
Manufacturing	\$90,650	\$34,176	\$15,183	1.79	0.13	0.98	0.16	0.81	-0.02
TCU	\$82,987	\$39,597	\$15,959	1.65	0.13	0.74	0.12	0.91	0.01
Wholesale	\$96,084	\$34,852	\$15,307	1.84	0.11	1.01	0.12	0.82	0.00
Retail	\$51,404	\$19,820	\$8,512	1.80	-0.05	0.95	0.01	0.85	-0.06
FIRE	\$114,428	\$37,083	\$16,244	1.95	0.18	1.13	0.15	0.83	0.03
Services	\$83,079	\$31,346	\$11,523	1.98	0.05	0.97	0.08	1.00	-0.03
All Sectors	\$82,207	\$31,477	\$11,992	1.93	0.06	0.96	0.10	0.97	-0.03

NOTE: Based on LEHD data from CA, IL, MD and NC.

TABLE 2
WORKER MOBILITY IN AND OUT OF INDUSTRIAL SECTORS, 1992-2003

	Number of Workers in 1992 and 2003	Proportion in Industry Sector		
		Only in 1992	Only in 2003	Both Years
Agriculture	578,036	39%	48%	13%
Mining	67,888	56%	29%	14%
Construction	1,511,595	32%	53%	14%
Manufacturing	5,145,894	44%	35%	21%
TCU	1,775,581	37%	44%	19%
Wholesale Trade	2,006,918	41%	47%	12%
Retail Trade	4,214,151	39%	49%	12%
FIRE	2,101,998	36%	47%	17%
Services	10,196,180	31%	51%	18%

NOTE: Based on LEHD data for CA, IL, MD and NC. Some rows may not sum to 100% due to rounding.

TABLE 3
CHANGES IN WORKFORCE COMPOSITION BETWEEN 1992 AND 2003

	Employment in 1992	Change in					Individual Fixed Effects (log points)
		Employment between 1992 and 2003	Proportion of Workforce				
			Male	14-29	30-49	50+	
Agriculture	300,709	17%	-7%	-6%	1%	5%	0.05
Mining	48,063	-39%	2%	0%	-11%	10%	0.08
Construction	704,268	46%	-1%	-5%	1%	4%	0.03
Manufacturing	3,357,441	-14%	2%	-7%	-1%	7%	0.06
TCU	991,212	14%	2%	-3%	-4%	7%	0.06
Wholesale Trade	1,066,376	11%	0%	-7%	0%	7%	0.06
Retail Trade	2,138,239	20%	0%	-5%	1%	4%	0.04
FIRE	1,111,889	21%	3%	-5%	-1%	6%	0.05
Services	4,998,570	41%	0%	-4%	-4%	7%	0.06

NOTE: Based on LEHD data for CA, IL, MD and NC.

TABLE 4
FIRM ENTRY AND EXIT RATES

	Number of Firms in 1992 and 2003	Proportion in Industry Sector		
		Only in 1992	Only in 2003	In Both Years
Agriculture	50,825	32%	39%	29%
Mining	2,135	45%	35%	20%
Construction	155,195	33%	45%	22%
Manufacturing	107,200	36%	37%	27%
TCU	56,355	35%	45%	20%
Wholesale Trade	143,414	36%	43%	21%
Retail Trade	263,093	40%	41%	20%
FIRE	120,763	33%	46%	22%
Services	686,606	31%	49%	19%

NOTE: Based on LEHD data for CA, IL, MD and NC. Some rows may not sum to 100% due to rounding.

TABLE 5
DECOMPOSITIONS OF CHANGES IN LOG EARNINGS AND EARNINGS INEQUALITY MEASURES, 1992-2003: ALL SECTORS

	(a) Levels							Change from 1992 to 2003 (8)
	2003	Sector Distribution	+Worker entry and exit	+ Change in observable characteristics	+ Firm entry and exit	+Sorting of firms and workers	1992	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
10 th percentile	9.379	9.393	9.403	9.245	8.261	9.082	9.276	0.103
50 th percentile	10.341	10.347	10.357	10.209	10.059	10.188	10.271	0.070
90 th percentile	11.322	11.322	11.337	11.254	11.299	11.242	11.155	0.167
90-10 difference	1.943	1.929	1.934	2.009	3.038	2.160	1.879	0.064
90-50 difference	0.981	0.975	0.980	1.045	1.240	1.054	0.884	0.097
50-10 difference	0.962	0.954	0.954	0.964	1.798	1.106	0.995	-0.033
	(b) Change in Statistic when also Accounting for							
	Sector Distribution	Worker entry and exit	Change in observable characteristics	Firm entry and exit	Sorting of firms and workers	Residual		
10 th percentile	-0.014	-0.010	0.158	0.984	-0.821	-0.194		
50 th percentile	-0.006	-0.010	0.148	0.150	-0.129	-0.083		
90 th percentile	0.000	-0.015	0.083	-0.045	0.057	0.087		
90-10 difference	0.014	-0.005	-0.075	-1.029	0.878	0.281		
90-50 difference	0.006	-0.005	-0.065	-0.195	0.186	0.170		
50-10 difference	0.008	0.000	-0.010	-0.834	0.692	0.111		

NOTE: Based on LEHD data for CA, IL, MD and NC.

TABLE 6
DECOMPOSITIONS OF CHANGES IN LOG EARNINGS AND EARNINGS INEQUALITY MEASURES, 1992-2003: SECTORS WITH
DECLINING INEQUALITY

	Change in Statistic when also Accounting for						1992	Change from 1992 to 2003
	2003	Worker entry and exit	Change in observable characteristics	Firm entry and exit	Sorting of firms and workers	Residual		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Agriculture								
10th percentile	9.12	-0.01	0.21	0.26	-0.20	0.01	8.84	0.28
50th percentile	9.86	-0.01	0.14	0.04	-0.02	0.01	9.70	0.16
90th percentile	10.70	-0.01	0.12	-0.11	0.10	0.04	10.56	0.14
90-10 difference	1.58	0.00	-0.08	-0.37	0.29	0.02	1.72	-0.14
90-50 difference	0.83	-0.01	-0.02	-0.15	0.12	0.03	0.85	-0.02
50-10 difference	0.75	0.00	-0.07	-0.23	0.18	0.00	0.87	-0.12
Mining								
10th percentile	10.02	0.01	0.22	1.34	-1.37	-0.17	9.99	0.03
50th percentile	10.73	0.01	0.11	0.29	-0.36	-0.06	10.74	-0.01
90th percentile	11.32	0.01	-0.03	0.13	-0.03	-0.15	11.39	-0.07
90-10 difference	1.31	-0.01	-0.25	-1.21	1.35	0.03	1.40	-0.09
90-50 difference	0.59	0.00	-0.14	-0.17	0.33	-0.09	0.65	-0.06
50-10 difference	0.72	0.00	-0.12	-1.04	1.01	0.12	0.75	-0.03
Construction								
10th percentile	9.61	-0.01	0.16	0.56	-0.36	-0.15	9.41	0.20
50th percentile	10.44	-0.02	0.12	0.16	-0.08	-0.09	10.35	0.09
90th percentile	11.20	-0.02	0.07	0.03	0.01	0.00	11.13	0.08
90-10 difference	1.60	-0.01	-0.09	-0.53	0.37	0.14	1.72	-0.12
90-50 difference	0.76	0.00	-0.06	-0.12	0.09	0.09	0.77	-0.01
50-10 difference	0.84	-0.01	-0.04	-0.41	0.28	0.06	0.94	-0.11

Retail Trade								
10th percentile	9.05	-0.01	0.11	0.48	-0.36	-0.10	8.93	0.12
50th percentile	9.89	-0.01	0.14	0.13	-0.06	-0.14	9.83	0.06
90th percentile	10.85	-0.01	0.12	-0.02	0.07	-0.09	10.78	0.07
90-10 difference	1.80	-0.01	0.01	-0.50	0.43	0.01	1.85	-0.05
90-50 difference	0.95	0.00	-0.02	-0.15	0.13	0.05	0.94	0.01
50-10 difference	0.85	0.00	0.03	-0.35	0.30	-0.04	0.90	-0.06

NOTE: Based on LEHD data for CA, IL, MD and NC.

TABLE 7
 DECOMPOSITIONS OF CHANGES IN LOG EARNINGS AND EARNINGS INEQUALITY MEASURES, 1992-2003: SECTORS WITH RISING
 INEQUALITY

	Change in Statistic when also Accounting for						1992	Change from 1992 to 2003
	2003	Worker entry and exit	Change in observable characteristics	Firm entry and exit	Sorting of firms and workers	Residual		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Manufacturing								
10th percentile	9.63	0.01	0.20	1.35	-1.29	-0.18	9.55	0.08
50th percentile	10.44	0.01	0.21	0.13	-0.25	-0.04	10.38	0.06
90th percentile	11.42	0.00	0.13	-0.08	-0.01	0.17	11.20	0.22
90-10 difference	1.79	0.00	-0.07	-1.43	1.28	0.35	1.65	0.14
90-50 difference	0.98	0.00	-0.08	-0.21	0.24	0.21	0.82	0.16
50-10 difference	0.81	0.00	0.01	-1.22	1.05	0.13	0.83	-0.02
TCU								
10th percentile	9.68	-0.01	0.17	0.70	-0.69	-0.20	9.70	-0.02
50th percentile	10.59	-0.01	0.09	0.13	-0.17	-0.06	10.61	-0.02
90th percentile	11.33	-0.01	0.07	0.01	-0.05	0.08	11.22	0.11
90-10 difference	1.65	0.01	-0.11	-0.69	0.65	0.28	1.52	0.13
90-50 difference	0.74	0.01	-0.03	-0.12	0.13	0.14	0.62	0.13
50-10 difference	0.91	0.00	-0.08	-0.57	0.52	0.14	0.90	0.01
Wholesale Trade								
10th percentile	9.64	0.00	0.17	0.46	-0.38	-0.17	9.55	0.09
50th percentile	10.46	0.00	0.15	0.02	-0.09	0.01	10.38	0.08
90th percentile	11.47	0.00	0.08	-0.38	0.29	0.21	11.28	0.20
90-10 difference	1.84	0.00	-0.10	-0.84	0.67	0.38	1.73	0.11
90-50 difference	1.01	0.00	-0.07	-0.40	0.38	0.20	0.90	0.12
50-10 difference	0.82	0.00	-0.03	-0.44	0.29	0.17	0.83	0.00

FIRE								
10th percentile	9.70	-0.01	0.18	0.79	-0.66	-0.13	9.53	0.17
50th percentile	10.52	-0.01	0.16	0.14	-0.08	-0.02	10.32	0.20
90th percentile	11.65	-0.01	0.09	-0.10	0.20	0.16	11.30	0.35
90-10 difference	1.95	0.00	-0.09	-0.89	0.86	0.30	1.77	0.18
90-50 difference	1.13	0.00	-0.07	-0.24	0.28	0.18	0.97	0.15
50-10 difference	0.83	0.00	-0.02	-0.65	0.58	0.11	0.80	0.03
Services								
10th percentile	9.35	0.02	0.13	1.27	-0.99	-0.32	9.24	0.11
50th percentile	10.35	0.02	0.11	0.19	-0.08	-0.16	10.27	0.08
90th percentile	11.33	0.01	0.04	0.00	0.11	0.00	11.16	0.17
90-10 difference	1.98	-0.01	-0.09	-1.26	1.10	0.32	1.92	0.05
90-50 difference	0.98	-0.01	-0.07	-0.18	0.18	0.16	0.89	0.08
50-10 difference	1.00	0.00	-0.03	-1.08	0.92	0.16	1.03	-0.03

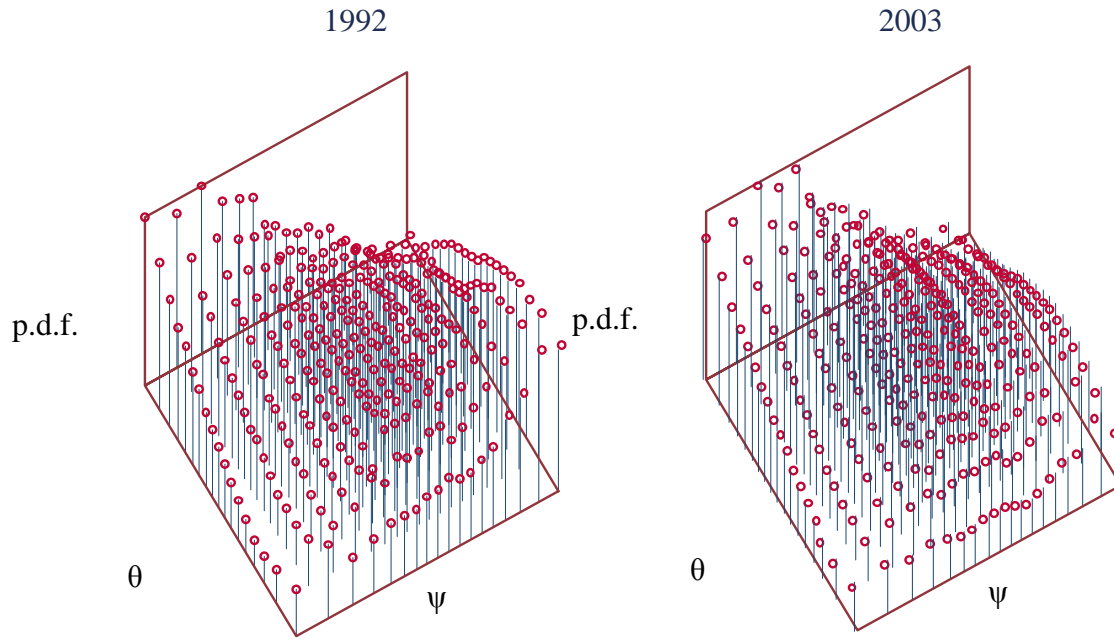
NOTE: Based on LEHD data for CA, IL, MD and NC.

TABLE 8
DECOMPOSITIONS OF CHANGES IN LOG EARNINGS AND EARNINGS INEQUALITY MEASURES, 1992-2003: EXCLUDING CALIFORNIA

	Change in Statistic when also Accounting for						1992 (7)	Change from 1992 to 2003 (8)
	2003 (1)	Worker entry and exit (2)	Change in observable characteristics (3)	Firm entry and exit (4)	Sorting of firms and workers (5)	Residual (6)		
Retail Trade								
10th percentile	9.72	-0.01	0.18	0.91	-0.78	-0.13	9.54	0.18
50th percentile	10.50	-0.01	0.18	0.13	-0.10	0.02	10.27	0.23
90th percentile	11.62	-0.01	0.11	-0.17	0.23	0.21	11.26	0.36
90-10 difference	1.91	0.00	-0.08	-1.08	1.01	0.33	1.72	0.18
90-50 difference	1.12	0.00	-0.08	-0.30	0.33	0.18	0.99	0.14
50-10 difference	0.78	0.00	0.00	-0.78	0.68	0.15	0.73	0.05

NOTE: Based on LEHD data for IL, MD and NC.

FIGURE 1
JOINT DISTRIBUTION OF WORKER HUMAN CAPITAL (θ) AND FIRM PAY POLICY (ψ) MATCH



NOTE: Based on LEHD data from CA, IL, MD and NC.

APPENDIX

TABLE A1

PARAMETER ESTIMATES FROM ABOWD ET AL. (2002) MODEL FOR LOG REAL ANNUALIZED EARNINGS USING POOLED LEHD DATA

	Variable Mean	Parameter Estimate
Constant		8.961
Male x		
Total Experience	10.428	0.132
Total Experience ² /100	2.757	-0.609
Total Experience ³ /1000	8.674	0.124
Total Experience ⁴ /10000	30.532	-0.010
Discontinuous Emp. in Year (0/1)	0.049	0.009
≥ 1 Full-Quarter of Emp. in Year (0/1)	0.424	-0.322
Earnings per Capita Index	0.485	0.618
Unemployment Rate (%)	2.894	0.015
Female x		
Total Experience	8.382	0.121
Total Experience ² /100	2.206	-0.616
Total Experience ³ /1000	6.909	0.144
Total Experience ⁴ /10000	24.163	-0.013
Discontinuous Emp. in Year (0/1)	0.033	0.091
≥ 1 Full-Quarter of Emp. in Year (0/1)	0.359	-0.350
Earnings per Capita Index	0.399	0.735
Unemployment Rate (%)	2.377	0.032
Observations		691,002,595
Workers		121,373,970
Firms (State Employer Identification Numbers)		7,968,638
R-Squared		0.799

NOTE: Includes annual data from LEHD for an unbalanced sample of 22 states. Experience is measured as potential experience (age-education-6) upon an individual's first appearance in the data, then as the sum of observed and potential experience in subsequent periods. Discontinuous employment occurs when a worker does not report positive earnings one or more quarters in a given year. A full quarter of employment occurs when a worker reports positive earnings at the same employer in quarters $t-1$, t , and $t+1$. The earnings per capita index is measured as $1+\ln(E_k/E_{2000})$, where E_k denotes national real per capita wage and salary earnings (obtained from the Bureau of Economic Analysis) in year k . Annual national unemployment rate data obtained from the Bureau of Labor Statistics. All parameter estimates are statistically significant at the 1% level.