Industry Wage Differentials: A Firm-Based Approach

David Card
UC Berkeley

Jesse Rothstein
UC Berkeley

Moises Yi
US Census Bureau

October 2022

ABSTRACT

We revisit the estimation of industry wage differentials using linked worker-employer data from the Longitudinal Employer-Household Dynamics program. Building on recent advances in the measurement of *employer* wage premiums from workers who move across employers, we define the industry wage effect as the employment-weighted average employer premium in the industry. We show that cross-sectional estimates of industry differentials dramatically overstate industry differentials, due to unmeasured worker heterogeneity. However, estimates based on industry movers significantly *understate* the true differentials. Job switchers tend to move between firms offering similar firm-specific pay premiums, so those who switch to an industry with higher average pay premiums typically come from higher-paying firms in their origin industry and move to lower-paying firms in their destination industry (and vice versa), attenuating the implied industry effects. Corrected estimates based on average employer premiums indicate substantial heterogeneity in narrowly-defined industry premiums, with a standard deviation of 0.122. Higher-pay industries have substantially higher-skilled workers, particularly in the dimension of skill that is unrelated to education. There is small but systematic variation in industry premiums across cities, with more variability in both pay premiums and worker sorting in cities with higher-wage firms and higher-skilled workers.

JEL Codes: J31, J62
Keywords: Industry wage differentials; AKM models; mover design

---

*We thank Wolfgang Dauth, Enrico Moretti, Pat Kline, and Thomas Lemieux for helpful comments and conversations. We are also grateful to Nicole Gandre and Richard Jin for outstanding research assistance. All results in this paper derived from the LEHD have been reviewed to ensure that no confidential information is disclosed (DRB clearance number CBDRB-FY22-422). Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau.

I. Introduction

Wages of similar workers vary across industries in predictable ways.[1] In a classic paper, Krueger and Summers (1988) (hereafter, KS) summarized the distribution of these industry wage effects and showed that they were relatively stable over time and across data sets. More controversially, they also argued that the wage effects from a cross-sectional regression were quite similar to the wage changes for between-industry job movers, once allowance was made for misclassification errors in industry. Since comparisons of job movers hold constant both observed and unobserved worker skill characteristics, their analysis suggested that observed cross-sectional wage differences between industries reflect causal pay premiums, rather than differences in unobserved worker abilities. They took the existence of such premiums as evidence against the hypothesis that the labor market is perfectly competitive, and in favor of efficiency wage models in which higher wages can increase profits by increasing work effort or reducing turnover.

Despite their careful analysis, KS's conclusions about the role of unobserved worker ability in measured industry wage differences have not been universally accepted. One reason is that their mover analysis was based on small samples, and included only 7 industries.[2] In addition, some later studies showed stronger evidence of sorting of workers with higher unobserved skills to higher-paying industries.[3] And finally, in the years after KS, economists became more interested in Roy-style models with industry-specific *match effects*, which confound the interpretation of the wage changes for industry movers (e.g., Gibbons et al., 2005).

[1] These patterns were so systematic that one of the questions in the "Knowledge of the World of Work" test in the National Longitudinal Survey asked whether unskilled laborers in steel mills have higher or lower average annual earnings than unskilled laborers in shoe factories (Kohen and Breinich, 1975).

[2] Their samples of workers matched across consecutive May Current Population Surveys had 18,122 observations; their sample from the 1984 Displaced Worker Survey had 2,318 observations.

[3] Gibbons and Katz (1992) extended KS's analysis of wage changes for displaced workers and replicated their main findings, but also showed results from another specification that suggested a bigger role for unobserved ability. Murphy and Topel (1990) regressed the wage changes for job movers in consecutive March Current Population Surveys on the change in their estimated industry effects from a cross-sectional model and found a coefficient of only 0.36 (though they did not try to correct for misclassification in the assignment of job changers to industry changes). Abowd, Kramarz and Margolis (1999) presented results from France which suggested that skill differences across industries were large, though their estimation method was later shown to have significant problems. Other studies, including Haisken-DeNew and Schmidt (1999), Goux and Maurin (1999) and Carruth et al. (2004) using longitudinal data from France and Britain, respectively, also found larger roles for unobserved ability.

In this paper, we re-examine the structure of industry wage premiums and the degree of skill sorting across industries, taking advantage of two post-KS innovations: the availability of administrative earnings data for U.S. workers from the Longitudinal Employer-Household Dynamics (LEHD) program; and a growing body of evidence derived from the two-way fixed effects specification proposed by Abowd, Kramarz and Margolis (1999) (hereafter, AKM), confirming the existence of *firm-specific* wage premiums in the labor market (see Card et al., 2018 for a survey of this work). As noted by AKM, in a model with establishment-specific pay premiums, the *industry wage effect* can be defined as a weighted average of the pay premiums for establishments in that industry. We adopt this "bottom-up" approach and measure the industry wage premiums for roughly 300 4-digit industries in the U.S. labor market using the estimated establishment premiums from AKM models fit to LEHD data for the largest Commuting Zones (CZ's) in the country.[4]

We address two main sets of questions. First, how do the economy-wide industry pay premiums derived from this bottom-up AKM approach differ from the estimated premiums from a simple cross-sectional regression, or from a two-way fixed effects model with person and *industry effects*, akin to KS's job movers estimates? Are KS's conclusions born out using the massive samples of the LEHD, or are average industry pay differences mainly due to unobserved worker ability? Second, how do industry pay premiums and the degree of skill sorting vary across *local labor markets*? Are premiums the same in different places (as was implied by KS), or do they vary systematically with characteristics of the local market. Is the degree of skill sorting linked to the dispersion in the industry pay premiums? Differences in industry pay premiums and the degree of sorting have strong implications for local inequality. Their patterns also shed light on alternative models of firm-specific pay setting.

Starting from a national perspective, our first key finding is that the estimated industry effects from a simple cross-sectional model significantly ***overstate*** the magnitude of the premiums from

[4] Haltiwanger, Hyatt, and Spletzer (2022a,b) also estimate industry wage differences as the average of AKM firm effects.

an AKM approach.[5] The source of this bias is pervasive skill sorting across industries that is highly correlated with the actual wage premiums paid in an industry (Haltiwanger, Hyatt, and Spletzer 2022a). Indeed, we find that elasticity of average worker skills in a 4-digit industry with respect to the industry's AKM-based pay premium is around 0.9. Only a fraction of these skill differences (around one-half) is captured by conventional human capital variables, leading to an upward bias in the dispersion in industry pay premiums from a cross-sectional approach. The high correlation between industry premiums and average worker skills provides further confirmation of the importance of assortative worker-firm matching in the labor market.

Our second key finding is that the estimated industry premiums from a model with person and industry effects (i.e., an "industry mover" design) are significantly ***attenuated*** relative to the effects from a bottom-up AKM approach. This source of this bias, which has been under-appreciated in the literature, is that industry movers are differentially selected from the firms in in their origin and destination industries. In a model with person and industry effects, the residual includes a term representing the gap between the pay premium at the actual workplace and the average premium in the associated industry – a term we call the "industry hierarchy effect."[6] Empirically, we find that when workers move across industries the industry hierarchy term tends to (partially) offset the change in industry premiums: workers who move up the industry ladder tend to come from higher-paying establishments in their origin industry, and move to lower-paying establishments in their destination industry. Likewise, those who move down the industry ladder tend to move from lower-paying workplaces in their origin industry to higher-paying workplaces in their destination industry. Consequently, the change in the firm hierarchy component for industry movers is negatively correlated with the change in their industry premiums, causing an attenuation bias in the estimated industry effects.

[5] This is consistent with the findings of Murphy and Topel (1990) and Haisken-DeNew and Schmidt (1999), though as we discuss below the movers-based estimates that these authors treated as revealing the true industry effects are in fact themselves biased. Our data set covers the 2010-2018 period, so we cannot test whether the same conclusions were true in the earlier periods covered by previous researchers.

[6] This term was explicitly noted by AKM in their equation 2.6.

The magnitude of the bias is large, even in a model with over 300 4-digit industries. We find that the standard deviation of industry premiums is about 50% larger using the bottom-up approach than is indicated by the KS-style estimates based on average wage changes of between-industry movers.

Turning to a local labor markets perspective, we consider the 50 largest CZs in the country (from Los Angeles to Norfolk, Virginia) plus approximately 10 aggregates of smaller and contiguous CZs.[7] We characterize the distribution of industry effects in a given market *c* by the coefficient $\theta_c^{\psi\psi}$ from a (weighted) regression of the industry premiums in that market on the corresponding national premiums – analogous to the "beta" coefficient in a CAPM model.[8] Higher values of $\theta_c^{\psi\psi}$ mean that industry premiums are widened in market *c* relative to the national structure, while lower values mean they are compressed. We find a fairly wide range in the estimates of $\theta_c^{\psi\psi}$, with a mean of 0.88 and a standard deviation of 0.08.[9] The distribution of industry premiums is wider in labor markets with more high-premium firms, but there is only a weak relationship to the relative supply of skilled workers, or to unionization rates, relative minimum wages, or city size.

We propose two alternative measures of the local degree of skill sorting. The first, $\theta_c^{\alpha\alpha}$, is the regression coefficient of the mean person effect for a given industry in CZ *c* on the mean person effect in that industry in the country as a whole. This measures skill dispersion across industries in the CZ relative to that in the nation. The second, $\theta_c^{\alpha\psi}$, is the regression coefficient of the mean person effect for a given industry in CZ *c* on the national industry pay premiums. This measures the local degree of assortative matching across industries, using national rather than local

[7] For example, the CZs around Boston, Hartford, Providence, and Manchester/Lowell in New England are all among the 50 largest in the country. One of our aggregate CZs includes all workers and firms in New England that are not in those four CZs. For disclosure reasons in this draft we cannot precisely delineate the residual regions.

[8] In a companion paper (Card, Rothstein, and Yi 2022), we studied CZ earnings premia, and concluded that the between-industry variation in these CZ premia was small relative to the average. However, this need not mean that between-CZ variation in industry premia is small relative to the average industry premia. Our approach here of studying how much more or less dispersed industry premia are in a CZ relative to the country as a whole is a way of isolating a systematic component of the CZ-by-industry premia.

[9] The R-squared of these regressions is also potentially important – however, these tend to be similar across markets, and clustered in the range of 0.6 to 0.8. This means that the rank order of the industry effects is fairly similar across markets.

industry premiums.[10] The interpretation of these sorting measures is similar to that of $\theta_c^{\psi\psi}$: Higher values mean that industries that have high skilled workers or high premiums nationally have even higher skilled workers in this CZ, relative to industries with lower premiums or less-skilled workers. Again, we find wide dispersion of skill sorting across CZs, with standard deviations of $\theta_c^{\alpha\alpha}$ and $\theta_c^{\alpha\psi}$ of 0.10 and 0.13, respectively. We also find that the two skill sorting measures are very strongly related to each other, but more weakly (though still positively) related to $\theta_c^{\psi\psi}$. Skill sorting is positively related to the share of jobs in both high-premium and high-skill industries, and to the average skill of workers in the CZ. It is also positively related to the local minimum wage and to the size of the local labor market (the latter consistent with Dauth et al. 2022).

Our findings contribute to the longstanding debate over whether estimated industry wage differentials from a cross-sectional model largely reflect causal pay premiums or are mainly driven by unobserved ability differences. In this debate, we come down part way between the position of KS, who argued that the industry effects in a simple cross-sectional wage model are largely causal, and that of Murphy and Topel (1990), who concluded that industry pay differences "… can be easily rationalized as a result of unobserved quality differences." In our data the "true" industry wage premiums calculated from underlying AKM models with person and establishment effects are only about 60% as large as those estimated from a cross-sectional model, but they are about half again larger than those estimated from a movers design. They are also quantitatively large, with a standard deviation of around 0.12 – remarkably similar to that estimated by KS.

The existence and magnitude of these systematic *industry-wide* pay premiums mean that the establishment-level premiums documented in many recent AKM-related papers cannot be fully explained by idiosyncratic firm-specific factors, such as monopsonistic markdowns (Card et al., 2018), or compensating differentials for firm-specific disamenities (Sorkin, 2018). They also lend support to a continuing focus on differences between industries in the analysis of policies related

[10] It can also be interpreted as the reduced form coefficient from an IV estimate of the CZ-specific assortative matching coefficient, using national industry premiums to instrument local premiums.

to trade (e.g., Harrison and Rodriguez-Clare, 2010), worker retraining (e.g., Katz et al., 2022), and regional disparities (e.g., Moretti, 2012).

Methodologically, we also contribute to the growing literature that uses mover designs to identify the relative contributions of two sides of a binary interaction – e.g., employer and employee contributions to wages (AKM), or patient and place contributions to health care spending (Finkelstein et al., 2016). Our analysis points to potential biases that can arise when units on one side of the interaction are aggregated. In the case of industry wage differences, we find that the observed wage changes for industry movers tend to understate the average wage differences between industries because of the non-random selectivity of the origin and destination firms of movers. We conjecture that similar biases could arise in other contexts that use a relatively course aggregation of units. For example, in an analysis of place effects on children's outcomes (e.g., Chetty and Hendren 2018), there may be an unmeasured component of local neighborhoods (Chetty et al. 2020) that varies systematically for movers between cities. If families tend to move between neighborhoods with relatively similar outcomes, this would attenuate estimated effects of the larger geographic units.

Finally, we contribute some new descriptive facts to the analysis of local labor markets, based on the extent to which the distribution of industry wage effects is compressed or widened across CZs, and the sorting of higher skilled workers to industries is attenuated or magnified. In larger CZs, with larger shares of highly skilled workers and a greater share of employment in high-premium industries, wage inequality is magnified both by a widening of the wage premiums across industries and by an increase in the assortative matching of higher skilled workers to high-premium industries. Interestingly, the dispersion in industry pay premiums and the degree of skill sorting are, if anything, higher in CZs with higher levels of unionization and with higher minimum wages. These institutional factors do not seem to moderate the *between-industry component* of wage inequality, though they still can potentially affect within-industry inequality.

II. Industry pay premiums

In this section we present our "bottom up" framework for estimating average industry pay premiums. We formalize the differences between these premiums and premiums derived in two alternative approaches: a two-way fixed effects model with person and industry effects; and a one-way fixed effects model with worker skill characteristics and industry effects. We then present a descriptive framework for relating the structure of industry wage premiums and the degree of skill sorting in different local labor markets to the corresponding constructs at the national level.

A. Basic AKM model

Building on AKM and an extensive body of subsequent work, we start with a two-way fixed effects model of earnings determination (for people with one employer per period) of the form:

$$y_{it} = \alpha_i + \delta_{f(i,t)} + X_{it}\beta_X + \varepsilon_{it}. \quad (1)$$

Here, y_{it} represents the logarithm of earnings of individual *i* in quarter t, α_i is a person effect that captures permanent differences in the earnings capacity of *i* that are equally rewarded in all jobs, X_{it} is a vector of time-varying personal and market-level variables (including age effects and calendar quarter effects), $\delta_{f(i,t)}$ is an establishment effect that captures the wage premium paid at *i's* workplace or firm in quarter t (indicated by the index function $f(i,t)$) and ε_{it} is a residual term. (We use workplace, firm and establishment interchangeably, but our LEHD data set identifies establishments.) This residual incorporates three conceptually distinct components: (i) any persistent <u>match effect</u> between the worker and her current workplace; (ii) any <u>person-specific</u> factors that cause earnings to vary from quarter to quarter on the same job, such as health shocks or family disruptions; and (iii) any <u>establishment-wide</u> time-varying factors that cause earnings to vary for employees at the workplace, such as product demand shocks that lead to overtime or short hours in the quarter.

In our empirical analysis, we start by estimating models like equation (1) by OLS, separately for each of a set of large CZs.[11] It is well known that this approach will only yield unbiased estimates of the worker and establishment effects under the so-called "exogenous mobility" assumption – that the error term ε_{it} is orthogonal to the full set of establishment identifiers representing the employment history of worker *i.* In the Appendix (not yet disclosed) we reproduce a series of specification tests proposed by Card, Heining and Kline (2013) and Card, Cardoso and Kline (2016) that address the plausibility of the exogenous mobility assumption. We conclude that although exogenous mobility can be rejected in our LEHD samples, the combination of equation (1) with exogenous mobility provides a relatively good approximation to the earnings outcomes of most workers. We therefore assume that we can obtain unbiased estimates of the establishment effects in equation (1).

B. Industry premiums

In the framework of equation (1), a reasonable definition of the wage effect (or wage premium) for industry *j* is just the weighted average of the establishment premiums for all establishments in that industry, where the weight is the relative number of person-quarter observations in that establishment (versus others in the same industry).[12] Specifically, letting $j(f)$ represent the industry of establishment *f*, we define the AKM-based industry wage effect for industry *j* as:

$$\psi_j \equiv \frac{\sum_{j(f)=j} N_f \delta_f}{\sum_{j(f)=j} N_f}, \qquad (2)$$

where N_f is the number of person-quarter observations for establishment *f*. A similar definition was proposed by AKM (1999).

We interpret ψ_j as defined in equation (2) as our preferred definition of the wage premium for industry *j*. It corresponds to the average premium in an industry relative to another: if a worker

[11] This approach means that we are identifying the establishment effects in (1) from within-market movers only. This allows us to abstract from differences in the wage structure between markets. To construct average establishment effects from the same industry in different markets, however, we have to impose a cross-market normalization assumption, explained in detail below.

[12] We refer to wage premiums and pay premiums interchangeably. Our data report quarterly earnings but not hours, so we cannot distinguish components coming from hours vs. hourly wages.

were to move from a randomly selected job in industry j to a randomly selected job in industry j', their wage would increase, on average, by $\psi_{j'} - \psi_j$.

Next, we ask how this preferred measure differs from the wage premiums estimated by the two main approaches in the existing literature.

C. Person and industry effects model

Consider first an approach based on a two-way fixed effects model with person and ***industry effects***:

$$y_{it} = \alpha_i + \psi_{j(f(i,t))} + X_{it}\beta_X + \underbrace{h_{f(i,t)} + \varepsilon_{it}}_{\hat{\varepsilon}_{it}}, \tag{3}$$

where $h_f \equiv \delta_f - \psi_{j(f)}$ is the difference between firm $f's$ wage premium and the average premium in its industry. We refer to h_f as the "firm hierarchy component" of the residual $\hat{\varepsilon}_{it}$ in a model with person and industry effects.

Note that a two-way fixed effects model with person and industry effects can be interpreted as a "movers" design for identifying industry wage premia, as in Krueger and Summers (1988). In such a model, identification of the industry effects is based on wage changes for people who move between industries. However, the average wage changes for industry movers do not necessarily identify the industry wage premia defined in (2), even if the AKM model's <u>firm</u> mobility assumptions are satisfied. The problem is that industry movers may be non-randomly selected with respect to the industry hierarchy components of their origin or destination firms.

For example, suppose that workers tend to move between firms with similar firm-specific wage premiums (i.e., stay in the same rung of a job ladder based on firm premiums).[13] In that case, workers who move from a lower-premium to a higher-premium industry will tend to come from higher-paying firms within their origin industry, and land at lower-paying firms within their

[13] The hypothesis that workers tend to move between firms with similar firm-specific pay premiums is supported by the observation that worker effects α_i are strongly correlated with firm premia $\delta_{f(i,t)}$ – see e.g., Kline et al. (2020). Such a high correlation will only emerge if the expected firm premium conditional on the worker's effect α_i is highly predictable.

destination industry.[14] Likewise, workers who move to a lower-premium industry will tend to come from lower-paying firms in their origin industry and move to higher paying firms in their destination industry. This will generate a negative correlation between the change in the hierarchy effect and the change in the average industry effect for industry movers, leading to an attenuation bias in the estimated industry premiums from an industry mover design.

We examine the importance of hierarchy effects by comparing estimates of industry premia from equations (2) and (3). We also take advantage of the fact that we can measure the hierarchy components h_f directly, and we examine the change in the average of *h* for workers moving across industries that differ in their $\psi_j{}'s$. This gives direct evidence that these hierarchy effects are important components of estimated industry effects in switchers-type designs.

D. Cross-sectional industry effects model

Next, consider an approach based on a cross-sectional regression model that controls for the observed skills (S_i) of worker *i.* This is the basic model used by KS and by much of the subsequent literature. Starting from equation (3), consider the projection of the person effect α_i on a worker's observed skills, their *X's* in a given period, and their observed industry in that period:

$$\alpha_i = S_i\pi_{\alpha S} + \pi_{j(i,t)} + X_{it}\beta_{\alpha X} + \ v_{it}\ . \quad (4)$$

In this equation, $\pi_{j(i,t)}$ represents the mean of the permanent skill component for workers observed in the industry of worker *i* in period *t*, adjusting for their observed skills (i.e., *S* and *X*). Loosely speaking, it represents "mean unobserved ability" in the industry. Given evidence in the recent AKM-based literature of strong positive assortative matching at the firm level between high-wage workers and high-premium firms (Card et al., 2018; Kline et al., 2020), we expect π_j to be bigger in high-premium industries.

[14] A little more formally, if we assume that the change in firm-specific wage premiums is small, so $|\Delta\delta_f| < k$, then $|\Delta\psi_{j(f)} + \Delta\delta_f| < k$, which implies that the changes in the industry effect and the hierarchy effect are negatively correlated.

Similarly, consider the projection of the establishment hierarchy component $h_{f(i,t)}$ on the same variables:

$$h_{f(i,t)} = S_i\pi_{hs} + \ \rho_{j(i,t)} + X_{it}\beta_{hX} + u_{f(i.t)}. \tag{5}$$

Note that even though the mean hierarchy effect in each industry across all worker-time observations is 0, the mean conditional on worker characteristics is not necessarily 0. Indeed, holding constant observed worker skills we expect the mean hierarchy effect to be negatively correlated with the average industry premium, for reasons discussed in the previous subsection.

Substituting (4) and (5) into (3), we obtain an equation relating earnings to observed skills (S_i and X_{it}), an industry effect $\kappa_{j(i,t)}$, and a residual term:

$$y_{it} = S_i\pi + \kappa_{j(i,t)} + X_{it}\beta + \ \tilde{\varepsilon}_{it} \tag{6}$$

where

$$\kappa_j = \psi_j + \pi_j + \rho_j. \tag{7}$$

There are two sources of bias in a cross-sectional estimate of the wage premium for a given industry relative to the AKM-based estimates: unmeasured worker skills; and unmeasured "firm quality" (represented by the hierarchy effect). We investigate both components below, although our analysis is somewhat constrained by the fact that education – the most important observable skill in most studies – is only available for a subset of observations in the LEHD.

E. Geographic variation in industry premiums and worker sorting

Krueger and Summers (1988) spawned a large literature investigating industry wage premiums in different countries (see Rycx and Tojerow, 2007, for a recent survey). While international comparisons of industry premiums provide some insights into the factors responsible for these premiums (see e.g., Tuelings and Hartog, 1998) there are at least two confounding issues with such comparisons. First, there can be important differences across countries in how earnings are measured, how industries are classified, and how worker skills are measured. Second, as noted above, pay premiums estimated from cross-sectional models include biases attributable to unobserved worker skills and unobserved firm quality that can vary across settings.

Our large national data base and bottom-up framework provide a novel opportunity to explore differences in the distributions of industry pay premiums across local markets, while abstracting from measurement problems and differences in the extent of skill sorting across industries. We take as "local markets" the larger CZs in the country, and define the average industry wage premium in CZ *c*, ψ_{jc}, using the CZ-specific analogue of equation (2). We also define the average skill of workers in industry *j* in market *c* by the mean of the estimated person effects among workers employed in that industry (weighting by the duration of employment), $\bar{\alpha}_{jc}$. We then ask how the distribution of industry premiums varies across markets, how skill sorting varies across markets, and how both phenomena are related to market-level characteristics.

Card, Rothstein, and Yi (2022) found that the ψ_{jc}s are highly correlated across locations, leaving relatively little variation in industry premiums across places. Moreover, 4-digit industry employment counts at the CZ level are often quite small, sometimes with just a few firms, making it difficult to measure ψ_{jc} with any precision. To extract some signal in CZ-specific premium structures, we devise single-dimensional measures that capture whether a CZ's industry premiums and/or skill sorting are larger or smaller than in the nation as a whole. First, we estimate a (weighted) regression of the industry premiums in CZ *c* on the corresponding *national* premiums:

$$\psi_{jc} = \mu_c^{\psi\psi} + \theta_c^{\psi\psi}\psi_j + \xi_{jc}^{\psi\psi}, \tag{8}$$

using as a weight for industry *j* the national share of employment in that industry. Our measure of the CZ industry premium structure is the slope coefficient $\theta_c^{\psi\psi}$. Analogously to the beta coefficient in a CAPM pricing model, $\theta_c^{\psi\psi} < 1$ means that the industry premiums in CZ *c* are compressed relative to the national structure, while $\theta_c^{\psi\psi} > 1$ means they are expanded.[15]

[15] It is straightforward to show that the normalization of ψ_{jc} and ψ_j to 0 for the restaurant industry is absorbed by the intercept in (8), and that $\theta_c^{\psi\psi}$ is unaffected by the choice of normalization.

We construct a second descriptive coefficient to capture analogous variation in relative skill sorting in different CZs. This is based on a (weighted) regression of $\overline{\alpha}_{jc}$ on $\overline{\alpha}_j$:

$$\overline{\alpha}_{jc} = \mu_c^{\alpha\alpha} + \theta_c^{\alpha\alpha}\overline{\alpha}_j + \xi_{jc}^{\alpha\alpha} \ . \tag{9}$$

A CZ has more (less) skill sorting across industries than in the national labor market as a whole if $\theta_c^{\alpha\alpha}$ is greater than (less than) 1.

We also constructed a third measure, capturing whether a CZ's workers are more completely sorted into industries with different wage premia. Here, we estimated a (weighted) regression of the mean person effect for workers in industry *j* and CZ *c* on the national premium in that industry:

$$\overline{\alpha}_{jc} = \mu_c^{\alpha\psi} + \theta_c^{\alpha\psi}\psi_j + \xi_{jc}^{\alpha\psi}. \tag{10}$$

The coefficient $\theta_c^{\alpha\psi}$ provides a simple summary of the extent of skill sorting in CZ *c*, using the national premiums for each industry to index the between-industry "job ladder."[16] This can be compared to the "national" version of this same model:

$$\overline{\alpha}_j = \mu^{\alpha\psi} + \theta^{\alpha\psi}\psi_j + \xi_j^{\alpha\psi}. \tag{11}$$

A city with $\theta_c^{\alpha\psi} > \theta^{\alpha\psi}$ has more skill sorting than in the nation as a whole, while a city with $\theta_c^{\alpha\psi} < \theta^{\alpha\psi}$ has less.

In fact, $\theta_c^{\alpha\psi}$ is closely related to our first measure of skill sorting, $\theta_c^{\alpha\alpha}$ from model (9). Substituting (11) into (9), it is readily seen that:

$$\theta_c^{\alpha\psi} = \theta^{\alpha\psi}\theta_c^{\alpha\alpha} + \tau_c^{\xi\alpha\alpha}, \tag{12}$$

[16] This coefficient is conceptually similar to ones estimated by Dauth et al. (2022).

where $\tau_c^{\xi\alpha\alpha}$ is the slope coefficient from a regression of the error term in equation (9) on ψ_j, the national industry wage premium. In our sample $\theta^{\alpha\psi} = 0.9$ (see below) so $\theta_c^{\alpha\psi}$ is (roughly speaking) a "noisy" version of $\theta_c^{\alpha\alpha}$.

The coefficient $\theta_c^{\alpha\psi}$ defined in equation (10) can also be interpreted in another way. Consider a CZ-specific regression of average worker skills in each industry on the corresponding <u>local</u> wage premium:

$$\overline{\alpha}_{jc} = \mu_c \ + \lambda_c \ \psi_{jc} + \xi_{jc}^0. \tag{13}$$

Although this equation can be estimated by OLS, this may be biased. It is now well known that the covariance between estimated worker and firm effects from AKM models is downward biased due to correlated sampling error (Kline et al. 2020). The same downward bias applies to measures aggregated to the industry-CZ level, particularly in small cells. An alternative is an instrumental variables (IV) approach, using ψ_j as an instrumental variable for ψ_{jc}. The IV estimator is:

$$\lambda_c^{iv} = \frac{\theta_c^{\alpha\psi}}{\theta_c^{\psi\psi}} \tag{14}$$

which implies that $\theta_c^{\alpha\psi} = \lambda_c^{iv}\theta_c^{\psi\psi}$. If the true coefficient λ_c is constant across CZ's, and the residual in (13) is uncorrelated with ψ_j (so the IV estimator is consistent) then one might expect $\theta_c^{\alpha\psi}$ to vary across CZ's in strict proportion to $\theta_c^{\psi\psi}$. Indeed, we find that the two are very strongly correlated.

Each of our three coefficients $\theta_c^{\psi\psi}$, $\theta_c^{\alpha\psi}$, and $\theta_c^{\alpha\alpha}$ capture a different aspect of the industry wage structure in a CZ (though as noted the latter two are closely related to each other). We conduct a descriptive analysis of the variation across markets in these three coefficients and their relationships to each other and to various characteristics of the local market, including the relative share of employment in higher-premium industries, the relative supply of higher-skilled workers, the size of the local market, and measures of the local rate of union coverage and the relative level of the minimum wage.

III. Data

Our analysis relies on the U.S. Census's Longitudinal Employer-Household Dynamics (LEHD) data. These are derived from quarterly earnings reports provided by employers to state unemployment insurance (UI) agencies, which are then assembled into a national data set. The core data set includes total wages paid by a given employer to each worker in a quarter, with identifiers for the employer firm and establishment (discussed below). This is supplemented with information on workers and employers collected from other sources (e.g., decennial census and ACS files, linked at the individual level; see Abowd et al., 2009). The LEHD covers about 95% of private sector employment, as well as state and local government employees, but excludes federal employees, members of the armed services, and self-employed workers. From 2010 forward it includes data from all 50 states. We focus on person-employer-quarter (PEQ) observations from 2010Q1 to 2018Q2 where the worker is between 22 and 62 years of age. In some analyses we further limit attention to individuals whose education has been measured in the ACS (2001-2017) and linked to LEHD.[17]

A limitation of the LEHD is that there is no information on job start or end dates, or on hours of work. To help screen out part-time jobs and/or partial-quarter job spells we exclude PEQs with earnings below $3,800 (roughly the earnings from a full-time job at the federal minimum wage), quarters where an individual had multiple jobs, and any *transitional* quarters (the first and last quarter of any person-employer spell). We also drop PEQs with unknown industry and/or establishment location. Finally, we drop all individuals with fewer than 8 remaining quarters of observed employment in our 8½ year sample window.

The UI data that form the LEHD are submitted by employers, and contain an identifier for the employing firm and the state, but not for the specific establishment. The Census Bureau

[17] Because some of the education measures predate the beginning of our sample, the education sample overrepresents older individuals.

supplements these data with worker residential addresses and the locations of all establishments, linked to the owning firm. These are used to impute establishments for workers employed in multi-establishment firms (Vilhuber 2018). We use the first of the multiple imputations available to assign PEQs to establishments and, via the establishment, to industries and commuting zones. Industry premiums estimated obtained by averaging (2) only over single-establishment firms (which are not subject to any type of imputation) are very similar to those that use all firms.

We use the 4-digit NAICS code for the establishment, with a total of 311 distinct values.[18] We also use the establishment location to assign the worker to a commuting zone. We focus on the 50 largest CZs, accounting for over half of the national population. We group all other CZs into approximately 10 composite groupings.

As noted above, our preferred estimates of industry effects derive from an AKM model with individual and "firm" (or strictly speaking, establishment) fixed effects. For computational tractability, we estimate this model separately for each CZ. In practice, this means that the coefficients on time-varying individual covariates (the β_Xs in equation 1) are allowed to vary by CZ, and the individual effect is in fact an individual-by-CZ effect. As a result, firm effects are identified solely from workers who move from one establishment to another within the same CZ, and are only estimated for establishments in the largest connected set in each CZ. This includes 98% of all PEQs in the original sample. We normalize the firm effects in each CZ to have mean zero (weighted by the PEQ count) across all firms in the "restaurants and other eating places" industry (NAICS code 7225). This is a large sector, with 2.7% of national employment; is non-tradeable, not geographically concentrated, and competitive; and does not require highly specialized skills. It is also typically one of the lowest-paying industries. This makes it a useful benchmark for other sectors.

[18] Industry codes are imputed to establishments in the LEHD using the procedures described in Vilhuber and McKinney (2014).

In some analyses we focus on patterns of wage changes around industry changes (for workers who remain in the same CZ).[19] For these event studies we limit attention to workers who switch industries only once in our sample, with a stable job in the same firm for at least 5 consecutive quarters before the switch and a stable job in another firm, in a new industry but the same CZ, for at least five consecutive quarters after the switch. Because many moves involve periods of non-employment, we allow up to 6 quarters of non-employment between leaving the origin firm and appearing in the destination firm.[20]

Table 1 presents some characteristics of our LEHD samples. The first column describes the connected sets used for estimation of the AKM model. Columns 2 and 3 divide this into a group that never switches industries (column 2) and another that switch industries at least once within our sample window (column 3). The two groups are fairly similar, though movers are about 4 years younger on average, have about 9% lower mean earnings, and are observed for fewer quarters. Finally, column 4 summarizes our event study sample for industry changers. This sample is similar to the broader industry switcher sample in column 3, but has mean earnings that are even higher than the industry stayers in column 2, likely reflecting the restriction to workers with relatively stable employment.

IV. Descriptive overview of industry pay differences

Before turning to causal estimates of industry pay differentials, we begin by presenting simple descriptive evidence on differences across industries from the American Community Survey. We pool data from 2010 through 2018 – the same period covered by our LEHD sample – and relate log hourly wages to the average education of workers in each industry. For this, we use the 262 industry codes that are available in the ACS – these generally map to the codes used in the LEHD, though the ACS combines some small industries.

[19] In Card, Rothstein and Yi (2022) we present an analysis of wage changes around moves between CZ's.

[20] Transitional quarters are considered non-employment when computing this gap. Thus, we allow workers to have no UI-related work for up to four quarters between the last quarter of their origin job spell and the first quarter of their destination spell. Note that a worker can appear multiple times in the event study sample if he or she qualifies in more than one CZ, but this is rare.

Figure 1A shows average education and average log hourly wages across industries. The first thing to note here is that there is wide dispersion in each: The standard deviation of mean industry log wages is 0.31, while the standard deviation of mean years of education across industries is 1.35. Second, the two are very strongly related to each other. Their (weighted) correlation across industries is 0.685; each additional year of mean education is associated with mean wages that are 0.157 log points higher. This is much larger than standard estimates of the causal return to education (typically around 0.10 per year), suggesting that there is something beyond individual sorting that drives differences in mean industry wages. The higher between-industry "return" could reflect industry wage premiums, unobserved skill differences, or both.

As an initial probe of this, we estimate a regression of log individual wages on industry fixed effects, controlling for just over 200 individual characteristics (including education, gender, race/ethnicity, and field of study), as well as indicators for the major CZs and residual regions of the country.[21] This specification is similar in spirit to the cross-sectional estimates in KS, though richer, taking advantage of variables such as immigrant source country and field of study that were not available to KS. Figure 1B replaces unadjusted mean industry wages with the adjusted estimated industry effects (which we normalize to mean zero). The standard deviation of the estimated industry effects is much lower, 0.180. The slope is also much reduced, to 0.057, though still highly significant. Evidently, industries that have more educated workers also pay higher wages conditional on worker observed characteristics.

As equation (7) indicates, however, the industry effects from our ACS sample contain two terms beyond the causal effect of the industry as defined above – one term reflecting unobserved differences among workers in different industries, and the other reflecting firm hierarchy effects. We next turn to the LEHD data to explore models that adjust for these factors.

[21] The controls include: year effects; a quartic in potential experience, interacted with female; indicators for 4 race/ethnicity groups, interacted with female; indicators for single years of education, interacted with female; foreign-born, and foreign born but arrived in the US before completing schooling; dummies for 4 major immigrant source regions, interacted with years since arrival in the US; and 15 field-of-highest-degree indicators (for people with a BA or higher), interacted with female. The R-squared of the model is 0.398.

V. National industry differentials

In this section we present results for our national analysis of industry pay differentials. We begin by estimating the AKM model (1). The model includes fixed effects for each worker and each employer (establishment) as well as controls for calendar quarter (fixed effects) and a cubic polynomial in the worker's age. We estimate the model separately for each CZ, allowing the time and age effects to vary freely across CZs. This means that the firm effects δ_f are identified only from workers who switch firms within CZs; the same person is allowed to have different α_is in each CZ in which he or she is observed. The AKM model requires a normalization; we impose that the mean of δ_f across all firms in the restaurant industry (NAICS code 7225) in each CZ is zero.

We aggregate the estimated firm effects δ_f to the industry level, pooling firms in all CZs and weighting by each firm's employment (equation (2)). This yields our preferred estimates of the industry wage differentials, ψ_j, normalized to zero for the restaurant industry. The estimated wage premiums, along with the average value of the estimated worker effects for employees in the industry and the share of person-quarter observations contributed by the industry, are reported in Appendix Table A-1. The median industry ranked by ψ_j (hospitals) has a pay differential of +0.24 relative to the restaurant industry; the 25th percentile industry (elementary and secondary schools) has a pay difference of 0.13; and the 75th percentile industry (management and technical consulting) has a pay difference of 0.33. Four relatively small industries have lower wage differentials than the restaurant industry (so their normalized $\psi_j's$ are negative): drinking places (bars and pubs) with $\psi_j = -0.09$; florists with $\psi_j = -0.06$; wine and liquor stores with $\psi_j = -0.01$; and personal care services with $\psi_j = -0.01$.[22] At the other end of the scale, the industry with the highest pay differential is coal mining (NAICS code 2121), with $\psi_j = 0.80$.

The weighted standard deviation of ψ_j is 0.12, which is quite close to the (sampling error-adjusted) standard deviation of 2-digit industry effects estimated by KS. Figure 2 shows a

[22] Together these four account for only about 0.3% of employment in the US, while restaurants account for about 3%. Note that to the extent that restaurant and drinking places workers fail to report their tip income, we may understate their true pay premiums.

histogram of the estimated ψ_j coefficients, using different colors for industries in each of 9 major 1-digit industry groups. The histogram is bell-shaped but skewed to the right, reflecting an upper tail of high-premium industries. As might be expected, manufacturing and FIRE/Administrative industries mainly contribute to the middle and upper parts of the distribution, while arts/entertainment/accommodation and education/health mainly contribute to the lower and middle parts of the distribution.

Nine of the ten highest ψ_j industries are in resource-related sectors (mining, petroleum refining, pipelines). Interestingly, the highest pay premium industry in the finance sector (securities and commodities exchanges, with a pay premium of 0.51) is only ranked #18, while investment and brokerage firms, the industry with the highest average worker skill as measured by the mean of the α_is for employees in the industry, is ranked #38, with a pay premium of 0.42.

Figure 3 shows how the estimated ψs compared to the mean education of workers in the industry.[23] The correlation is positive but not statistically significant, and the slope is less than one third of what we saw in Figure 1B (where we plotted the industry effects from a cross-sectional wage model against mean education in the industry). Evidently the industry effects based on a cross-sectional model are biased by unobserved worker skills that are positively correlated with education; removing this unobserved skill component eliminates most of the relationship between industry wage premiums and worker education. Note, moreover, that there is enormous vertical range in the figure – industries with average worker education around 13, for example, have industry pay premia that range from -0.1 to +0.7.

Figure 4 shows the relationship between industry premiums and average wages. The relationship is very tight, with an R^2 of 0.72 – higher premium industries are largely higher average wage industries. The slope of average earnings with respect to ψ_j is 1.93, indicating a strong pattern of assortative matching across industries between high-skilled workers and high-premium industries that magnifies wage inequality. Figure 5 shows this more directly. Here, we plot the average value

[23] To construct this figure, we crosswalk the NAICS industries available in the LEHD to the somewhat coarser groupings reported in the ACS, grouping each together to permit common definitions. The figure reports means for 222 industry groups.

of α for workers in each industry against the associated industry wage premium ψ_j, providing a visualization of equation (11) above. The estimated slope (i.e., the coefficient $\theta^{\alpha\psi}$) is 0.90.[24]

Table 2 presents regression versions of several of these analyses, with industries as the unit of observation. We begin, in columns 1 and 2, by regressing first $\bar{\alpha}_j$ and then ψ_j on mean unadjusted log wages in the industry. The slopes of these models correspond to the share of the variation in industry wages that is attributable to worker sorting and industry premiums, respectively. They indicate that 62% is due to differences in workers and 37% to differences in premiums. (These do not add exactly to 100% because a small portion is due to differences in the time-varying observables in (1).)

Column 3 shows a regression of $\bar{\alpha}_j$ on ψ_j, and is the model plotted in Figure 5. Each 0.10 log point increase in an industry's premium is associated with a 0.09 log point increase in the skill index of the workers in that industry, magnifying the effect of the premium.

Columns 4-8 shift samples to the 222 industry groupings in our ACS-LEHD crosswalk, as in Figure 3. We begin by replicating the model from column (3) in this sample; it is essentially identical. In Column 5, we regress the average α for workers in each industry from the LEHD on the average education of workers in the industry, measured in the ACS. The slope is 0.10, very similar to standard estimates of the return to education.[25] Education explains almost exactly half of the variation in industry average αs, suggesting that there is substantial skill sorting not accounted for by observed measures. Column 6 adds a control for the industry premium ψ_j. Interestingly, the industry premium is strongly and significantly related to $\bar{\alpha}_j$ even after controlling for education – there is substantial sorting on the unobserved component of worker skills.

In columns 7 and 8 we illustrate this further by decomposing $\bar{\alpha}_j$ into two components – the portion predicted based on industry mean education and the remainder (i.e., the predicted value

[24] The slope in Figure 4, 1.93, reflects the sum of a contribution of 1.0 from industry premiums, 0.90 from average worker effects, and a very small amount (0.03) from the time varying covariates X in (1), which are slightly positively related to the industry effects.
[25] For example, if we fit a conventional Mincer equation to our ACS data, including education, a cubic in experience, and controls for Black race, Hispanic ethnicity, and immigrant status, the coefficient on education is 0.107 for males and 0.110 for females.

and the residual from the model in column 4). We regress each of these separately on ψ_j. The residual component of worker skill is much more strongly related to industry premiums than is the average education of workers in the industry.

A. Assessing the validity of the decomposition

As has been widely noted, interpreting the firm effects in the AKM decomposition (1) as reflecting the causal effects of firms requires strong assumptions, most notably that person and firm effects are additively separable and that firm-to-firm mobility is exogenous with respect to the error term ε_{it}. We defer a comprehensive exploration of the validity of these assumptions to the Appendix (not yet disclosed), but we present some basic evidence here, focusing on the interpretation of the industry averages of the firm effects.

Panel A of Figure 6 shows an event study of changes in age-adjusted earnings in the quarters leading up to and following a move from one firm to a new firm in a different industry. This analysis is based on our "event study" sample of workers who were stably employed at a single firm for at least five quarters, moved to a new firm in a new industry, and were stably employed there as well for at least five quarters. We divide industries into quartiles based on their estimated effects ψ_j, and show eight series, corresponding to moves originating at firms in the top and bottom quartiles. The graph shows age-adjusted earnings by "event time," the number of quarters prior to a move (negative values) or following arrival in the new firm (positive values, with zero corresponding to the first quarter the worker is observed at the new firm).

Three facts are apparent here. First, workers' earnings seem to change in accordance with the nature of their move. Those who move from low-wage industries to high-wage industries (labeled "1-4" and "1-3" on the plot) see substantial increases in their earnings, while those who move from high-wage to low-wage industries (labeled "4-1" and "4-2") see substantial declines. By comparison workers who move to industries with wage differentials not that much different from their origin industry ("1-1" and "4-4" moves) see relatively small average changes in their earnings when they move.

Second, most of the change is observed in the quarter of the move. There is no sign of differential trends in earnings leading up to a move, which would indicate a violation of the AKM model's exogeneity assumption. There is a little more evidence of an adjustment process following a move: all of the lines slope up slightly between quarter 0 and quarter 2. This pattern is similar to one uncovered in a similar analysis of between-CZ moves in Card, Rothstein, and Yi (2022), where we find that movers to a new CZ take 2-3 quarters to fully adjust to their new earnings level.

Third, we see that a worker's initial wage level varies not just with the origin industry, as we would expect from the industry wage differentials that we estimate, but also with the destination industry. Among workers who originate in high-wage industries, those who will later move to a lower-wage industry earn substantially less in the origin industry than those who will remain in high-wage industries, even a full year before the move. Similarly, among workers originating in low-wage industries, those who will eventually move to high-wage industries earn more than those who will not. This arises from substantial worker sorting: A worker's destination industry is informative about his or her α_i, with workers who will later work in higher-wage industries tending to have higher α_is. Thus, the figure reflects a micro-level version of the strong positive correlation between α_i and $\psi_{j(f(i,t))}$ seen in Figure 5.

The large differences in α_is across groups in panel A necessitate a zoomed-out scale, which makes it difficult to see smaller differences across groups. In Panel B, we repeat the exercise but plot the mean of $\hat{\varepsilon}_{it}$, the error term in the AKM decomposition, by event time quarter and group. Ideally, there would be no systematic variation in this. This is not quite the case. We see that all eight groups' wage residuals decline in the first quarter following a move, by as much as 0.08 log points, then trend upward over the next few quarters. The tendency for negative residuals post-move is especially pronounced for movers from low-wage to high-wage industries (i.e., the groups transitioning from quartile 1 or 2 to quartile 4), suggesting that these workers do not immediately see the full increase associated with the new industry premium.

Finally, Panel C of the Figure shows the event study variation in the firm hierarchy effect, h_f. This reflects the premium offered by the firm where the worker works, relative to the average premium in the industry. This is stable in the quarters leading up to and following a move, as our event study sample is limited to workers who remain in the same firms for five quarters before and after the focal move. However, we see that h_f changes fairly substantially at the time of moves, in a way that is negatively correlated with the change in industry effects: Workers who move from low-wage industries to high-wage industries tend to leave from firms that paid better than average for their industries and to move to firms that paid worse than average in the new industries. The reverse is true for workers who move from high-wage to low-wage industries – hierarchy effects increase by a full 0.15 log points. As we discussed in Section II.C, the change in hierarchy effects represents an omitted variable in an industry movers analysis that does not account for firm effects, and the pattern seen here indicates that such an analysis will tend to understate differences in earnings effects across industries.

Figure 7 provides another look at this phenomenon. Here, we divide industries into 20 vingtiles based on their estimated pay premiums, and construct 400 cells corresponding to moves between firms in an origin vingtile v of the premium distribution and firms in a destination vingtile v'. We construct the mean change in the age-adjusted wage and in different components of it between the last quarter at the origin firm and the first quarter at the destination firm. We then plot these against the change in ψ_j from the origin to the destination industry.

In Panel A of Figure 7 we plot the mean age-adjusted wage change for each of the 400 mover groups against the change in the mean industry pay premium for each group. The 45-degree line shows the change that would be expected if a typical mover simply exchanged her old industry's premium for the premium of her new industry, with no change in the average hierarchy effect or the average AKM residual. We see that earnings changes are substantially attenuated relative to that benchmark – workers who move to higher- (lower-) premium industries gain (lose) much less than the difference in premiums would imply. Indeed, for relatively "small" moves (where $|\psi_{j'} - \psi_j|$ is small), the wage changes are generally close to zero.

By construction, the change in the mean age-adjusted wage for a mover from vingtile v to v' consists of three components: (1) the change in the mean of the industry pay premiums between vingtiles v and v'; (2) the change in the average within-industry hierarchy component (i.e., the change in the gap between the average firm premium received by a typical mover and the average firm premium paid to all workers in the industry); and (3) the change in the average AKM residual in the pay of movers. Panel B of Figure 7 shows the changes in the AKM residuals for the movers in each origin-destination pair. Consistent with the pattern in Panel B of Figure 6, there is tendency for movers with the largest upward change in their industry effects (i.e., with the largest positive values on the x-axis) to experience somewhat negative residuals in the first quarter after their move. However, the contribution of changes in the AKM residuals to the "flattening" observed in Panel A is small.

Panel C shows the change in the firm hierarchy effects. Here we see a much more substantial downward slope, as expected from the pattern in Panel C of Figure 6. Workers who move to higher-premium industries systematically lose firm hierarchy effects, enough so to offset about one-third of the gain in industry premiums, with a similar improvement in hierarchy effects for workers moving to lower-premium industries. Such a pattern leads to substantial attenuation of industry effects in industry-switcher analyses like KS.

Panel D shows the change in earnings net of the hierarchy effect (that is, in $y_{it} - X_{it}\hat{\beta} - h_{f(i,t)}$). This is quite close to the 45-degree line, indicating that the estimated industry effects do a good job of predicting earnings changes for between-industry movers, once the change in firm hierarchy is accounted for.[26]

As we discussed earlier, the firm hierarchy effect is easily understood in a model of job ladders, where workers seek to climb the ladder to firms with higher premiums, either by moving within industry to a firm with a higher *h* or across industries to an industry with a higher ψ. This would suggest that workers may tend to move up the *h* distribution as they gain experience. We explore this in Table 3. Here, we regress the firm hierarchy effect for worker *i* in quarter *t* on a quadratic in the number of quarters of experience that the worker has in industry *j(f(i,t))*. We estimate this separately for younger workers, aged 26 or less at the beginning of 2010, and for older workers. There are two reasons to expect that the experience effect will be larger for younger workers. First, young workers switch jobs more often, and may be more actively engaged in climbing the

[26] Note that the variable on the y axis is the sum of the change in the industry effect and the change in the AKM residual. Thus, the tendency for the points on far right side of the graph to fall below the 45 degree line is just a reflection of the pattern of residuals in Panel B.

job ladder. Second, because we are able to measure industry experience only during the period covered by our sample, our experience measures are right-censored for many of the older workers in our sample.

We indeed find positive, statistically significant returns to industry experience, larger for younger than for older workers. The quadratic curves are concave downward, indicating that the experience effect maxes out at around 18 – 20 quarters of experience in an industry. However, the magnitude of the effects is quite small. The total increase in the hierarchy effect associated with five years of accumulated industry experience amounts to just a 1% increase in earnings for younger workers, and about two-thirds of that for older workers. Thus, while there is evidence supporting a within-industry job ladder based on industry-specific experience, the magnitudes are small, and it appears that changes in hierarchy effects are larger and more systematic for between-industry than for within-industry moves.

B. Comparison to other strategies

Figure 8 shows the relationship between cross-sectional estimates of industry differentials considered in the previous section (on the vertical axis) and the firm-movers-based estimates that we obtain from the AKM model (on the horizontal axis). Here, we use a simpler cross-sectional model that can be estimated in the LEHD data, with controls for calendar time, age, gender, race, ethnicity, and nativity but not for education. Despite this, we estimate industry effects that are strongly correlated with those from our preferred AKM-based approach (ρ=0.78) – the two approaches rank industries similarly. However, the cross-sectional approach overstates the variation in industry premiums relative to the more fully controlled AKM-based approach – the slope of the former estimates against the latter is 1.63 (S.E. 0.13). As noted in Section 2, the primary bias in the cross-sectional model is due to unobserved ability sorting. This figure shows that sorting to be substantial.

Table 4 compares a range of different cross-sectional and industry movers estimates of ψ_j to those from our preferred "bottom-up" AKM specification. In each column, we present a bivariate regression of the ψ_j estimates from a single alternative model on the preferred estimates.

Regressions are weighted by the number of person-quarter observations in the industry. A coefficient greater than 1 indicates that the alternative specification overstates the magnitude of industry premiums obtained from our preferred specification, while a coefficient less than one indicates that the alternative specification understates the magnitude. We also show standard deviations of the industry premiums from each specification; these can be compared to the standard deviation of 0.12 of our preferred estimates.

In columns 2-4, we consider cross-sectional specifications with varying controls. Column 2 controls only for the time-varying controls included in our main AKM specification – age and calendar time. Column 3 adds controls for time-invariant Xs available in the main LEHD data – sex, race, ethnicity, and an indicator for foreign born (this corresponds to the specification shown in Figure 8). Column 4 adds CZ fixed effects. In each case, we find that the cross-sectional models dramatically overstate the magnitude of industry differentials, by more than 60%, and that as a consequence they overstate the variability of these differentials by a factor of 2 or more. All of these are consistent with dramatic bias in the cross-sectional estimates from the limited observability of the permanent determinants of workers' earnings.

Columns 5-7 consider specifications like equation (3), with person and industry fixed effects but not firm controls. These are industry movers specifications of the style estimated by KS and many subsequent authors, and as noted in section 2 are subject to bias from omitted firm hierarchy effects. In column 5, we include in X just age and calendar time; in column 6 we add CZ fixed effects; and in column 7 we estimate a full set of industry-by-CZ effects, then average these to the industry level to estimate ψ_j. These models consistently <u>under-estimate</u> the magnitude of the industry premiums, with coefficients around 0.62 and standard deviations about two-thirds as large as those from our preferred specification. Consistent with Figures 6 and 7, these results show that hierarchy effects are important, and that failure to adjust for them substantially biases estimated industry premiums toward the mean.

C. Differences by education

Our estimates so far focus on a single wage premium per industry that applies to all workers.

However, different kinds of workers may select into firms at different points in a within-industry distribution, implying that the average effect of the industry may differ across groups of workers. To explore this, we estimate separate industry premiums for college and non-college workers. We continue to rely on a pooled AKM specification (1), in which firms have constant effects on their college and non-college workers' log earnings, but modify (2) to use weights N_j that are specific to each education group. This yields separate industry premiums reflecting differences in the firm-specific premiums offered by firms employing college workers and the firm employing non-college workers.

Of course if the firm distributions of the two education groups are the same, the two premiums will be identical. If, on the other hand, more- and less-educated workers work at completely different firms within the same industry, the two premiums could be quite different. Empirically, the degree of segregation of more- and less-educated workers between firms within the same industry is similar to the (relatively high) degrees of segregation of whites versus nonwhites, and of female versus male workers. Thus, there is substantial leeway for the wage premiums of the two education groups to differ within industries.

Figure 9 explores the relationship between these education-specific premiums and the pooled premium described by equation (2). They are extremely highly correlated – 0.98 for non-college workers and over 0.99 for college workers. Evidently, industry wage premiums are quite similar for the two groups, though the slopes in Figure 9 indicate slightly larger premiums for college than for non-college workers.

The degree of similarity is reduced when we look at the sorting of workers *within* each of the two education groups to higher- and lower-premium industries. Figure 10 shows how the mean person effect for workers in each industry relates to the pooled industry differential, separately for college and non-college workers. Not surprisingly, the mean of the person effects is higher for college than for non-college workers, both overall and within industries. But the scatter of points in the figures shows that there is clearly more sorting of high-α college workers into high-premium industries than there is for high-α non-college workers. This echoes a similar finding regarding the sorting of college and non-college workers between CZs with higher and lower average pay premiums in Card, Rothstein, and Yi (2022). There, we find that geographic sorting within the college-educated workforce is quite systematic, whereas non-college workers are much more evenly distributed across CZs.

VI. Local industry differentials

Although most of the studies of industry pay differentials that followed KS focused on measurement issues and the potential importance of unobserved worker skills, a number of studies tried to use differences in the patterns of industry differentials across countries or over time to say something about the explanation for these differences. Our setting provides a novel opportunity to pursue this agenda, focusing on differences in pay differentials and worker sorting across major CZs in the US. Relative to simple cross-country comparisons we have two key advantages. First, we can sidestep issues of data comparability. Second, our AKM-based approach allows us to clearly delineate between the average pay premiums offered by firms in different industries and the relative skills of the workers in those industries.

As noted above, we summarize the patterns of industry wage premiums and the degree of skill sorting across industries in different CZ's using three simple regression coefficients:

i) $\theta_c^{\psi\psi} \equiv \frac{d\psi_{jc}}{d\psi_j}$, the slope of the relation between ψ_{jc} and ψ_j (the national industry premium);

ii) $\theta_c^{\alpha\alpha} = \frac{d\bar{\alpha}_{jc}}{d\bar{\alpha}_j}$, the slope of the relation between $\bar{\alpha}_{jc}$ and $\bar{\alpha}_j$ (the national mean of skill); and

iii) $\theta_c^{\alpha\psi} = \frac{d\bar{\alpha}_{jc}}{d\psi_j}$, the slope of the relation between $\bar{\alpha}_{jc}$ and ψ_j.

Rows 1 and 2 of Table 5 summarize these slopes. The (weighted) national means of the three coefficients across our set of major CZs and residual geographic areas are 0.88, 0.86, and 0.80, respectively, with standard deviations of 0.08, 0.10, and 0.13, respectively.[27]

[27] Recall that $\theta_c^{\psi\psi} < 1$ indicates that CZ *c* has industry premiums that are compressed relative to the national premiums, with a similar interpretation for $\theta_c^{\alpha\alpha}$. That the means of these are well below 1 is a minor version of a

The CZ-level regressions that generate these coefficients (not reported in Table 5) have average R^2 coefficients of 0.75 for $\theta_c^{\psi\psi}$, 0.83 for $\theta_c^{\alpha\alpha}$, and 0.31 for $\theta_c^{\alpha\psi}$. The relatively high average R^2 for the models generating $\theta_c^{\psi\psi}$ means that the industry wage premiums in a typical CZ are very highly correlated with their national analogues (typical correlation coefficient around 0.87). Likewise, the 0.83 average R^2 for the models generating $\theta_c^{\alpha\alpha}$ means that CZ-specific measures of average skill in an industry are quite highly correlated with the corresponding national skill measures (typical correlation coefficient around 0.91). The lower average R^2 coefficients for the models generating $\theta_c^{\alpha\psi}$ have to be interpreted differently, since even at the national level, the "benchmark" version of this model (equation 11, above) has an R^2 of only 0.356. Thus, a typical CZ-specific model relating local skill means to the national industry wage premiums has an R^2 that is about 87% as large as that for the corresponding national model.

The lower rows of Table 5 show simple descriptive regressions relating the three coefficients to each other. The samples for these regressions are the largest CZs, weighted by the number of person-quarter observations in our sample in the CZ, and the regressions control for fixed effects for the four census regions.[28] For each regression we show the coefficient and robust standard error, and the within-region R-squared – the explained portion of the variance of θ_c remaining after partialling out the region controls.

Looking at the regression coefficients and R-squared measures, we see that the strongest relationship among the three coefficients is between $\theta_c^{\alpha\alpha}$ and $\theta_c^{\alpha\psi}$. As noted in the derivation of equation (12), above, this is what we would expect, since the two coefficients are essentially different measures of the same skill sorting phenomenon. They are correlated 0.86 with each other, and we have found that they have similar relationships with other CZ characteristics.

Simpson's paradox – part of the national variation in ψ_j and α_j derives from differences across CZs that specialize in different industries, so it is possible for every CZ to have more a more compressed within-CZ premium structure than the country as a whole.

[28] Due to Census disclosure rules, the samples actually include the residual geographic areas along with the largest CZs. However, we add a set of indicators to "dummy out" the residual regions. Regressions of our three θ_c variables on dummies for the residual regions and 4 region indicators have R-squared coefficients of 0.26 ($\theta_c^{\psi\psi}$), 0.57 ($\theta_c^{\alpha\alpha}$), and 0.60 ($\theta_c^{\alpha\psi}$). The. R-squared statistics pertain only to the. larger CZs, excluding the composite areas.

Accordingly, our subsequent analyses focus on $\theta_c^{\alpha\alpha}$; results for $\theta_c^{\alpha\psi}$ are similar when suitably scaled.

Somewhat unexpectedly, there is only a relatively weak (marginally statistically significant) relationship between $\theta_c^{\alpha\alpha}$ and $\theta_c^{\psi\psi}$. The relative degree of skill sorting across industries in a CZ does not appear to be strongly determined by the relative compression or stretching of the wage premiums for different industries in that CZ. However, the relationship between $\theta_c^{\psi\psi}$ and $\theta_c^{\alpha\psi}$ is somewhat stronger – as would be expected from equation (14).

In Panel A of Figure 11 we plot industry wage premiums against national wage premiums, showing just the 10 CZs with the highest values of $\theta_c^{\psi\psi}$ and the 10 CZs with the lowest values of $\theta_c^{\psi\psi}$. There is a clear separation between the scatters for these two sets of CZs, illustrating the systematic compression of premiums in the low-$\theta_c^{\psi\psi}$ CZs relative to the high-$\theta_c^{\psi\psi}$ CZs. The slope of the scatterplot is 0.82 for the low-$\theta_c^{\psi\psi}$ CZs and 1.07 for the high-$\theta_c^{\psi\psi}$ CZs. In Panel B of the figure we plot the industry average values of $\bar{\alpha}_{jc}$ for the same two groups of CZs against $\bar{\alpha}_j$ (the national mean of the person effects in the corresponding industry). The scatter of points for the high-$\theta_c^{\psi\psi}$ CZs has a clearly steeper slope than the scatter for the low-$\theta_c^{\psi\psi}$ CZs, suggesting that $\theta_c^{\alpha\alpha}$ is higher, on average in the former set of CZs.

Table 6 relates the θ_cs to other CZ characteristics. The first row shows relationships to CZ size, which is a key characteristic in many theoretical and empirical models of local labor markets (Moretti, 2010). There is no relationship between CZ size and the dispersion of industry premia. However, there is significantly more skill sorting in larger CZs. This is consistent with the finding of Dauth et al. (2022) that high-quality workers are more likely to be matched to high-wage firms in bigger cities. We also find a positive, significant relationship between city size and $\theta_c^{\alpha\psi}$, which is more directly related to the Dauth et al. (2022) analysis but is not reported in Table 6.

The next rows relate the θ_cs to the average pay premiums and average worker effects in the CZ. We see in column 3 that CZs with higher average pay premiums have more spread in their industry premiums. This is partly mechanical -- our normalizing assumption that $\psi_{jc} = 0$ for the restaurant industry means that our measure of average local pay premiums is strongly correlated with the spread in premiums.[29] As shown in column 4, CZs with higher average pay premiums also have more sorting of high skilled workers to high-premium industries. Since widening of industry wage premiums and greater skill sorting both magnify within-CZ wage inequality, the implication is that places with higher average pay premiums tend to have higher inequality in wages.

A CZ could have high average pay premiums either because it has an unusual concentration of employment in industries with high average pay premiums, or because firms in the CZ tend to pay higher premiums than firms in the same industry in other CZs. Likewise a CZ could have high average worker skill because it has relatively high share of employment in industries that tend to attract high-α workers, or because in that particular CZ, $\bar{\alpha}_{jc} > \bar{\alpha}_j$ for most sectors. To explore the effects of industry composition, we use the local employment shares of each industry and national average pay premiums and worker skills in each industry to construct "expected" local average pay premiums and local average worker skills:

$$\bar{\psi}_c^{exp} = \sum_j w_{ic}\psi_j\,; \quad \bar{\alpha}_c^{exp} = \sum_j w_{ic}\bar{\alpha}_j$$

An advantage of these measures is that our normalization choice amounts to an additive constant that does not vary across CZs, so is absorbed by the constants in our θ regressions. These are presented in the next rows of Table 6. We see a similar pattern of positive relationships, with substantially larger (but more imprecisely estimated) coefficients for the expected average pay premiums and expected average skills than for the actual averages. This suggests that local industry structure plays a potentially important role in the effects seen in the upper rows of the table.

[29] Recall that the restaurant industry is very near the bottom of the ψ_j distribution; the same is true in each CZ. Thus, the average normalized premium in a CZ reflects the distance between the average firm in the CZ and the restaurant industry, and is increased whenever the spread increases – even if this derives from lower pay in restaurants rather than higher pay elsewhere.

We explore this further in Appendix Table A-2, where we decompose mean pay premiums and mean worker skills into their expected means, based on local industry shares and national characteristics of the industry, and the local deviations from these expected means:

$$\bar{\psi}_c = \bar{\psi}_c^{exp} + (\bar{\psi}_c - \bar{\psi}_c^{exp}); \quad \bar{\alpha}_c = \bar{\alpha}_c^{exp} + (\bar{\alpha}_c - \bar{\alpha}_c^{exp})$$

Here we find that in general both the expected component and local deviation of average pay premiums and average worker skills affect $\theta_c^{\psi\psi}$ and $\theta_c^{\alpha\alpha}$, but that the expected component has a larger effect. The one exception to this rule is the connection between average worker skills and the local dispersion in pay premiums $\theta_c^{\psi\psi}$. In this case, only the part of $\bar{\alpha}_c$ based on industry structure matters, while the effect of the local deviation, $(\bar{\alpha}_c - \bar{\alpha}_c^{exp})$ has a weak <u>negative</u> effect. This points to the possibility that, holding local industry structure constant, an increase in the supply of highly skilled workers may (slightly) compress local industry differentials.

The final two rows of Table 6 explore two institutional features that might influence wage premia or worker sorting in a CZ, unionization and the minimum wage. We first consider the share of workers in a CZ who are covered by a union, computed from data assembled by Hirsch and Macpherson (2003). We might expect unions to both raise wages and influence skill sorting in a CZ, though the exact nature of the effect on the θ_c slopes likely depends on the specifics of which industries are unionized. In fact, we find little relationship between the CZ unionization rate and either of the θ_cs.

Ex ante predictions are somewhat clearer for the minimum wage: We expect a higher minimum wage to raise wages in the lowest-wage, lowest-premium industries, reducing $\theta_c^{\psi\psi}$. It might also attract higher-skill workers to those industries, reducing $\theta_c^{\alpha\alpha}$. We construct mean minimum wages over the 2010 to 2018 period covered by our data, using data from Vaghul and Zipperer (2016, 2021). Where there are local minimum wages in the principal city of a CZ we use those; otherwise we use the state minimum wage. The resulting minimum wage measures are positively related to both θ_cs, significantly so for $\theta_c^{\alpha\alpha}$. We do not have a good account for this result. One possibility is that the minimum wage effect is confounded by other CZ differences; our 50-CZ cross-sectional analysis does not permit the most credible identification strategies from the minimum wage literature (Cengiz et al. 2019; Dube et al. 2010).

More generally, the results in Table 6 are merely correlations. We see them as indicating (1) that there are systematic differences in industry premia and between-industry worker sorting across CZs; (2) that with suitable parameterization these can be measured with enough precision to study; and (3) that they are clearly related to the local labor market structure. More work is needed to understand how the structure of the local labor market generates the variation in industry premia and skill sorting examined here.

VII. Conclusion

In this paper, we have used comprehensive universe data on U.S. workers to revisit the measurement of industry pay premiums. In light of the modern literature on firm differentials spawned by Abowd, Kramarz, and Margolis (1999), we define the industry pay premium as the weighted average of the premiums for each of the firms in an industry. We show that neither cross-sectional estimates nor two-way fixed effects estimates with controls for workers and industries (i.e., industry movers models) identify these effects. Cross-sectional estimates are substantially biased by the inability to control comprehensively for worker ability, while industry mover estimates are subject to what we call "firm hierarchy bias" – workers who switch industries tend to move among firms that are more similar to each other than are the average firms in their origin and destination industries. As a consequence, cross-sectional estimates overstate the dispersion of industry pay premiums, while movers designs understate it.

Our corrected estimates of industry pay premiums are constructed from the bottom up, by first estimating firm wage effects via the AKM model and then averaging to the industry level. These indicate substantial dispersion across industries, with a standard deviation across 311 4-digit NAICS industries of 0.12. Resource-related industries tend to have very high pay premiums and hospitality, education, and health have lower premiums, with traditional high-wage industries like finance and manufacturing in the upper middle. Industry premiums are strongly related to the average skill of workers in the industry, as measured by the AKM worker effects, but not to their average education – there is more sorting into industries on the basis of unmeasured determinants of earnings than on education. Premiums are quite similar for more- and less-educated workers in the same industry, though college-educated workers are more thoroughly sorted across industries on the basis of their unobserved skill than are non-college workers.

We also explore variation in industry premiums across the 50 largest commuting zones in the U.S. In a companion paper (Card, Rothstein, and Yi, 2022), we found that industry-by-CZ premiums were well described by an additive model with CZ and industry effects, with relatively little variation in the relative return to different industries across CZs. While the variation is limited, it is not zero. We isolate a measurable component of this variation that captures the degree to which industry premiums in a CZ are stretched or compressed relative to their national pattern. We show that this varies systematically with CZ characteristics: CZs with more high-wage firms offer greater between-industry pay dispersion, while worker sorting is related to CZ size, average firm premiums, and average worker skill. We find no relationship between firm premia or worker sorting and the CZ unionization rate, while we find a counterintuitive positive relationship of worker sorting with the minimum wage.

There are several open questions for future work. One is to dig deeper into institutional features that might influence local industry premiums or worker sorting. With only 50 CZs there is limited variation to identify this, but this setting is more promising than the cross-country analyses with much lower sample sizes that have been the focus of many similar investigations to date. Another is to explore the determinants of what we have called the firm hierarchy effect – why do some firms pay much more than their industry competitors while others pay much less? Third, it would be useful to understand whether industry premia are the same for all workers – we found evidence that they were basically identical for high- and low-education workers, but other heterogeneity (by gender, race, age, or occupation) would also be of interest.

Finally, we expect that the firm hierarchy bias that we found to be a major problem with estimates of industry premiums based on models with worker and industry fixed effects (i.e., movers designs) is a more general phenomenon. It is not uncommon to use fixed effects for aggregates, when there may be unmodeled heterogeneity within those aggregates. For example, analyses of geographic mobility often include county, commuting zone, or health service region fixed effects, when there is actually variation across much smaller neighborhoods or other within-region units (e.g., hospitals). Insofar as movers across regions tend to move between neighborhoods that are more similar to each other than are the regions as a whole, hierarchy bias will tend to attenuate between region differentials. Further research is needed to understand the scope and magnitude of this bias.

References

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* (Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts, eds.): 149-230. University of Chicago Press.

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High wage workers and high wage firms." *Econometrica* 67, no. 2: 251-333.

Card, David. 1999. "The causal effect of education on earnings." *Handbook of Labor Economics* vol. 3 (Orley Ashenfelter and David Card, eds.): 1801-1863. Elsevier/North Holland.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and labor market inequality: Evidence and some theory." *Journal of Labor Economics* 36, no. S1: S13-S70.

Card, David, Ana Rute Cardoso, and Patrick Kline. 2016. "Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women." *The Quarterly Journal of Economics* 131, no. 2: 633-686.

Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace heterogeneity and the rise of West German wage inequality." *The Quarterly Journal of Economics* 128, no. 3: 967-1015.

Card, David, Jesse Rothstein, and Moises Yi. March 2022. "Location, location, location." Working paper.

Carruth, Alan, William Collier, and Andy Dickerson. 2004. "Inter-industry wage differences and individual heterogeneity." *Oxford Bulletin of Economics and Statistics* 66, no. 5: 811-846.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The effect of minimum wages on low-wage jobs." *The Quarterly Journal of Economics* 134, no. 3: 1405-1454.

Chetty, Raj, and Nathaniel Hendren. 2018. "The impacts of neighborhoods on intergenerational mobility II: County-level estimates." *The Quarterly Journal of Economics* 133, no. 3: 1163-1228.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2018. “The opportunity atlas: Mapping the childhood roots of social mobility.” *National Bureau of Economic Research Working Paper,* no. w25147.

Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti and Jens Suedekum. 2022. “Matching in cities,” *Journal of the European Economic Association* 20, no. 4: 1478–1521.

Dube, Arindrajit, T. William Lester, and Michael Reich. 2010. "Minimum wage effects across state borders: Estimates using contiguous counties." *The Review of Economics and Statistics* 92, no. 4: 945-964.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of geographic variation in health care: Evidence from patient migration." *The Quarterly Journal of Economics* 131, no. 4: 1681-1726.

Gibbons, Robert, and Lawrence Katz. 1992. "Does unmeasured ability explain inter-industry wage differentials?." *The Review of Economic Studies* 59, no. 3: 515-535.

Gibbons, Robert, Lawrence F. Katz, Thomas Lemieux and Daniel Parent. 2005. "Comparative advantage, learning, and sectoral wage determination." *Journal of Labor Economics* 23, no.4: 681-723.

Goux, Dominique, and Eric Maurin. 1999. "Persistence of interindustry wage differentials: a reexamination using matched worker-firm panel data." *Journal of Labor Economics* 17, no. 3: 492-533.

Haisken-DeNew, John P. and Christoph M. Schmidt. 1999. “Industry wage differentials revisited: A longitudinal comparison of Germany and USA (1984-1996).” *IZA Discussion Paper,* no. 98.

Haltiwanger, John C., Henry R. Hyatt, and James Spletzer. 2022a. “Industries, mega firms, and increasing inequality.” *National Bureau of Economic Research Working Paper*, no. w29920.

Haltiwanger, John, Henry R. Hyatt, and James R. Spletzer. 2022b. "Sources of Increasing Earnings Inequality: Reconciling Survey and Administrative Data." Working paper.

Harrison, Ann, and Andrés Rodríguez-Clare. 2010. "Trade, foreign investment, and industrial policy for developing countries." *Handbook of Development Economics* 5: 4039-4214.

Hirsch, Barry T., and David A. Macpherson. 2003. "Union membership and coverage database from the current population survey: Note." *ILR Review* 56, no. 2: 349-354.

Katz, Lawrence F., Jonathan Roth, Richard Hendra, and Kelsey Schaberg. 2022. "Why do sectoral employment programs work? Lessons from WorkAdvance." *Journal of Labor Economics* 40. S1: S249-S291.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten. 2020. "Leave-out estimation of variance components." *Econometrica* 88, no. 5: 1859-1898.

Kohen, Andrew I., and Susan C. Breinich. 1975. "Knowledge of the world of work: a test of occupational information for young men." *Journal of Vocational Behavior* 6, no. 1: 133-144.

Krueger, Alan B., and Lawrence H. Summers. 1988. "Efficiency wages and the inter-industry wage structure." *Econometrica: Journal of the Econometric Society*: 259-293.

Moretti, Enrico. 2012. *The New Geography of Jobs*. Houghton Mifflin Harcourt.

Murphy, Kevin M., and Robert H. Topel. 1990. "Efficiency wages reconsidered: Theory and evidence." *Advances in the Theory and Measurement of Unemployment* (Yoram Weiss and Gideon Fishelson, eds.): 204-240. Palgrave Macmillan, London.

Rycx, François and Ilan Tojerow. 2007. "Inter-industry wage differentials: What do we know?." *Reflets et perspectives de la vie économique*, XLVI: 13-22. https://doi.org/10.3917/rpve.462.0013

Sorkin, Isaac. 2018. "Ranking firms using revealed preference." *The Quarterly Journal of Economics* 133, no. 3: 1331-1393.

Teulings, Coen and Joop Hartog. 1998. *Corporatism or Competition?: Labour contracts, institutions and wage structures in international comparison*. Cambridge University Press.

Vaghul, Kavya, and Ben Zipperer. September 2016. "Historial state and sub-state minimum wage data." *Washington Center for Equitable Growth,* Working paper.

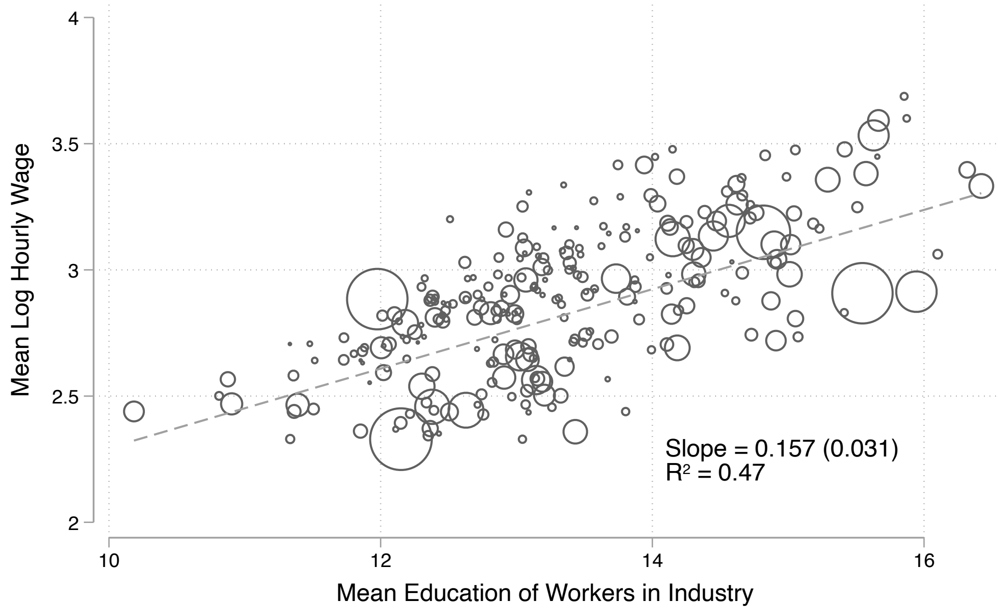
Vaghul, Kavya, and Ben Zipperer. December 10, 2021."State and sub-state historical minimum wage data." https://github.com/benzipperer/historicalminwage. Accessed August 15, 2022.

Vilhuber, Lars. 2018. "LEHD Infrastructure S2014 files in the FSRDC." *US Census Bureau, Center for Economic Studies Discussion Papers* CES 1, no. 2: 3.

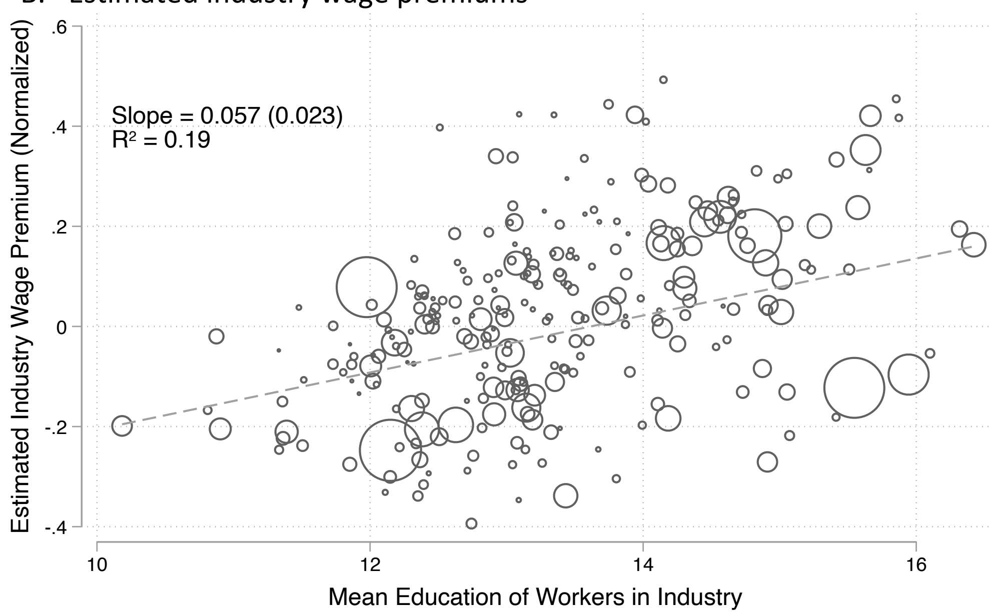
Vilhuber, Lars, and Kevin McKinney. 2014. "LEHD Infrastructure files in the Census RDC-Overview." *Center for Economic Studies, US Census Bureau Working Papers*: 14-26.

Figure 1. Average education of workers and average wages by industry, American Community Survey data

A. Mean log hourly wages

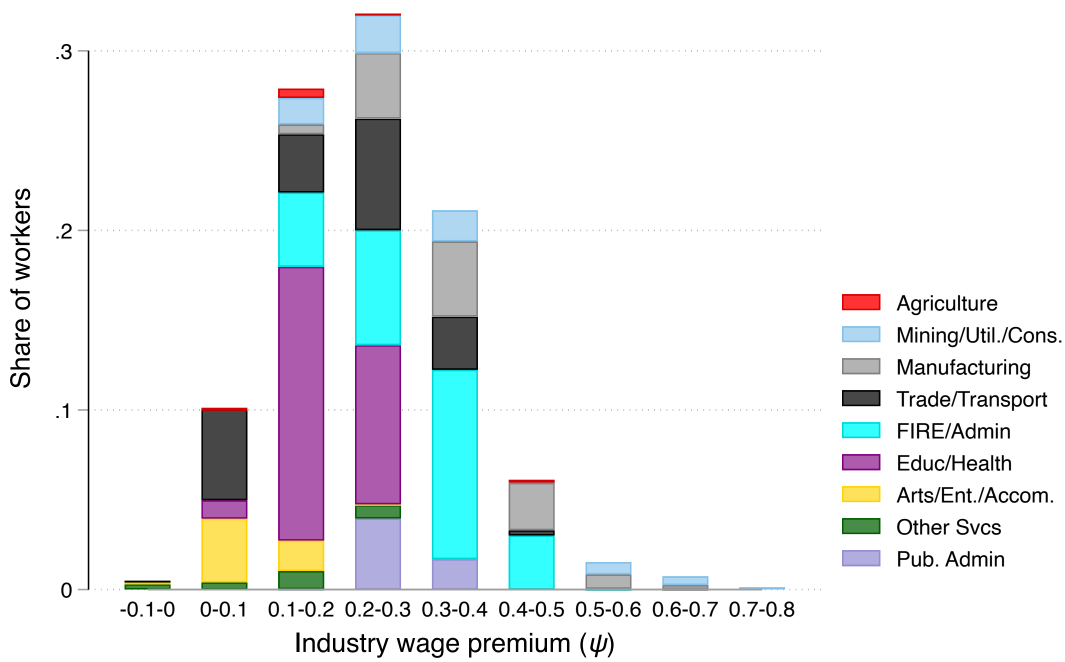


B. Estimated industry wage premiums



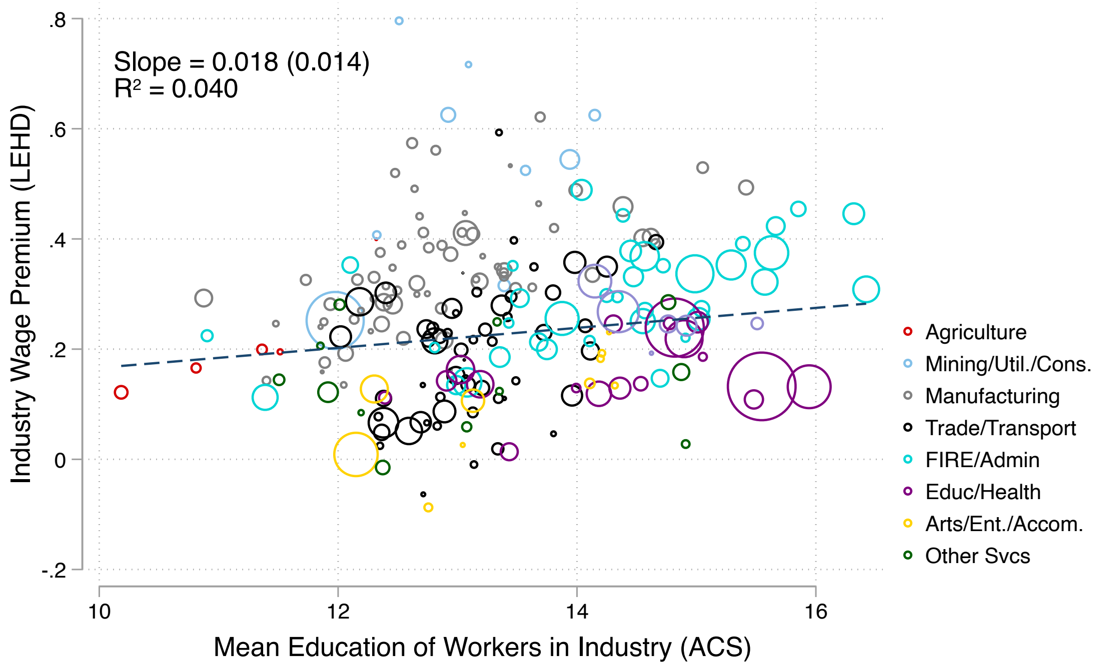
Notes: Samples are pooled 2010-2018 one-year public use samples from the American Community Survey. Individuals aged greater than 62 or with potential experience less than 2 are excluded. Each point corresponds to one of 262 industries that appear in the 2018 data; earlier data are crosswalked to these. Industries are weighted by the number of (weighted) observations. Industry wage premiums are industry fixed effects from a sample-weighted regression that controls for age, education (dummies), college degree field (for college graduates), a quartic in potential experience, and race/ethnicity (five categories), each interacted with gender; indicators for immigration from three regions; immigrant region * years since arrival; separate education indicators for immigrants and for US-educated immigrants; an indicator for presence in one of the 50 largest CZs; and calendar year indicators. Figures show weighted industry-level regressions, with robust standard errors.

Figure 2. Histogram of estimated industry wage premiums



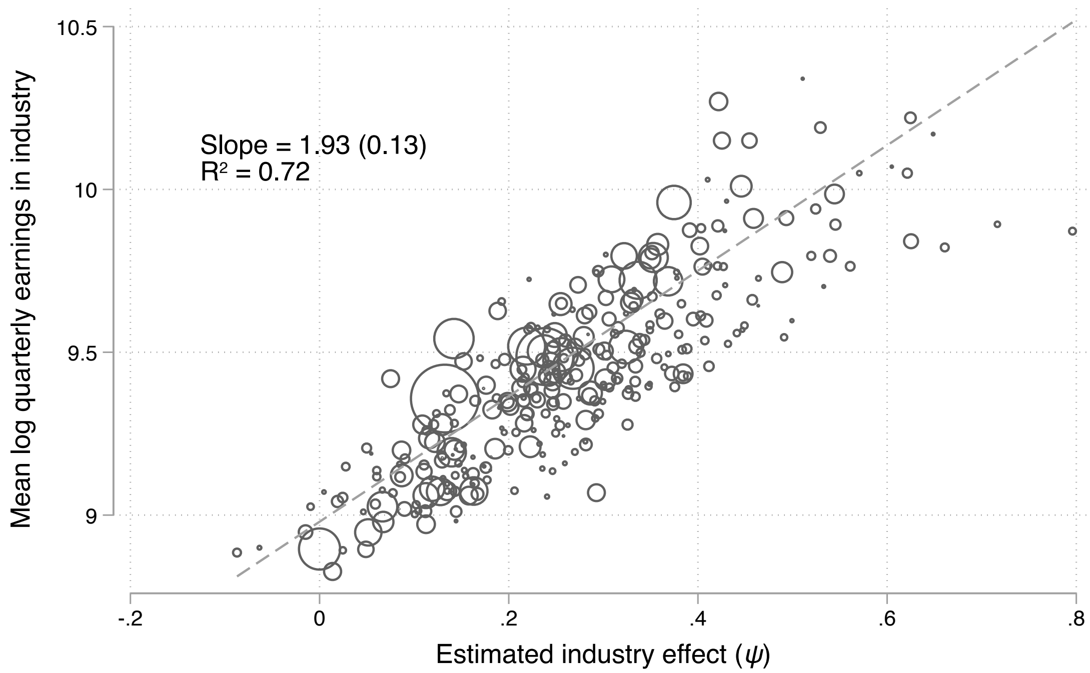
Notes: Figure shows the weighted histogram of estimated industry wage premiums, derived from the "bottom-up" estimator described in the text. N=311 industries are weighted by the number of person-quarter observations. Colors represent the contributions coming from one-digit industry groupings.

Figure 3. Industry wage premium and mean education



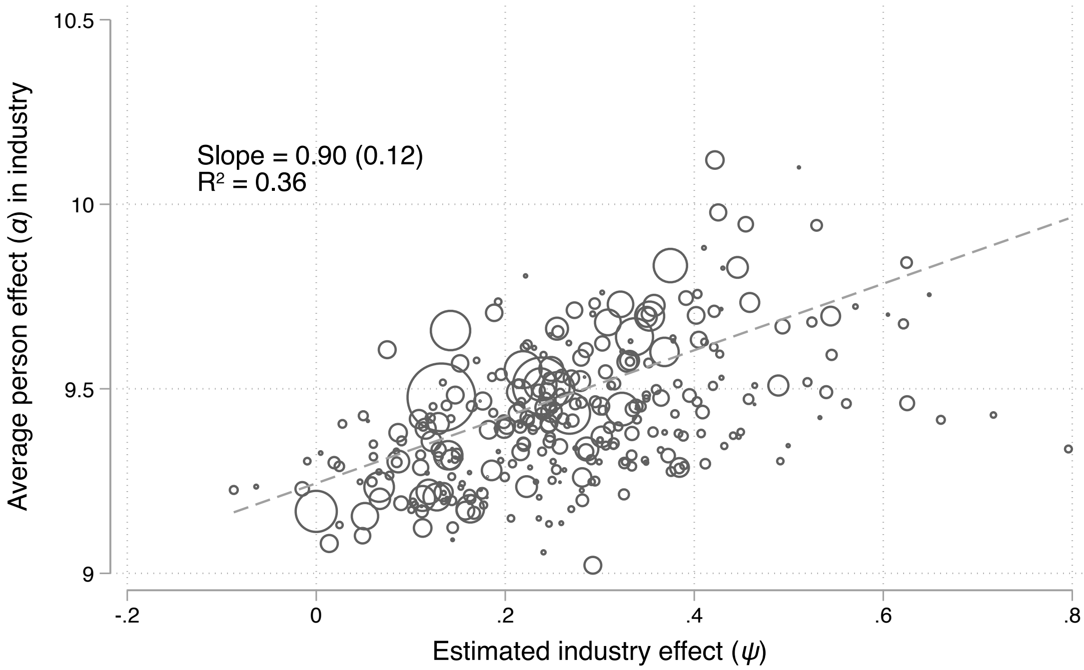
Notes: Figure shows estimated industry wage premia from our "bottom-up" estimator applied to LEHD data against mean education in the industry from the ACS. Industries are grouped into 222 that could be defined consistently in both datasets, averaging over smaller groups where necessary. Regression line is weighted by the number of person-quarter observations in the LEHD sample, and robust standard error is reported.

Figure 4. Industry differentials and mean earnings



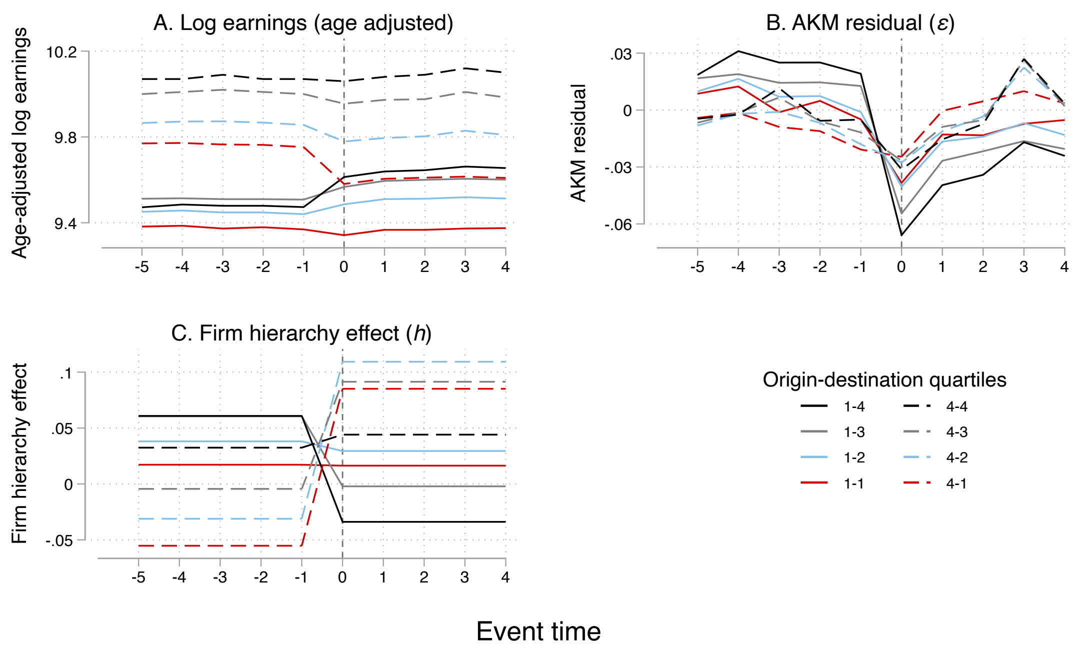
Notes: Figure shows mean log quarterly earnings and estimated industry differential from our "bottom-up" estimator. N=311 industries are weighted by the number of person-quarter observations in our sample. Regression line is weighted and robust standard error is reported.

Figure 5. Industry differentials and average worker effects



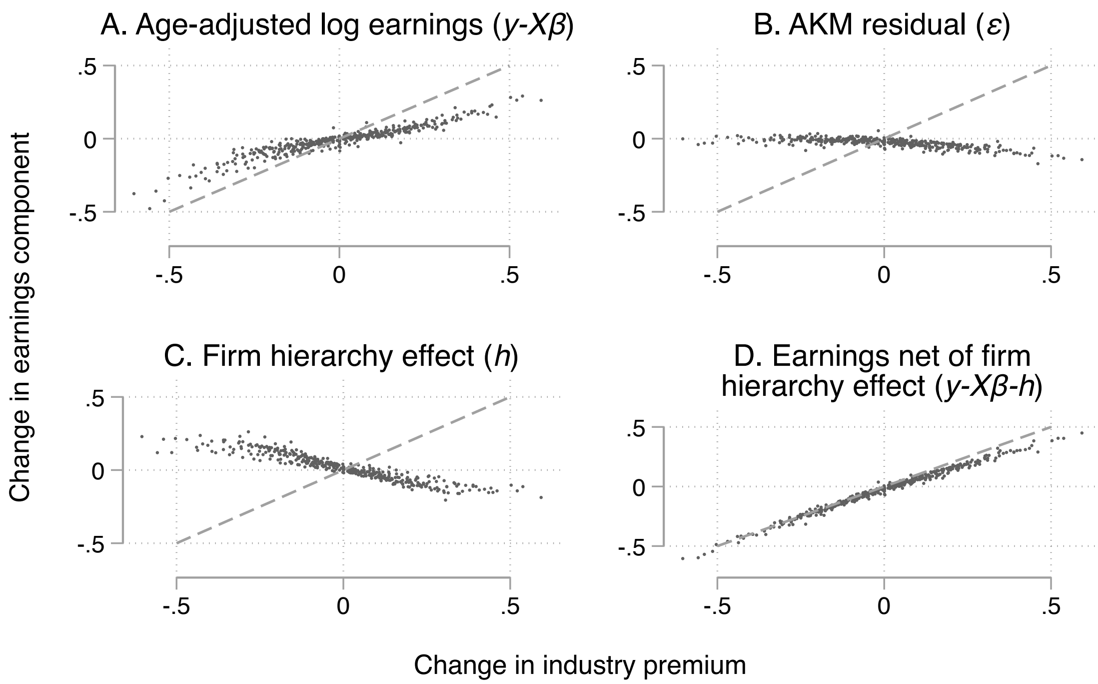
Notes: Figure shows mean of estimated person effect and estimated industry differential from our "bottom-up" estimator. N=311 industries are weighted by the number of person-quarter observations in our sample. Regression line is weighted and robust standard error is reported.

Figure 6. Event studies for earnings of industry movers



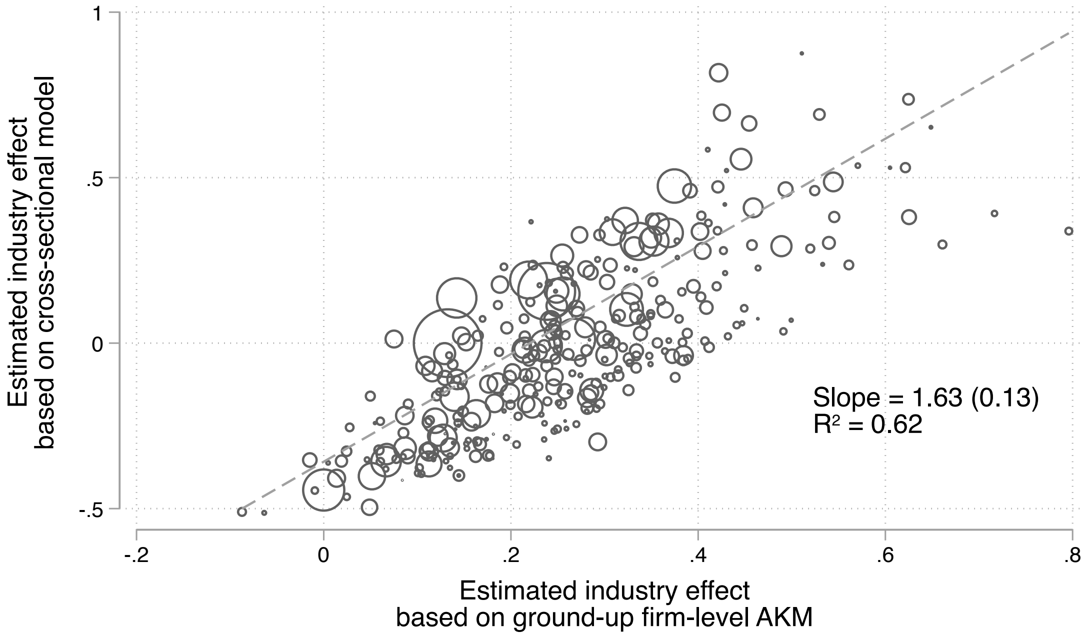
Notes: Figures show event-time means for workers who move between industries within CZs and originate in industries with estimated industry premiums in the top or bottom quartile. See text for definition of the AKM residual and the firm hierarchy effect.

Figure 7. Average earnings changes for industry movers



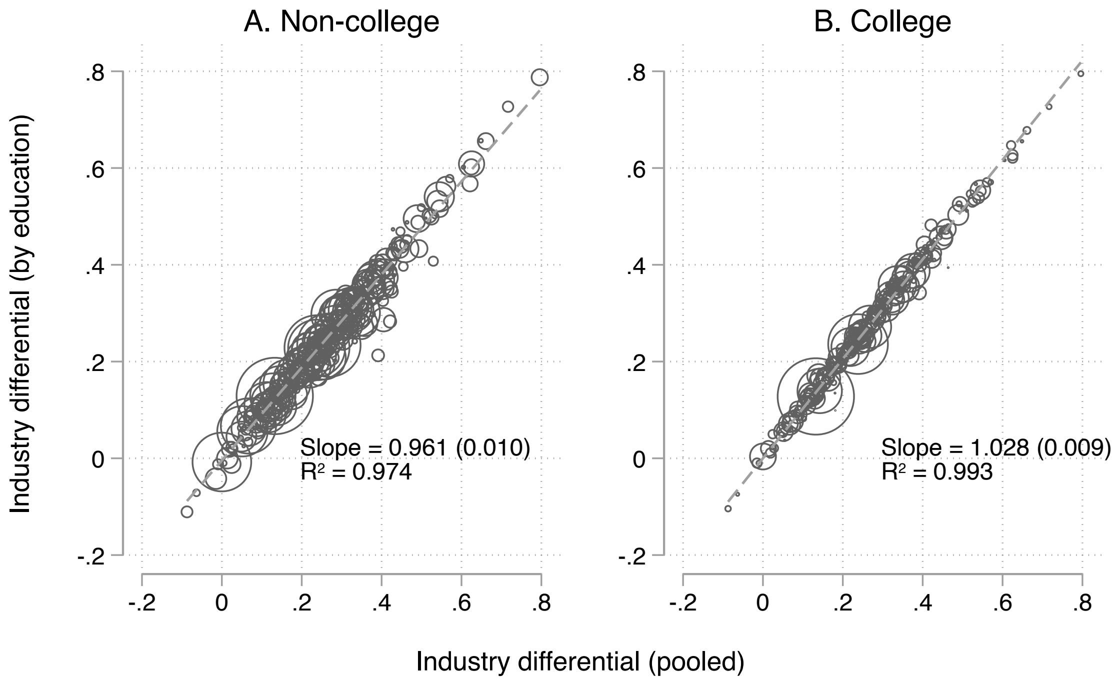
Notes: Within-CZ, between-industry movers are grouped into 400 cells based on vingtiles of their origin and destination industry premiums, from our "bottom-up" estimator. Plot shows, for each cell, the average change in each outcome between the final pre-move quarter and the first post-move quarter.

Figure 8. Comparing cross-sectional and AKM-based estimates of industry premia



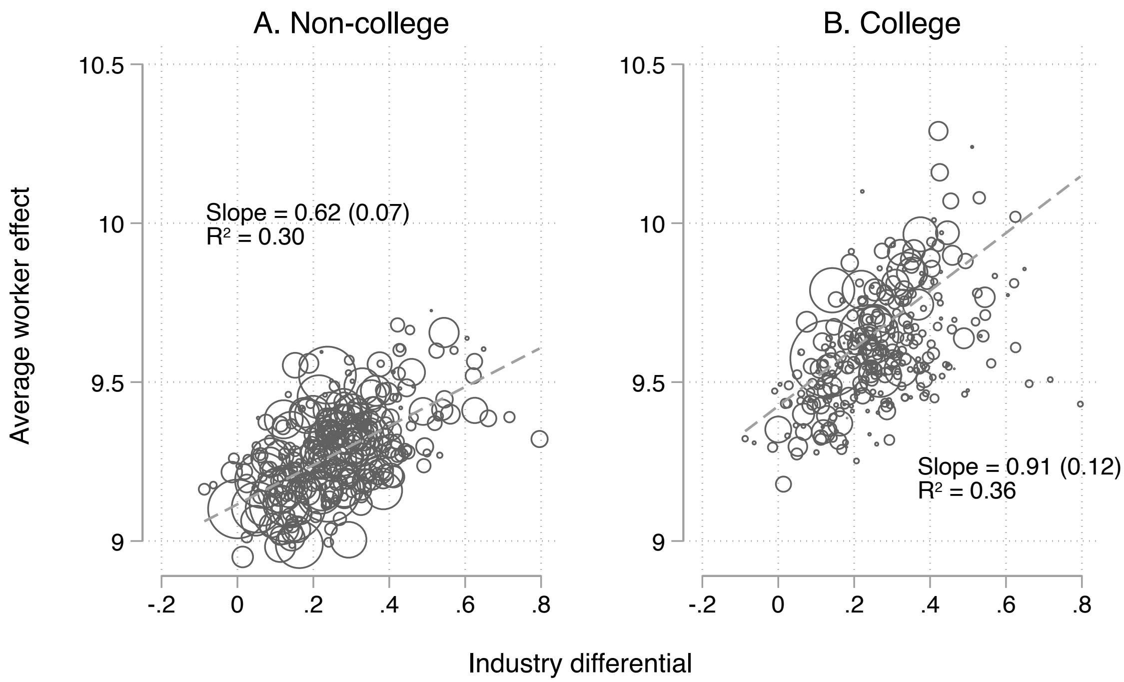
Notes: Estimated industry effects on the X-axis are from our ground-up, firm-level AKM estimator. Those on the Y-axis are from a cross-sectional model estimated on the LEHD data, controlling for gender, race, ethnicity, foreign-born status, age (cubic), and calendar quarter. Regression line is fit to the 311 industries and weighted by the number of person-quarter observations; robust standard error is reported.

Figure 9. Pooled vs. separate estimates of industry premiums by education



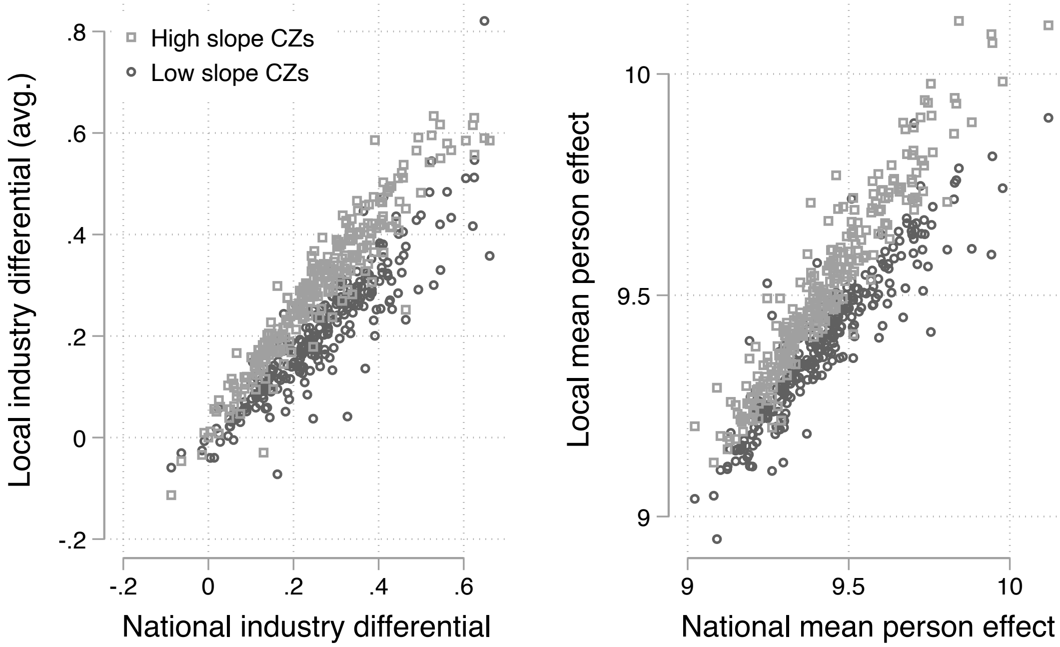
Note: College group includes workers with some college or more. Pooled industry differentials are from our base bottom-up model, averaging AKM firm premiums weighted by the number of person-quarter observations in our main sample. Education-specific industry differentials use the same AKM firm premiums, but weight firms by the number of workers who could be matched to education information from the Decennial Census or ACS and who are of the indicated education group. Firms with no matched workers are omitted. Regression lines are weighted by the number of person-quarter observations in the industry-education group, and robust standard errors are reported.

Figure 10. Average worker effects by education and industry



Note: See notes to Figure 9. Vertical axis is the average of the estimated person effects from the firm-level AKM model, across all workers in the industry who could be matched to education information and were of the indicated education group.

Figure 11. Local industry premiums and worker sorting



Note: High/low slope CZs refer to the 10 CZs with the highest/lowest $\theta_c^{\psi\psi}$s. Each point represents the Ψ_{jc}s averaged across the 10 CZs in each group. To meet disclosure rules, we grouped 38 of the smallest industries into an aggregate industry "Other," accounting for a small fraction of employment.

Table 1. Summary statistics

	Full sample	Industry stayers	Industry switchers	Event study sample
	(1)	(2)	(3)	(4)
Ln(quarterly earnings)	15,510	16,050	14,630	16,350
	(18,020)	(19,710)	(14,860)	(45,670)
Age	42	44	40	40
	(11)	(11)	(10)	(11)
Female	0.47	0.48	0.46	0.46
Foreign born	0.16	0.16	0.16	0.14
Number of CZs in which observed				
1	0.79	0.82	0.72	1.00
2	0.17	0.14	0.22	0.00
3+	0.04	0.03	0.05	0.00
Number of industry switches (within CZs)				
0	0.62	1.00	0.00	0.00
1	0.26	0.00	0.68	1.00
2+	0.12	0.00	0.32	0.00
Quarters observed	25.9	27.2	23.7	10.0
	(7.7)	(7.8)	(7.0)	(0.3)
Number of person-quarter observations (millions)	2,505	1,544	960.4	87.4
Number of unique people (millions)	111.7	65.7	46.1	8.7

Notes: Sample means; standard deviations in parentheses. "Industry switchers" are people observed in more than one industry within a single CZ. "Stayers" may be observed in multiple CZs, potentially in different industries in each, but are observed in only one industry per CZ. Event study sample is workers who switch industries exactly once within a CZ, and are observed for at least five continuous quarters within the CZ before and after the switch.

Table 2. Relationships between worker skill and industry effects

Sample	LEHD			Linked LEHD-ACS				
Dependent variable:	$\bar{\alpha}_j$	ψ_j	$\bar{\alpha}_j$	$\bar{\alpha}_j$	$\bar{\alpha}_j$	$\bar{\alpha}_j$	$E[\bar{\alpha}_j \mid \overline{educ_j}]$	$\bar{\alpha}_j - E[\bar{\alpha}_j \mid \overline{educ_j}]$
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
$\bar{y}_j$	0.62	0.37						
	(0.02)	(0.02)						
ψ_j			0.90	0.91		0.73	0.21	0.70
			(0.12)	(0.14)		(0.08)	(0.14)	(0.08)
Average education					0.10	0.08		
					(0.02)	(0.01)		
N	311	311	311	222	222	222	222	222
R2	0.87	0.72	0.36	0.36	0.49	0.71	0.04	0.42

Note: Observations are industries, using NAICS 4-digit industries in the LEHD data (columns 1-3) and groups that can be defined uniformly in the LEHD and ACS in columns 4-8. Standard errors are robust.

Table 3. Worker experience and the industry hierarchy effect

	Young workers		Older workers	
	(1)	**(2)**	**(3)**	**(4)**
Number of quarters in industry/10	0.012	0.010	0.007	0.006
	(0.002)	(0.001)	(0.001)	(0.001)
(Number of quarters in industry/10)^2	-0.0033	-0.0029	-0.0016	-0.0016
	(0.0006)	(0.0003)	(0.0003)	(0.0002)
Controls for worker, CZ, industry, time FEs	N	Y	N	Y
N (millions of person-quarter observations)	89.8	89.8	421.8	421.8
R2 (adj.)	0.0004	0.7340	0.0002	0.8370
Experience (in quarters) at which slope=0	18.1	17.2	21.8	18.3
Cumulative effect of 5 years of experience	0.011	0.008	0.008	0.005

Notes: Dependent variable in all columns is the hierarchy effect, the difference between the AKM estimate of the firm effect and the industry mean firm effect. Young workers are those who were not yet 26 at the beginning of 2010; older workers are all others in our main sample. Industry experience is the number of quarters to date that the worker has been observed in the industry; this count continues if a worker returns to the same industry after leaving. Standard errors are clustered at the industry level.

Table 4. Comparisons of industry effects from alternative models

	Preferred model	Cross-sectional models			Movers models		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Alternative model controls for:							
Time-varying controls		X	X	X	X	X	X
Time invariant controls			X	X			
CZ FEs				X		X	
Industry-by-CZ FEs							X
Individual FEs					X	X	X
Standard deviation of industry effects	0.122	0.271	0.254	0.240	0.079	0.079	0.082
Regression of alternative model estimates	1.00	1.86	1.63	1.61	0.62	0.62	0.66
on preferred model estimates (N=311)		(0.12)	(0.13)	(0.12)	(0.02)	(0.02)	(0.01)
R^2 (adj)		0.707	0.614	0.672	0.929	0.924	0.954

Note: Preferred model is "ground-up" model, based on averages of firm effects from AKM specification. Regressions are of industry effects from alternative model on industry effects from preferred model, and are weighted by the number of person-quarter observations in the industry. Time-varying controls are a cubic in age and calendar quarter indicators. Time-invariant controls are indicators for female, race (4 categories), ethnicity (Hispanic), and foreign born. In column 7, the alternative model includes industry-by-CZ fixed effects; these are then averaged to the industry level using CZ person-quarter observation counts as weights.

Table 5. Summaries of CZ-level gradients

	$\theta^{\psi\psi}$	$\theta^{\alpha\alpha}$	$\theta^{\alpha\psi}$
	(1)	**(3)**	**(2)**
Mean	0.880	0.863	0.799
SD	0.082	0.097	0.134
Regressions			
$\theta^{\psi\psi}$			
Coeff.		0.36	0.62
SE		(0.21)	(0.20)
R^2 (within-region)		0.14	0.26
$\theta^{\alpha\alpha}$			
Coeff.	0.43		1.14
SE	(0.22)		(0.14)
R^2 (within-region)	0.14		0.74
$\theta^{\alpha\psi}$			
Coeff.	0.42	0.65	
SE	(0.09)	(0.06)	
R^2 (within-region)	0.26	0.74	

Note: θs are slopes of a CZ's estimated industry effects or industry average person effects with respect to the national industry effect or industry average person effect. Summary statistics are for their across-CZ distribution. Regressions are CZ-level regressions of one θ slope (indicated by the column header) on another (row header); each coefficient is from a separate regression. Regressions control for indicators for four regions and for each of approximately 10 composite CZs. N≈60. Reported R2 indicate the share of the within-region variation explained, among the 50 non-composite CZs. Summary statistics and regressions are weighted by the number of person-quarter observations in the CZ, and standard errors are heteroskedasticity-robust.

Table 6. Bivariate relationships between CZ characteristics and industry slopes

	Mean	Regressions			
	[SD]	$\theta^{\psi\psi}$		$\theta^{\alpha\alpha}$	
			R^2		R^2
	(1)	(2)	(3)	(4)	(5)
ln(city size)	17.91	0.03	0.04	0.09	0.50
	[0.84]	(0.03)		(0.02)	
Mean firm effect in CZ($\bar{\psi}_c$)	0.24	1.86	0.72	0.88	0.18
	[0.04]	(0.22)		(0.36)	
Average person effect in CZ($\bar{\alpha}_c$)	9.46	0.20	0.04	0.66	0.67
	[0.13]	(0.22)		(0.07)	
Expected average pay premiums and average worker skills, given industry composition:					
Expected mean firm effect in CZ	0.24	3.98	0.28	2.62	0.14
	[0.01]	(1.02)		(0.86)	
Expected mean person effect in CZ	9.46	1.35	0.13	2.36	0.48
	[0.03]	(0.48)		(0.39)	
Institutional features					
Union coverage	0.13	0.45	0.02	0.61	0.05
	[0.06]	(0.53)		(0.36)	
ln(minimum wage)	2.09	0.24	0.06	0.42	0.23
	[0.11]	(0.14)		(0.10)	

Note: Each coefficient is from a regression of the column variable on the row variable, with indicators for four regions and for each of approximately 10 composite Czs. N≈60. Regressions are weighted by the number of person-quarter observations in the CZ, and standard errors are heteroskedasticity-robust.

Appendix Table A-1: List of Industries, with Mean Eearnings, Mean Worker Effect, and Average Pay Premium

NAICS	Industry Description	Mean Log Earnings	Mean Worker Effect (normalized to mean 0)	Average Pay Premium	Percent of Workforce
	Mean across industries	9.434	0.000	0.236	
	SD	0.278	0.185	0.122	
1111	Oilseed & Grain Farming	9.216	-0.145	0.152	0.03
1112	Vegetable & Melon Farming	9.120	-0.213	0.155	0.04
1113	Fruit & Tree Nut Farming	9.013	-0.270	0.104	0.06
1114	Greenhouse, Nursery & Floriculture Prodn.	9.032	-0.263	0.102	0.08
1119	Other Crop Farming	9.113	-0.215	0.131	0.03
1121	Cattle Ranching & Farming	9.108	-0.271	0.177	0.08
1122	Hog & Pig Farming	9.159	-0.320	0.259	0.02
1123	Poultry & Egg Prodn.	9.143	-0.307	0.236	0.03
1124	Sheep & Goat Farming	9.019	-0.254	0.084	0.00
1125	Aquaculture	9.185	-0.152	0.141	0.00
1129	Other Animal Prodn.	9.150	-0.229	0.174	0.01
1131	Timber Tract Operations	9.555	0.076	0.284	0.00
1132	Forest Nurseries & Gathering of Forest Prods.	9.176	-0.159	0.130	0.00
1133	Logging	9.254	-0.150	0.195	0.04
1141	Fishing	9.643	0.002	0.464	0.00
1142	Hunting & Trapping	9.179	-0.194	0.181	0.00
1151	Support Activities for Crop Prodn.	9.098	-0.255	0.164	0.11
1152	Support Activities for Animal Prodn.	9.178	-0.183	0.162	0.02
1153	Support Activities for Forestry	9.330	-0.063	0.190	0.01
2111	Oil & Gas Extraction	10.220	0.386	0.625	0.18
2121	Coal Mining	9.872	-0.119	0.796	0.07
2122	Metal Ore Mining	9.893	-0.027	0.717	0.05
2123	Nonmetallic Mineral Mining & Quarrying	9.536	-0.077	0.407	0.09
2131	Support Activities for Mining	9.841	0.005	0.625	0.30
2211	Electric Power Generation, Transmission & Distn	9.986	0.241	0.544	0.55
2212	Natural Gas Distribution	9.940	0.225	0.525	0.13
2213	Water, Sewage & Other Systems	9.575	0.058	0.316	0.21
2361	Residential Building Construction	9.399	0.011	0.176	0.46
2362	Nonresidential Building Construction	9.652	0.117	0.329	0.62
2371	Utility System Construction	9.596	0.019	0.365	0.37
2372	Land Subdivision	9.631	0.168	0.267	0.03
2373	Highway, Street & Bridge Construction	9.504	-0.003	0.301	0.45
2379	Other Heavy & Civil Engineering Construction	9.649	0.058	0.383	0.09
2381	Foundation, Structure & Building Ext. Contractors	9.347	-0.064	0.199	0.53
2382	Building Equipment Contractors	9.503	0.056	0.237	1.64
2383	Building Finishing Contractors	9.324	-0.067	0.183	0.49
2389	Other Specialty Trade Contractors	9.409	-0.049	0.246	0.45
3111	Animal Food Mfg	9.454	-0.122	0.365	0.05
3112	Grain & Oilseed Milling	9.582	-0.073	0.449	0.07

Appendix Table A-1: List of Industries, with Mean Earnings, Mean Worker Effect, and Average Pay Premium (continued)

NAICS	Industry Description	Mean Log Earnings	Mean Worker Effect (normalized to mean 0)	Average Pay Premium	Percent of Workforce
3113	Sugar & Confectionery Product Mfg	9.395	-0.149	0.348	0.06
3114	Fruit & Vegetable Preserving & Specialty Food Mfg	9.278	-0.242	0.326	0.16
3115	Dairy Product Mfg	9.457	-0.159	0.412	0.14
3116	Animal Slaughtering & Pross.	9.069	-0.434	0.293	0.44
3117	Seafood Product Preparation & Packaging	9.226	-0.231	0.281	0.02
3118	Bakeries & Tortilla Mfg	9.217	-0.258	0.282	0.21
3119	Other Food Mfg	9.373	-0.157	0.326	0.17
3121	Beverage Mfg	9.452	-0.061	0.310	0.18
3122	Tobacco Mfg	9.702	-0.034	0.533	0.01
3131	Fiber, Yarn & Thread Mills	9.057	-0.399	0.241	0.03
3132	Fabric Mills	9.194	-0.282	0.270	0.05
3133	Textile & Fabric Finishing & Fabric Coating Mills	9.219	-0.208	0.233	0.03
3141	Textile Furnishings Mills	9.135	-0.322	0.246	0.05
3149	Other Textile Product Mills	9.074	-0.261	0.135	0.05
3151	Apparel Knitting Mills	8.982	-0.365	0.144	0.01
3152	Cut & Sew Apparel Mfg	9.122	-0.194	0.143	0.07
3159	Apparel Accessories & Other Apparel Mfg	9.072	-0.252	0.138	0.01
3161	Leather & Hide Tanning & Finishing	9.243	-0.209	0.258	0.00
3162	Footwear Mfg	9.094	-0.264	0.162	0.01
3169	Other Leather & Allied Product Mfg	9.030	-0.281	0.118	0.01
3211	Sawmills & Wood Preservation	9.252	-0.204	0.250	0.08
3212	Veneer, Plywood & Engineered Wood Product Mfg	9.297	-0.208	0.292	0.07
3219	Other Wood Product Mfg	9.146	-0.239	0.176	0.17
3221	Pulp, Paper & Paperboard Mills	9.764	0.004	0.561	0.12
3222	Converted Paper Product Mfg	9.435	-0.137	0.372	0.29
3231	Printing & Related Support Activities	9.282	-0.127	0.216	0.42
3241	Petroleum & Coal Prods. Mfg	10.050	0.220	0.621	0.13
3251	Basic Chemical Mfg	9.892	0.136	0.545	0.16
3252	Resin, Synthetic Rubber, Fibers & Filaments Mfg	9.796	0.062	0.520	0.10
3253	Pesticide, Fertilizer & Other Agric.l Chemical Mfg	9.727	0.053	0.464	0.04
3254	Pharmaceutical & Medicine Mfg	9.912	0.213	0.493	0.31
3255	Paint, Coating & Adhesive Mfg	9.568	0.018	0.349	0.06
3256	Soap, Cleaning Compound & Toilet Prep. Mfg	9.499	-0.039	0.339	0.10
3259	Other Chemical Product & Preparation Mfg	9.555	-0.024	0.380	0.09
3261	Plastics Product Mfg	9.292	-0.196	0.281	0.53
3262	Rubber Product Mfg	9.427	-0.172	0.384	0.14
3271	Clay Product & Refractory Mfg	9.349	-0.154	0.299	0.04
3272	Glass & Glass Product Mfg	9.427	-0.164	0.389	0.08
3273	Cement & Concrete Product Mfg	9.420	-0.104	0.315	0.17
3274	Lime & Gypsum Product Mfg	9.597	-0.110	0.499	0.02
3279	Other Nonmetallic Mineral Product Mfg	9.414	-0.110	0.315	0.07
3311	Iron & Steel Mills & Ferroalloy Mfg	9.822	-0.040	0.661	0.10

Appendix Table A-1: List of Industries, with Mean Earnings, Mean Worker Effect, and Average Pay Premium (continued)

NAICS	Industry Description	Mean Log Earnings	Mean Worker Effect (normalized to mean 0)	Average Pay Premium	Percent of Workforce
3312	Steel Product Mfg from Purchased Steel	9.526	-0.111	0.432	0.06
3313	Alumina & Aluminum Prodn. & Pross.	9.546	-0.152	0.491	0.06
3314	Nonferrous Metal (except Alum.) Prodn. & Pross.	9.559	-0.083	0.441	0.07
3315	Foundries	9.394	-0.180	0.376	0.12
3321	Forging & Stamping	9.392	-0.110	0.306	0.10
3322	Cutlery & Handtool Mfg	9.394	-0.094	0.301	0.04
3323	Architectural & Structural Metals Mfg	9.349	-0.112	0.258	0.33
3324	Boiler, Tank & Shipping Container Mfg	9.508	-0.077	0.383	0.09
3325	Hardware Mfg	9.358	-0.117	0.274	0.02
3326	Spring & Wire Product Mfg	9.296	-0.153	0.251	0.04
3327	Mchn. Shops; Turned Product; Screw, Nut & Bolt Mfg	9.383	-0.055	0.246	0.34
3328	Coating, Engraving, Heat Treating & Allied Activities	9.275	-0.175	0.254	0.12
3329	Other Fabricated Metal Product Mfg	9.458	-0.077	0.334	0.27
3331	Agriculture, Construction & Mining Machinery Mfg	9.598	-0.019	0.409	0.25
3332	Industrial Machinery Mfg	9.641	0.118	0.332	0.11
3333	Commercial & Service Industry Machinery Mfg	9.557	0.053	0.312	0.09
3334	HVAC & Commercial Refrigeration Equipment Mfg	9.351	-0.148	0.291	0.13
3335	Metalworking Machinery Mfg	9.475	0.003	0.274	0.18
3336	Engine, Turbine & Power Transmission Equip. Mfg	9.675	0.052	0.420	0.10
3339	Other General Purpose Machinery Mfg	9.536	-0.003	0.338	0.27
3341	Computer & Peripheral Equipment Mfg	10.190	0.487	0.530	0.18
3342	Communications Equipment Mfg	9.881	0.301	0.403	0.11
3343	Audio & Video Equipment Mfg	9.692	0.173	0.333	0.02
3344	Semiconductor & Other Electronic Component Mfg	9.763	0.177	0.405	0.40
3345	Nav., Measuring, Medical & Control Instruments Mfg	9.826	0.243	0.402	0.44
3346	Mfg & Reproducing Magnetic & Optical Media	9.728	0.174	0.378	0.02
3351	Electric Lighting Equipment Mfg	9.401	-0.095	0.300	0.05
3352	Hhld. Appliance Mfg	9.386	-0.154	0.327	0.07
3353	Electrical Equipment Mfg	9.538	-0.004	0.344	0.15
3359	Other Electrical Equipment & Component Mfg	9.481	-0.073	0.356	0.14
3361	Motor Veh. Mfg	9.796	0.035	0.540	0.22
3362	Motor Veh. Body & Trailer Mfg	9.311	-0.206	0.295	0.12
3363	Motor Veh. Parts Mfg	9.434	-0.169	0.384	0.55
3364	Aerospace Product & Parts Mfg	9.911	0.278	0.459	0.56
3365	Railroad Rolling Stock Mfg	9.569	-0.086	0.447	0.03
3366	Ship & Boat Building	9.511	-0.084	0.388	0.13
3369	Other Transportation Equipment Mfg	9.495	-0.074	0.369	0.03
3371	Hhld & Institutional Furniture & Kitchen Cabinet Mfg	9.127	-0.245	0.162	0.20
3372	Office Furniture (including Fixtures) Mfg	9.312	-0.126	0.239	0.10
3379	Other Furniture Related Product Mfg	9.186	-0.250	0.236	0.03
3391	Medical Equipment & Supplies Mfg	9.518	-0.011	0.335	0.31
3399	Other Miscellaneous Mfg	9.311	-0.107	0.220	0.25

Appendix Table A-1: List of Industries, with Mean Earnings, Mean Worker Effect, and Average Pay Premium (continued)

NAICS	Industry Description	Mean Log Earnings	Mean Worker Effect (normalized to mean 0)	Average Pay Premium	Percent of Workforce
4231	Mo. Vehs. & Parts & Supplies Mcht. Whlsalers	9.390	-0.034	0.223	0.28
4232	Furniture & Home Furnishing Mcht. Whlsalers	9.423	0.009	0.217	0.09
4233	Lumber & Construction Materials Mcht. Whlsalers	9.424	-0.023	0.238	0.19
4234	Prof. & Comm. Equip. & Supplies Mcht. Whlsalers	9.788	0.239	0.350	0.63
4235	Metal & Mineral (except Petroleum) Mcht. Whlsalers	9.491	-0.014	0.304	0.12
4236	Hhld. Appliances & Elect. Goods Mcht. Whlsalers	9.667	0.167	0.303	0.32
4237	Hardware, Plumbing, Heating Equip. Mcht. Whlsalers	9.477	0.039	0.236	0.23
4238	Machinery, Equipment & Supplies Mcht. Whlsalers	9.547	0.065	0.279	0.63
4239	Miscellaneous Durable Goods Mcht. Whlsalers	9.352	-0.058	0.216	0.24
4241	Paper & Paper Product Mcht. Whlsalers	9.482	0.033	0.258	0.11
4242	Drugs & Druggists' Sundries Mcht. Whlsalers	9.888	0.254	0.421	0.19
4243	Apparel, Piece Goods & Notions Mcht. Whlsalers	9.447	0.058	0.214	0.12
4244	Grocery & Related Product Mcht. Whlsalers	9.416	-0.086	0.302	0.66
4245	Farm Product Raw Material Mcht. Whlsalers	9.410	-0.054	0.266	0.05
4246	Chemical & Allied Prods. Mcht. Whlsalers	9.671	0.114	0.352	0.12
4247	Petroleum & Petroleum Prods. Mcht. Whlsalers	9.584	0.027	0.349	0.08
4248	Beer, Wine & Alcoholic Beverage Mcht. Whlsalers	9.509	0.009	0.295	0.19
4249	Miscellaneous Nondurable Goods Mcht. Whlsalers	9.348	-0.045	0.199	0.26
4251	Wholesale Electronic Markets & Agents & Brokers	9.831	0.270	0.358	0.70
4411	Automobile Dealers	9.448	0.035	0.215	0.99
4412	Other Motor Veh. Dealers	9.260	-0.058	0.113	0.11
4413	Automotive Parts, Accessories & Tire Stores	9.134	-0.169	0.111	0.37
4421	Furniture Stores	9.209	-0.132	0.149	0.16
4422	Home Furnishings Stores	9.278	-0.035	0.121	0.13
4431	Electronics & Appliance Stores	9.354	-0.048	0.230	0.36
4441	Building Material & Supplies Dealers	9.122	-0.153	0.087	0.75
4442	Lawn & Garden Equipment & Supplies Stores	9.118	-0.140	0.060	0.09
4451	Grocery Stores	9.026	-0.222	0.067	1.38
4452	Specialty Food Stores	9.068	-0.191	0.078	0.09
4453	Beer, Wine & Liquor Stores	9.026	-0.152	-0.009	0.07
4461	Health & Personal Care Stores	9.236	-0.064	0.116	0.63
4471	Gasoline Stations	8.895	-0.354	0.049	0.37
4481	Clothing Stores	9.168	-0.121	0.130	0.32
4482	Shoe Stores	9.159	-0.148	0.147	0.06
4483	Jewelry, Luggage & Leather Goods Stores	9.282	-0.037	0.143	0.08
4511	Sporting Goods, Hobby & Musical Instrument Stores	9.042	-0.156	0.019	0.20
4512	Book Stores & News Dealers	9.010	-0.208	0.046	0.04
4521	Department Stores	8.979	-0.254	0.067	0.65
4529	General Merch Stores, Wrhse. Clubs & Supercenters	8.947	-0.301	0.052	1.09
4531	Florists	8.900	-0.221	-0.064	0.02
4532	Office Supplies, Stationery & Gift Stores	9.154	-0.135	0.111	0.13
4533	Used Merchandise Stores	8.892	-0.325	0.025	0.06

Appendix Table A-1: List of Industries, with Mean Earnings, Mean Worker Effect, and Average Pay Premium (continued)

NAICS	Industry Description	Mean Log Earnings	Mean Worker Effect (normalized to mean 0)	Average Pay Premium	Percent of Workforce
4539	Other Miscellaneous Store Retailers	9.117	-0.155	0.085	0.15
4541	Electronic Shopping & Mail-Order Houses	9.454	0.016	0.242	0.26
4542	Vending Machine Operators	9.096	-0.238	0.135	0.03
4543	Direct Selling Establishments	9.359	-0.067	0.229	0.11
4811	Scheduled Air Transportation	9.627	0.250	0.188	0.42
4812	Nonscheduled Air Transportation	9.745	0.247	0.293	0.04
4821	Rail Transportation	9.485	-0.010	0.338	0.00
4831	Deep Sea, Coastal & Great Lakes Water Trans.	9.746	0.182	0.378	0.04
4832	Inland Water Transportation	9.706	0.074	0.429	0.03
4841	General Freight Trucking	9.375	-0.119	0.286	0.83
4842	Specialized Freight Trucking	9.367	-0.126	0.285	0.34
4851	Urban Transit Systems	9.602	0.026	0.395	0.25
4852	Interurban & Rural Bus Transportation	9.276	-0.176	0.263	0.02
4853	Taxi & Limousine Service	9.077	-0.181	0.066	0.04
4854	School & Employee Bus Transportation	9.054	-0.166	0.024	0.14
4855	Charter Bus Industry	9.115	-0.184	0.117	0.02
4859	Other Transit & Ground Passenger Transportation	9.003	-0.284	0.101	0.06
4861	Pipeline Transportation of Crude Oil	10.170	0.299	0.649	0.01
4862	Pipeline Transportation of Natural Gas	10.050	0.267	0.571	0.03
4869	Other Pipeline Transportation	10.070	0.245	0.605	0.01
4871	Scenic & Sightseeing Transportation, Land	9.203	-0.119	0.144	0.01
4872	Scenic & Sightseeing Transportation, Water	9.189	-0.043	0.055	0.01
4879	Scenic & Sightseeing Transportation, Other	9.389	0.011	0.173	0.00
4881	Support Activities for Air Transportation	9.421	-0.037	0.261	0.18
4882	Support Activities for Rail Transportation	9.403	-0.156	0.349	0.03
4883	Support Activities for Water Transportation	9.763	0.138	0.427	0.08
4884	Support Activities for Road Transportation	9.254	-0.159	0.208	0.08
4885	Freight Transportation Arrangement	9.448	-0.001	0.248	0.18
4889	Other Support Activities for Transportation	9.261	-0.169	0.226	0.02
4911	Postal Service	9.138	-0.250	0.181	0.00
4921	Couriers & Express Delivery Services	9.472	0.113	0.152	0.42
4922	Local Messengers & Local Delivery	9.138	-0.223	0.153	0.04
4931	Warehousing & Storage	9.210	-0.221	0.223	0.68
5111	Newspaper, Periodical, Book & Directory Publishers	9.518	0.073	0.270	0.37
5112	Software Publishers	10.150	0.490	0.455	0.33
5121	Motion Picture & Video Industries	9.749	0.275	0.295	0.17
5122	Sound Recording Industries	9.617	0.191	0.248	0.01
5151	Radio & Television Broadcasting	9.650	0.199	0.256	0.19
5152	Cable & Other Subscription Programming	9.766	0.171	0.411	0.07
5171	Wired Telecommunications Carriers	9.746	0.053	0.489	0.64
5172	Wireless Telecommunications Carriers	9.661	0.016	0.458	0.15
5174	Satellite Telecommunications	9.873	0.260	0.429	0.01

Appendix Table A-1: List of Industries, with Mean Earnings, Mean Worker Effect, and Average Pay Premium (continued)

NAICS	Industry Description	Mean Log Earnings	Mean Worker Effect (normalized to mean 0)	Average Pay Premium	Percent of Workforce
5179	Other Telecommunications	9.765	0.157	0.421	0.09
5182	Data Pross., Hosting & Related Services	9.806	0.246	0.351	0.28
5191	Other Information Services	9.875	0.290	0.391	0.29
5211	Monetary Authorities-Central Bank	10.030	0.426	0.410	0.02
5221	Depository Credit Intermediation	9.493	0.045	0.256	1.71
5222	Nondepository Credit Intermediation	9.663	0.121	0.332	0.57
5223	Activities Related to Credit Intermediation	9.602	0.090	0.306	0.27
5231	Sec. & Commidity Contracts Intermed. & Brokerage	10.270	0.664	0.422	0.48
5232	Sec. & Commidity Exchanges	10.340	0.644	0.511	0.01
5239	Other Financial Investment Activities	10.150	0.522	0.425	0.41
5241	Insurance Carriers	9.718	0.144	0.369	1.28
5242	Agencies, Brokerages & Insurance Related Activities	9.553	0.098	0.249	0.89
5251	Insurance & Employee Benefit Funds	9.619	0.145	0.324	0.03
5259	Other Investment Pools & Funds	9.964	0.371	0.430	0.02
5311	Lessors of Real Estate	9.333	-0.056	0.202	0.42
5312	Offices of Real Estate Agents & Brokers	9.478	0.083	0.196	0.20
5313	Activities Related to Real Estate	9.390	-0.022	0.213	0.48
5321	Automotive Equipment Rental & Leasing	9.348	-0.087	0.248	0.14
5322	Consumer Goods Rental	9.199	-0.195	0.200	0.11
5323	General Rental Centers	9.361	-0.055	0.216	0.03
5324	Comm. Industrial Equipment Rental & Leasing	9.618	0.042	0.360	0.12
5331	Lessors of Nonfin. Intangible Assets (except Copyrt)	9.800	0.305	0.303	0.02
5411	Legal Services	9.724	0.224	0.309	1.04
5412	Accounting, Tax Prep, Bookkeeping & Payroll Svcs	9.648	0.207	0.255	0.76
5413	Architectural, Engineering & Related Services	9.791	0.241	0.353	1.33
5414	Specialized Design Services	9.570	0.157	0.221	0.10
5415	Computer Systems Design & Related Services	9.960	0.378	0.375	1.76
5416	Management, Scientific & Technical Consulting Svcs	9.796	0.273	0.322	1.02
5417	Scientific Research & Development Services	10.010	0.373	0.446	0.68
5418	Advertising, Public Relations & Related Services	9.707	0.257	0.273	0.38
5419	Other Professional, Scientific & Technical Services	9.372	0.027	0.147	0.45
5511	Management of Companies & Enterprises	9.721	0.185	0.337	2.16
5611	Office Administrative Services	9.613	0.128	0.280	0.39
5612	Facilities Support Services	9.364	-0.166	0.334	0.11
5613	Employment Services	9.193	-0.136	0.140	1.28
5614	Business Support Services	9.204	-0.177	0.186	0.60
5615	Travel Arrangement & Reservation Services	9.409	-0.005	0.215	0.17
5616	Investigation & Security Services	9.074	-0.236	0.135	0.52
5617	Services to Buildings & Dwellings	9.060	-0.253	0.113	0.98
5619	Other Support Services	9.385	-0.042	0.224	0.19
5621	Waste Collection	9.410	-0.136	0.334	0.15
5622	Waste Treatment & Disposal	9.612	0.009	0.403	0.13

Appendix Table A-1: List of Industries, with Mean Earnings, Mean Worker Effect, and Average Pay Premium (continued)

NAICS	Industry Description	Mean Log Earnings	Mean Worker Effect (normalized to mean 0)	Average Pay Premium	Percent of Workforce
5629	Remediation & Other Waste Management Services	9.465	-0.058	0.320	0.11
6111	Elementary & Secondary Schools	9.358	0.021	0.132	7.22
6112	Junior Colleges	9.419	0.150	0.075	0.45
6113	Colleges, Universities & Professional Schools	9.542	0.202	0.142	2.44
6114	Business Schools & Computer & Mgmt. Training	9.571	0.136	0.241	0.06
6115	Technical & Trade Schools	9.359	-0.021	0.187	0.10
6116	Other Schools & Instruction	9.206	-0.029	0.050	0.13
6117	Educational Support Services	9.464	0.076	0.186	0.10
6211	Offices of Physicians	9.519	0.096	0.219	2.15
6212	Offices of Dentists	9.278	-0.050	0.129	0.70
6213	Offices of Other Health Practitioners	9.278	-0.038	0.109	0.52
6214	Outpatient Care Centers	9.438	-0.018	0.249	0.67
6215	Medical & Diagnostic Laboratories	9.430	-0.042	0.271	0.23
6216	Home Health Care Services	9.200	-0.148	0.143	0.60
6219	Other Ambulatory Health Care Services	9.311	-0.104	0.219	0.22
6221	General Medical & Surgical Hospitals	9.487	0.052	0.238	5.08
6222	Psychiatric & Substance Abuse Hospitals	9.341	-0.099	0.247	0.21
6223	Specialty (except Psych. & Substance Abuse) Hsptls	9.536	0.079	0.260	0.23
6231	Nursing Care Facilities (Skilled Nursing Facilities)	9.074	-0.281	0.163	1.19
6232	Resid. Facilities (Disab., Mental Health, Sub. Abuse)	9.060	-0.286	0.158	0.49
6233	Assisted Living & Comm Care Facilities for the Elderly	8.972	-0.333	0.113	0.49
6239	Other Residential Care Facilities	9.073	-0.259	0.143	0.13
6241	Individual & Family Services	9.081	-0.235	0.119	0.94
6242	Food, Housing, Emergency & Other Relief Services	9.183	-0.139	0.129	0.11
6243	Vocational Rehabilitation Services	9.012	-0.288	0.112	0.22
6244	Child Day Care Services	8.827	-0.375	0.014	0.45
7111	Performing Arts Companies	9.374	0.061	0.134	0.06
7112	Spectator Sports	9.656	0.280	0.193	0.07
7113	Promoters of Perf. Arts, Sports & Similar Events	9.482	0.121	0.170	0.05
7114	Agents & Managers for Artists, Athletes, Entertainers	9.724	0.350	0.222	0.02
7115	Independent Artists, Writers & Performers	9.576	0.155	0.231	0.03
7121	Museums, Historical Sites & Similar Institutions	9.323	-0.001	0.138	0.15
7131	Amusement Parks & Arcades	9.137	-0.106	0.060	0.09
7132	Gambling Industries	9.065	-0.292	0.167	0.22
7139	Other Amusement & Recreation Industries	9.199	-0.075	0.087	0.48
7211	Traveler Accommodation	9.072	-0.249	0.128	1.15
7212	RV (Recreational Vehicle) Parks & Recreational Camps	9.071	-0.130	0.005	0.02
7213	Room & Boarding Houses, Dormitories & Camps	9.027	-0.274	0.111	0.01
7223	Special Food Services	9.019	-0.266	0.090	0.29
7224	Drinking Places (Alcoholic Beverages)	8.885	-0.230	-0.087	0.10
7225	Restaurants & Other Eating Places	8.896	-0.288	0.000	2.66
8111	Automotive R&M	9.224	-0.098	0.122	0.60

Appendix Table A-1: List of Industries, with Mean Earnings, Mean Worker Effect, and Average Pay Premium (continued)

NAICS	Industry Description	Mean Log Earnings	Mean Worker Effect (normalized to mean 0)	Average Pay Premium	Percent of Workforce
8112	Electronic & Precision Equipment R&M	9.458	0.012	0.250	0.09
8113	Commercial & Industrial Machinery & Equip. R&M	9.494	0.007	0.281	0.17
8114	Personal & Hhld. Goods R&M	9.156	-0.125	0.085	0.04
8121	Personal Care Services	8.948	-0.227	-0.015	0.29
8122	Death Care Services	9.312	-0.004	0.124	0.07
8123	Drycleaning & Laundry Services	9.011	-0.332	0.144	0.18
8129	Other Personal Services	9.034	-0.208	0.059	0.13
8131	Religious Organizations	9.149	-0.051	0.028	0.09
8132	Grantmaking & Giving Services	9.577	0.164	0.224	0.12
8133	Social Advocacy Organizations	9.351	-0.003	0.165	0.16
8134	Civic & Social Organizations	9.173	-0.097	0.090	0.13
8139	Business, Professional, Labor, Political Organizations	9.624	0.149	0.285	0.31
8141	Private Hhld.s	9.075	-0.307	0.206	0.07
9211	Executive, Legislative & General Government Support	9.451	-0.022	0.268	2.66
9221	Justice, Public Order & Safety Activities	9.516	-0.011	0.324	1.69
9231	Administration of Human Resource Programs	9.428	-0.007	0.243	0.63
9241	Administration of Environmental Quality Programs	9.515	0.072	0.247	0.21
9251	Admin. of Housing, Planning & Community Dev.	9.433	-0.011	0.251	0.08
9261	Administration of Economic Programs	9.475	0.038	0.245	0.38
9281	National Security & International Affairs	9.267	-0.126	0.193	0.02

Table A-2. Testing the role of industry composition

	$\theta^{\psi\psi}$		$\theta^{\alpha\alpha}$	
	(1)	**(2)**	**(3)**	**(4)**
Mean pay premium in CZ $(\bar{\psi}_c)$				
Expected given industry shares	2.23		1.97	
	(0.98)		(0.86)	
Actual - expected	1.78		0.66	
	(0.24)		(0.42)	
Average person effect in CZ $(\bar{\alpha}_c)$				
Expected given industry shares		1.92		0.84
		(1.23)		(0.33)
Actual - expected		-0.23		0.62
		(0.39)		(0.11)
R-squared	0.80	0.39	0.67	0.86
Adjusted R-squared	0.74	0.19	0.56	0.81

Note: N≈60. Each coefficient is from a regression of the column variable on the row variables, with indicators for four regions and for each of approximately 10 composite CZs. Standard errors are approximated.