Industry Wage Differentials: A Firm-Based Approach

David Card	Jesse Rothstein	Moises Yi
UC Berkeley	UC Berkeley	US Census Bureau

July 2023

ABSTRACT

We revisit the estimation of industry wage differentials using linked employer-employee data from the U.S. LEHD program. Building on recent advances in the measurement of *employer* wage premiums, we define the industry wage effect as the employment-weighted average workplace premium in that industry. We show that cross-sectional estimates of industry differentials *overstate* the pay premiums due to unmeasured worker heterogeneity. Conversely, estimates based on industry movers *understate* the true premiums, due to unmeasured heterogeneity in pay premiums within industries. Industry movers who switch to higher-premium industries tend to leave firms in the origin sector that pay above-average premiums and move to firms in the destination sector with below-average premiums (and vice versa), attenuating the measured industry effects. Our preferred estimates reveal substantial heterogeneity in narrowly-defined industry premiums, with a standard deviation of 12%. On average, workers in higher-paying industries have higher observed and unobserved skills, widening between-industry wage inequality. There are also small but systematic differences in industry premiums across cities, with a wider distribution of pay premiums and more worker sorting in cities with more high-premium firms and high-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, as well as seminar and conference participants at UC Irvine, CU Denver, GW, BYU, and SOLE for helpful comments and conversations. We are also grateful to Nicole Gandre and Richard Jin for outstanding research assistance. Card and Rothstein are grateful to Open Philanthropy for research support. Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau. The Census Bureau has ensured appropriate access and use of confidential data and has reviewed these results for disclosure avoidance protection (Project 6000266: CBDRB-FY22-422, CBDRB-FY23-0342, CBDRB-FY23-0425).

I. Introduction

Wages of workers with similar characteristics vary across industries. 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.[1] More controversially, they also argued that the industry premiums from a cross-sectional regression model were quite similar to those estimated from between-industry job movers, once allowance was made for misclassification errors in industry. Since comparisons of job movers hold constant worker characteristics, their analysis suggested that the industry wage effects from a cross-sectional model represent causal pay premiums, rather than differences in unmeasured worker skills. KS took the existence of such premiums as evidence against a perfectly competitive labor market model, and in favor of efficiency wage models in which employers set wages to incentivize worker effort or reduce turnover.

Despite their careful empirical 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 relatively small samples and included only 7 industries.[2] 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 wage changes for industry movers (e.g., Gibbons et al., 2005).

[1] These patterns are 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 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 Germany, France and Britain, respectively, also found larger roles for unobserved ability.

In this paper, we revisit the estimation of industry wage premiums, 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 of Abowd, Kramarz and Margolis (1999; hereafter, *AKM*), confirming the existence of *workplace-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 "ground-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 find significant wage premiums for working in different industries. One-third of the variation in establishment-level pay premiums is between rather than within detailed industries (using 4-digit NAICS industries). The standard deviation of industry premiums is 0.122.[5] The gap between the highest-premium industry (coal mining) and the lowest-premium industry (drinking places) is 88 log points. There is substantial variation both across and within broadly-defined sectors. Across the 20 two-digit sectors, the standard deviation in average industry premiums is 0.10. Within the manufacturing sector, the standard deviation across the 86 4-digit industries is 0.102, with a gap of 54 log points between the top (iron and steel mills) and bottom (other leather products) industries; variability is similar across detailed industries in other sectors.

We use these estimates to address two sets of questions. First, how do the economy-wide industry pay premiums from our ground-up AKM approach differ from the estimated premiums from a simple OLS regression, or from a two-way fixed effects model with person and industry (but not firm) effects, akin to KS's industry mover design? How large are the biases from

---

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

[5] This almost exactly matches the standard deviation of 0.121 obtained for West Germany in the 2002-2009 period by Card, Heining, and Kline (2013; Table VI) using a similar ground-up approach.

unmeasured worker skills in industry premiums derived from a cross sectional model? Can we ignore heterogeneity in pay premiums *within* industries when focusing on wage changes for industry movers? Second, how do industry pay premiums and the degree of sorting of high skilled workers to different industries vary across local labor markets? Are the premiums the same in different cities, or do they vary systematically with characteristics of the local market? Is the degree of skill sorting linked to the dispersion in local pay premiums? Differences in industry pay premiums and the degree of sorting have strong implications for local wage inequality. Their patterns also shed light on models of pay determination.

Starting from a national perspective, our first key finding is that estimated industry effects from a richly specified cross-sectional model ***overstate*** the magnitude of the premiums from an AKM approach by about 20%.[6] The source of this bias is systematic skill sorting across industries that is highly correlated with the actual wage premiums paid in an industry (Haltiwanger, Hyatt, and Spletzer 2022a). We find that a regression of the mean person effects from our AKM model on the industry wage premiums across 4-digit industries has a slope of 0.90, implying strong assortative skill matching across sectors. Only about 70% of the cross-industry variance in worker skills from an AKM model is explained by an index of observed skill characteristics, leading to an upward bias in the dispersion in pay premiums from a cross-sectional approach.

Our second key finding is that the estimated industry premiums from a model with person and industry effects (i.e., a KS-style "industry mover" design) are significantly ***attenuated*** relative to the effects from a ground-up AKM approach. This source of this bias, which has been under-appreciated in the literature, is unobserved heterogeneity in the wage premiums paid by different firms in the same industry. 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 "hierarchy effect."[7]

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

Empirically, we find that when workers move across industries the 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. Symmetrically, 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. This pattern arises naturally in a model of on-the-job search where workers care about the level of wages but not the specific source of any premium. It means that the change in the hierarchy component for industry movers is negatively correlated with the change in the industry premium, causing an attenuation bias in the estimated industry effects in an industry mover design.

The magnitude of the bias is large. We find that the standard deviation of industry premiums is about 50% larger using the ground-up approach than is indicated by estimates based on between-industry movers that ignore within-industry heterogeneity in pay premiums.

Turning to a local labor markets perspective, we study the 50 largest CZs in the country (from Los Angeles to Norfolk, Virginia). We characterize the distribution of industry effects in a given market *c* by the coefficient β_c^1 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] Values of β_c^1 above/below 1 imply that the distribution of industry premiums is widened/compressed in market *c* relative to the national structure. We find modest differences in β_c^1 across CZ's, with a mean of 0.88 and a standard deviation of 0.08.[9] We also find that industry premiums are expanded in labor markets with more high-premium firms, but are only

[8] In a companion paper (Card, Rothstein, and Yi 2023), we study CZ earnings premia, and conclude that the between-industry variation in these CZ premia is 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.

weakly related to the relative supply of skilled workers, or to unionization rates, relative minimum wages, or city size.

We propose two complementary measures of the local degree of skill sorting. The first, β_c^2, 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. The second, β_c^3, relates the mean person effects in an industry and CZ to the national average pay premium for that industry. Higher values of these coefficients mean that industries with more skilled workers (or higher wage premiums) at the national level have even more skilled workers in a given CZ. Again, we find modest variation in the degree of skill sorting across CZs, with standard deviations of β_c^2 and β_c^3 of 0.10 and 0.13, respectively. The two sorting measures are very highly correlated with each other, but only weakly correlated with β_c^1. The degree of local skill sorting is higher in CZ's with more employment in high-premium and high-skill industries, and in CZ's with more high-skilled workers. It is also positively related to the local minimum wage rate, and to the size of the local labor market (consistent with Dauth et al., 2022).

Our findings contribute to the longstanding debate over whether the industry wage effects from a cross-sectional model reflect causal pay premiums or unobserved ability differences. In this debate, we come down part way between the position of KS, who argued that they are mostly causal, and that of Murphy and Topel (1990), who argued that they were mainly driven by unobserved worker abilities. In our data the industry wage premiums from a rich cross sectional model overstate the true premiums by about 20%. But the industry effects from an industry mover design – which both KS and Murphy and Topel (1990) assumed were a benchmark for the ground truth – understate the true premiums by about 50%.

The existence and magnitude of systematic *industry-wide* pay premiums mean that the establishment-level premiums documented in many recent AKM-related papers cannot be fully rationalized by firm-specific factors such as local labor market power. We also find that the wage premiums for different manufacturing industries are *positively* correlated with the markdowns of wages relative to marginal products estimated by Yeh, Macaluso, and Hershbein (2022). This suggests that industry wage premiums are not a consequence of industry-wide differences in labor market power, though there may be other latent factors (such as industry profitability) that drive both wage premiums and estimated markdowns.

The presence of systematic industry premiums lends support to a continuing focus on differences between industries in the analysis of policies related to trade (e.g., Harrison and Rodriguez-Clare, 2010), worker retraining (e.g., Katz et al., 2022), and regional disparities (e.g., Moretti, 2012). It may also support industrial policies to promote the growth of high-premium industries, though for this more work is needed on the *sources* of variation in industry premiums.

Methodologically, we 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. We conjecture that similar biases could arise in other contexts that use a relatively coarse aggregation of units. For example, in an analysis of place effects on children's outcomes (e.g., Chetty and Hendren 2018a,b), there may be unmeasured neighborhood effects (Chetty et al. 2020) that vary systematically for movers between cities. If families tend to move between neighborhoods with relatively similar outcomes, this would attenuate the estimated effects of city-wide factors.

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 bigger 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 between higher skilled workers and higher premium industries. Interestingly, the dispersion in pay premiums and the degree of skill sorting are, if anything, higher in CZs with greater unionization and higher minimum wages. These factors do not appear to moderate the *between-industry component* of wage inequality, though they may still affect within-industry inequality.

II. Industry pay premiums

In this section we present our ground-up framework for estimating industry pay premiums. We formalize the differences between these premiums and those derived from 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 of the form:

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

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 in quarter *t* (indicated by the index function $f(i,t)$) and ε_{it} is a residual term. (We use the terms workplace, firm, and establishment interchangeably; in our empirical analysis, *f* identifies establishments.) This residual incorporates three distinct components: (i) any persistent match effect between the worker and the employer; (ii) any person-specific transitory factors that cause earnings to vary over time on the same job, such as health shocks or family disruptions; and (iii) any establishment-wide transitory factors that cause earnings to vary over time, such as product demand shocks that lead to changes in hours of work.

In our empirical analysis, we start by estimating models like equation (1) by OLS, separately for each of a set of large CZs.[10] 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 in the employment history of worker *i.* In Card et al. (2023) 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 in our setting. We present versions of several of these tests here, focusing on implications for movers across industries. We conclude that although exogenous mobility can be rejected in our LEHD samples, the combination of additively separable worker and firm effects and 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 natural definition of the wage premium for industry *j* is 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).[11] 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).

[10] This approach means that we are identifying the establishment effects in (1) from within-market, between-establishment 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.

[11] We refer to "wage" premiums and "pay" premiums interchangeably. Our data include quarterly earnings but not hours, so we cannot distinguish components coming from hours vs. hourly wages. Our estimation sample includes several restrictions intended to minimize the number of part-time workers included.

We interpret ψ_j in equation (2) as our preferred definition of the wage premium for industry *j*. If a worker 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 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***. We can rearrange (1) as

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

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 "hierarchy component" of the residual $\tilde{\varepsilon}_{it}$ in (3).

Note that a two-way fixed effects model with person and industry effects can be interpreted as an "industry movers" design, as in Krueger and Summers (1988). In such a model, identification is based on wage changes for people who move between industries. Importantly, however, the average wage changes for industry movers do not necessarily identify the industry wage premia defined in (2), even if the assumption of exogenous mobility across firms is satisfied. The problem is that industry movers may be non-randomly selected with respect to the hierarchy components of their origin or destination firms.

For example, consider a simple model of on-the-job search, in which workers are more likely to accept a job offer that comes from a firm that pays a higher wage premium than their current firm.[12] In that case, a worker in a higher-premium industry who obtains an offer at a firm in a lower-premium industry will tend to accept that offer only when the change in the hierarchy terms between the incumbent and prospective firms is large enough to outweigh the decline in

[12] This is precisely the policy for job searchers in the model developed by Burdett and Mortensen (1998) -- see Manning (2003).

the average industry premium. By contrast, a worker in a low-premium industry considering an offer from a higher-premium industry will be much less selective about the hierarchy term of the new offer (unless it is so much worse than the current job's hierarchy term that the total wage change is positive). This process induces a negative correlation between the change in the industry effect and the change in the hierarchy term among industry movers, leading to an attenuation bias in the estimated industry effects from an industry mover design.[13]

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 examine the changes in the hierarchy term for workers moving across industries. As expected, we find that changes in the hierarchy term are strongly negatively correlated with the changes in estimated industry premia.[14]

We note that the error component in earnings created by a hierarchy effect has many of the same properties as an industry-specific match effect in worker productivity. Specifically, Gibbons et al. (2005) show that the change in industry-specific match effects for industry movers will tend to be negatively correlated with the change in mean wage premiums, just like the change in hierarchy effects. Hierarchy effects, however, are shared by all the workers at a given firm, rather than being worker-specific, so they are directly observable.

D. Cross-sectional industry effects model

Next, consider an approach based on a cross-sectional regression model that controls for the observed skills of worker *i.* Starting from equation (3), consider the projection of the person

---

[13] A little more formally, if we assume that accepted job offers have $\Delta\delta_f = \Delta\psi_{j(f)} + \Delta h_f > 0$, then the conditional covariance between $\Delta\psi_{j(f)}$ and Δh_f is more negative than the unconditional covariance: $cov(\Delta\psi_{j(f)}, \Delta h_f \mid \Delta\delta_f > 0) < cov(\Delta\psi_{j(f)}, \Delta h_f)$, creating omitted variables bias in a first-differenced version of (3). Note that if the firm effects δ_f reflect in part compensating differentials for non-wage job amenities, workers may voluntarily switch to jobs with lower δ_f. This would reduce the attenuation, potentially to zero if compensating differentials are the sole source of variation in firm premiums. Thus, evidence that industry movers estimates are attenuated can be taken as a demonstration that firm and industry premiums do not consist solely of compensating differentials.

[14] We also find a negative correlation between the change in the industry effect and the change in the worker-firm match effect (the component of ε_{it} that is stable within an employment spell). This has a similar interpretation in terms of selection from the set of possible job offers, but it is small in magnitude and does much less to offset changes in industry premiums.

effect α_i on a worker's observed skills (S_i), their *X's* (age and year effects), and their observed industry:

$$\alpha_i = S_i\lambda_\alpha + \pi_{j(i,t)} + X_{it}\theta_\alpha + v_{it\,.} \tag{4}$$

In this equation, $\pi_{j(f(i,t))}$ represents the mean of the person effects for workers in *i's* industry, controlling for her observed characteristics (S_i and X_{it}). Loosely speaking, it represents "mean unobserved ability" in the industry. Given evidence in the recent literature of strong positive assortative matching 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.

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

$$h_{f(i,t)} = S_i\lambda_h + \rho_{j(f(i,t))} + X_{it}\theta_h + u_{f(i,t)}. \tag{5}$$

Note that even though the mean of the hierarchy term across all person-quarter observations in an industry is 0, the mean conditional on worker characteristics is not necessarily 0. Indeed, holding constant observed worker skills we expect the mean hierarchy term 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, an industry effect $\kappa_{j(i,t)}$, and a residual term:

$$y_{it} = S_i(\lambda_\alpha + \lambda_h) + \kappa_{j(i,t)} + X_{it}(\theta + \theta_\alpha + \theta_h) + \eta_{it} \tag{6}$$

where

$$\kappa_{j(i,t)} = \psi_{j(f(i,t))} + \pi_{j(i,t)} + \rho_{j(f(i,t))}. \tag{7}$$

Equation (7) shows that 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 (represented by $\pi_{j(i,t)}$) which presumably lead to an upward bias in the estimated premiums for higher-paying industries, and unmeasured "firm quality" within the industry (represented by $\rho_{j(f(i,t))}$), which presumably works in the opposite direction. We investigate the net effect of these biases by comparing the estimated industry wage effects from a richly specified cross sectional model (estimated on data from the American Community Survey, ACS), to the wage effects from our ground-up approach.

E. Geographic variation in industry premiums and worker sorting

Krueger and Summers' (1988) paper spawned a large literature studying the structure of industry wage premiums in different countries (see Rycx and Tojerow, 2007, for a survey). While international comparisons of industry premiums can highlight some of the factors responsible for these premiums (see, e.g., Tuelings and Hartog, 1998) there are at least two problems 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 may vary across settings.

Our large national data base provides a novel opportunity to explore differences in the distributions of industry pay premiums across local markets, while abstracting from measurement problems and biases due to unmeasured skills or firm quality. 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 in that market (weighting by their quarters 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 are related to market-level characteristics.

Four-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. Moreover, Card, Rothstein, and Yi (2023) note that the $\psi_{jc}{}'$s from different CZ's are highly correlated, leaving relatively little variation in industry premiums across places. To extract some signal in the CZ-specific data we focus on 1-parameter models that measure how a CZ's industry premiums and/or degree of skill sorting compare with patterns in the nation as a whole. First, we estimate a (weighted) regression of the industry premiums in CZ *c* on the corresponding *national* premiums:

$$\psi_{jc} = \mu_c^1 + \psi_j \beta_c^1 + \xi_{jc}^1, \tag{8}$$

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 β_c^1. Analogously to the beta coefficient in a CAPM model, $\beta_c^1 < 1$ means that the industry premiums in CZ *c* are compressed relative to the national structure, while $\beta_c^1 > 1$ means they are expanded.[15]

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 $\overline{\alpha}_j$:

$$\overline{\alpha}_{jc} = \mu_c^2 + \overline{\alpha}_j \beta_c^2 + \xi_{jc}^2 \, . \tag{9}$$

A CZ has more (less) skill sorting across industries than in the national labor market as a whole if β_c^2 is greater than (less than) 1.

We also construct a third measure, capturing whether a CZ's workers are more or less 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:

$$\overline{\alpha}_{jc} = \mu_c^3 + \psi_j \beta_c^3 + \xi_{jc}^3 \tag{10}$$

[15] β_c^1 is unaffected by the choice of normalization of the market-specific AKM models. However, our estimate of the national industry premium incorporates the premiums for CZ c, so there is a potential correlation between the right and left hand sides of our estimating equation that biases the estimate of β_c^1 upward. This bias is on the order of the share of employment in CZ *c*, so we believe it can be safely ignored.

The coefficient β_c^3 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. This can be compared to the national version of this same model:

$$\overline{\alpha}_j = \mu^3 + \psi_j \beta^3 + \xi_j, \tag{11}$$

which, as we show below, has $\hat{\beta}^3 \approx 0.9$. A city with $\beta_c^3 > \beta^3$ has more skill sorting than in the nation as a whole, while a city with $\beta_c^3 < \beta^3$ has less.[16]

The coefficients β_c^1, β_c^2, and β_c^3 capture different aspects of the industry wage structure in a CZ. We conduct a descriptive analysis of their relationships to each other and to characteristics of the CZ, including the relative share of employment in higher-premium industries, the relative supply of higher-skilled workers, the size of the local market, the local rate of union coverage, and the relative level of the minimum wage.

III. Data

Our analysis relies on data from the Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program. These data are derived from quarterly earnings reports provided by employers to state unemployment insurance (UI) agencies. The core data set includes total wages paid by a given employer to each worker in a quarter. This is supplemented with information on employers and workers derived from other sources (e.g., the decennial census and ACS files) – 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

[16] The β_c^3 coefficient is conceptually similar to ones estimated by Dauth et al. (2022). It is also closely related to our first measure of skill sorting, β_c^2. Substituting (11) into (9), $\beta_c^3 = \beta^3 \beta_c^2 + \tau_c$, where τ_c is the slope coefficient from a regression of the error term in equation (9) on ψ_j, the national industry wage premium. Since β^3 is close to 1, β_c^3 is (roughly speaking) a "noisy" version of β_c^2.

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 all *transitional* quarters (the first or last quarter of any person-employer spell). We also drop PEQs with an unknown industry and/or establishment location. Finally, we drop individuals with fewer than 8 quarters of earnings that satisfy the previous restrictions over our 8½ year sample window.

The UI data in the LEHD contain an identifier for the employing firm and the state, but not for the specific establishment of firms with multiple establishments in the same state. The Census Bureau uses data on workers' residential addresses and the locations of establishments owned by the firm to impute establishments for workers employed at multi-establishment firms (Vilhuber 2018). We use the first of the multiple imputations available to assign PEQs to establishments. Industry premiums that are estimated by averaging establishment premiums for single-establishment firms (which are not subject to any type of imputation) are very similar to those that use the data for all establishments – see the analyses in Card, Rothstein, and Yi (2023).

We classify establishments by 4-digit NAICS codes, with a total of 311 distinct values.[18] We also use the establishment location to assign workers to commuting zones. We focus on the 50 largest CZs, which together account for over half of the national population. We group the

[17] Because some of the education measures predate the beginning of our sample and we use only ACS interviews when people were at least 30 years old, the education sample over-represents older individuals.

[18] Industry codes are imputed to establishments in the LEHD using the procedures described in Vilhuber and McKinney (2014). In some analyses we compare the LEHD to data from the American Community Survey (ACS). Here, we either use the 262 Census Industry Codes (CIC's) that are available in the ACS, which combine some small industries, or we link these ACS aggregates to their constituent LEHD industries. One LEHD industry (NAICS 5211, Central Banks) cannot be mapped to the ACS, and is omitted from this linkage.

remaining CZs into approximately 10 regional composites.[19] For simplicity, in the remainder of the paper we refer to the 50 larger CZ's and the regional composites as "CZ's".

We estimate our worker-firm AKM model separately by CZ. In practice, this means that the coefficients on time-varying individual covariates (the θs in equation 1) are allowed to vary by CZ, and the individual effect is actually an individual-by-CZ effect. As a result, the 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.[20] 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 non-tradeable sector (2.7% of national employment) that is present in all CZ's. It is also one of the lowest-paid industries, with jobs that are open to a wide variety of workers. This makes it a useful benchmark for other sectors.

In some analyses below we focus on patterns of wage changes around industry switches, for workers who remain in the same CZ.[21] For these event studies we limit attention to workers who switch industries once while in a given CZ, have a stable job at the same firm for at least 5 consecutive quarters before the switch and another stable job (at a different establishment) for at least five consecutive quarters afterwards. 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.[22]

Table 1 shows some characteristics of our various 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

[19] 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 composite CZs includes all workers and firms in New England that are not in those four CZs. For disclosure reasons we cannot precisely delineate the composite regions.
[20] Approximately 60% of PEQs in the connected set come from workers observed to switch firms at least once within the sample.
[21] In Card, Rothstein and Yi (2023) we present an analysis of wage changes around moves between CZ's.
[22] 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.

that never switches industries (column 2) and another that switches industries at least once within our sample window (column 3). The two subgroups 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, reflecting the restriction to workers with stable employment histories before and after their sole industry switch.

IV. Descriptive overview of industry pay differences

Before turning to causal estimates of industry pay differentials, we begin by presenting some 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 of the 262 ACS Census Industry Codes (CIC's).

Figure 1A shows how mean log hourly wages in an industry are related to mean education in that sector. The first thing to note here is that there is wide dispersion in each variable: The standard deviation of industry wages is 0.31, while the standard deviation of mean education is 1.35. Second, average wages and average education are strongly correlated (ρ = 0.69); each additional year of mean education is associated with mean wages that are 15.7% higher. This is 5-6 percentage points higher than a Mincerian estimate of the return to education in the ACS data (10% per year), suggesting that wage differences across industries reflect both education differences across workers and other ecological factors that vary across sectors.

To probe this, we estimate a regression of log individual wages on industry fixed effects, controlling for just over 200 individual characteristics (including years of education, gender, race/ethnicity, country of origin, and field of study), as well as indicators for the major CZs and residual regions of the country.[23] This specification is similar in spirit to the cross-sectional

[23] 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;

estimates in KS, though richer, taking advantage of variables such as field of study that were not available to KS. Figure 1B replaces unadjusted mean industry wages with the resulting estimated industry effects (which we normalize to mean zero). The standard deviation of the estimated industry effects is much lower, 0.180, and the slope is also much reduced, to 0.057. It remains highly significant, confirming that industries with more educated workers pay more, even controlling for the returns to the individual education of their workers.

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* skill differences among workers in different industries, and the other reflecting hierarchy effects. We next turn to the LEHD data to explore models that adjust for these factors.

V. National industry differentials

In this section we present the main results for our national analysis of industry pay differentials; robustness tests and evidence on the validity of the specification are presented in the next section.

a. Summary of AKM Model

We begin by estimating the AKM model (1). The model includes fixed effects for workers and employers (establishments) as well as fixed effects for calendar quarters 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; as noted above, we set the mean of $\hat{\delta}_f$ across all firms in the restaurant industry (NAICS code 7225) in each CZ to zero.

Appendix Table A1 presents some summary statistics on the estimated components of the AKM

---

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.

model and their variance shares.[24] Consistent with existing results for Washington State (Lachowska et at., 2023) and for many other countries (e.g., Germany, Portugal, and Brazil) we find that person effects account for the majority of the variance in earnings in our LEHD sample (roughly 68%) while establishment effects account for 11% and the covariance between worker and establishment effects accounts for 7%. The standard deviation of estimated worker-establishment "match effects" (the mean residual within a single match) is 0.089. A regression of the average person effects for the employees at a workplace ($\bar{\alpha}_f$) on the estimated pay premium at the establishment has a coefficient of 0.39, implying a significant degree of positive assortative matching between high wage workers and high premium firms.[25] If we instead regress establishment mean person effects on the establishment's estimated hierarchy effect, the coefficient is 0.14.

A key question from the perspective of industry-wide pay structures is what fraction of estimated pay premiums and average person effects at different establishments are explained by industry affiliation. The answer, based on the adjusted R-squared statistics from regressions of $\hat{\delta}_f$ and $\bar{\alpha}_f$ on 4-digit industry dummies, is 33% and 12%, respectively. Thus, industry affiliation is an important determinant of both establishment pay premiums and the average quality of workers, but there is a good deal of variation within industry in both dimensions.

b. Two- and Three-Digit Industry Averages

For most of our remaining analysis we aggregate the estimated firm effects δ_f from model (1) to the 4-digit industry level. Before turning to that analysis, however, we present a brief overview

---

[24] We compute variances and covariances of the components separately for each of the approximately 60 CZs on which we estimated separate AKM models. The upper panel of the table reports weighted means of these CZ-level variances and covariances; other statistics are computed from these means. For comparison, the lower panel of the table reproduces corresponding estimates for a single national model from Card, Rothstein, and Yi (2023). The variance of establishment effects is slightly larger here, as it includes between-CZ variation. The variance decomposition is otherwise very similar. If we regress CZ-by-CZ estimates of industry effects used here on the national estimates from Card et al. (2023), the slope is 0.96.

[25] This regression is estimated on a sample that stacks observations from all CZs. When we estimate this regression from the averages of CZ-level variance decompositions in Table A-1, thus weighting CZs by their sizes rather than by the variance of establishment premiums, the weighted mean coefficient is 0.33. Last, when we use the national estimates from Card, Rothstein, and Yi (2023) reported in the lower panel of Table A-1, thus including between-CZ as well as within-CZ variation, the coefficient is 0.43.

of the industry differentials across broader industry categories.

Table 2 shows a variety of summary statistics for 2-digit industries based on our ground-up approach applied to the LEHD (columns 1-4) as well as parallel results based on ACS data and the cross-sectional model used in Figure 1b (matched to 2-digit NAICS codes). The mean levels of worker skills from the two approaches (columns 2 and 6) are highly correlated across sectors (ρ=0.90) as are the estimated sectoral wage premiums (columns 3 and 7; ρ=0.91), though the standard deviation in premiums from the cross-sectional approach is about 50% higher than from the AKM-based approach (see the bottom row of the table).

Recall that in our AKM approach we adopt the normalization that pay premiums are zero for the restaurant sector. Consistent with the reasoning underlying that choice, the Accommodation and Food Services sector, NAICS 72 (of which restaurants comprise 86% of total employment) has the lowest average earnings across all two-digit industries. It also has both the lowest level of worker skills and the lowest average wage premium, using either our AKM-based approach or the cross-sectional model.[26] The highest-premium sector using either approach is mining and extraction, with utilities second. These industries do not have particularly high person-effect workers, but their industry premiums are enough to make them the highest-paid sectors overall.

Table 3 shows the premiums for 18 3-digit industries within manufacturing – a sector that is widely studied in part because of the availability of data on outputs and inputs. Again, we see substantial variation in wages, workers skills, and wage premiums across subsectors of manufacturing, with a standard deviation of industry effects from our AKM-based approach of 0.09, and of 0.10 from a cross-sectional model. We also see, as in Table 2, that measures of worker skills from the two approaches are highly correlated across sectors (ρ=0.98), as are the estimated average wage premiums (ρ=0.92). The lowest-premium manufacturing industry is Apparel, while the highest is Petroleum and Coal products; the difference between these is 47 log points using the AKM-based premiums, and 54 log points using the cross-sectional

[26] We normalize the pay premiums from our cross-sectional approach to have the same mean in NAICS 72 as in our AKM-based approach.

premiums.

A natural question is what industry characteristics correlate with the industry wage premiums that we identify. Casual inspection of Tables 2 and 3 points to a few candidates: A portion of the industry wage differentials may reflect compensating differentials for the working conditions in mining and other extractive industries. Even in these sectors, however, some of the wage advantage may also be due to union coverage or other factors.[27] Premiums also seem to be higher in industries with large amounts of capital per worker, perhaps reflecting rent-sharing or efficiency wage premiums that are paid to prevent shirking when workers have control over complex machinery and equipment.

Another potential contributor to industry wage premiums is variation in employer market power in wage setting. Appendix Figure A1 shows the relationship between the 3-digit premiums in Table 3 and the log of the median industry-specific monopsonistic wage markdowns – defined as the ratio of wages to marginal products, so lower values correspond to steeper markdowns – estimated in a recent paper by Yeh, Macaluso, and Hershbein (2022).[28] Contrary to the hypothesis that greater monopsonistic exploitation *lowers* wages, the estimated industry wage premiums are *negatively* correlated with the estimated markdowns (unweighted correlation = -0.69): Industries that Yeh et al. (2020) estimate to mark wages down further offer *higher* premiums than do those where wages hew closer to marginal products. This suggests that some other underlying factor, like industry profitability, is correlated with both the average wage premium in an industry and the average markdowns estimated by Yeh et al. (2022).

c. Industry Differentials at the 4-Digit Level

Appendix Table A2 presents estimated wage premiums and average worker effects at the 4-digit

[27] See Lavetti and Schmutte (2023) for a recent attempt to measure the size of the compensating wage differentials for safety risk. Even in their setting, with high quality administrative wage data and detailed fatality statistics at the firm level, the size of the compensating differentials is small.

[28] In principle these markdowns represent the average logarithmic gap between the average wage paid to employees and the average value marginal product of labor at a firm. As noted by Bond, Hashemi, Kaplan, and Zoch (2021), however, the interpretation of these estimated markdowns is clouded by the fact that the underlying production functions used by Yeh et al. (2022) are estimated using <u>revenue</u> rather than physical output.

industry level. 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. Only four industries, all relatively small, 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 $\psi_j = -0.01$.[29] At the other end of the scale, the industry with the highest pay differential is coal mining, with $\psi_j = 0.80$.

The weighted standard deviation of ψ_j is 0.12, which is remarkably close to the (sampling error-adjusted) standard deviation of 2-digit industry effects estimated by KS. Figure 2 shows a weighted histogram of the estimated ψ_j coefficients, using different colors for industries in each of 9 major 1-digit industry groups (identified by the first digit in the first column of Table 2). 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 are mainly represented in the middle and upper parts of the distribution, while arts/entertainment/accommodation and education/health are represented in 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 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 by ψ_j, with a pay premium of 0.42.

Figure 3A shows the relationship between industry premiums and average wages. The relationship is very tight, with an R^2 of 0.72 – higher premium industries tend to be higher average wage industries. The slope of average earnings with respect to ψ_j is 1.93, indicating a

[29]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.

strong pattern of assortative matching across industries between high-skilled workers and high-premium industries that magnifies wage inequality. Figure 3B shows this more directly. Here, we plot $\bar{\alpha}_j$ (the mean of the person effects in a sector) against the associated industry premium ψ_j, providing a visualization of equation (11) above. The estimated slope (i.e., the coefficient β^3) is 0.90.

Table 4 presents regression versions of several of these analyses, with industries as the unit of observation.[30] 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 (measured in our LEHD estimation sample). 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.[31]

Column 3 shows a regression of $\bar{\alpha}_j$ on ψ_j, summarizing the scatterplot in Figure 3B, with a slope of 0.90. We note that this slope is considerably above the corresponding slope at the firm level noted above (0.39), reflecting a combination of downward bias in the covariance between firm effects and workers effects that is mitigated by aggregating to the industry level and more sorting of workers across industries than across firms (within industries).[32]

In Columns 4-8 we combine industry-level estimates from the LEHD with parallel estimates from the ACS, using a crosswalk between the 4-digit NAICS codes used in the LEHD and the slightly more aggregated NAICS codes used in the ACS. Column 4 reproduces the model in column 3, omitting the one sector (NAICS 5211, Central Banks) that cannot be matched to the ACS. In Column 5, we regress the average α for workers in each industry from the LEHD on the average

[30] This analysis is related to earlier work by Abowd et al. (2005) and Abowd, Lengermann, and McKinney (2003), who measure differences across industries in the distribution of AKM person effects.
[31] These do not add exactly to 100% because a small portion is due to differences in the time-varying observables in (1). Similarly, the slope in Figure 3A, 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).
[32] As emphasized by Kline et al. (2020), estimation errors in the person and firm effects of the AKM model will lead to a downward bias in the covariance between the estimated firm effects and the person effects of the associated employees and an upward bias in the estimated variance of firm effects. Given the length of our sample period, however, results in two recent studies that compare corrected and uncorrected variance decompositions (Gerard et al., 2021, and Lachowska et at., 2023) suggest that this bias will be modest -- on the order of 10-15%.

of the "skill index" formed from the observed characteristics in a cross sectional model fit to the ACS (the same model used to estimate the industry effects in Figure 1B, but fit to 222 NAICS industries/industry aggregates rather than to 262 CIC's). The slope is very close to 1 (1.04, with a standard error of 0.07), implying that the component of skills in α_i that is <u>not predictable</u> by observed worker characteristics in the ACS is nearly orthogonal to the component represented by observed worker characteristics. Moreover, the R-squared is 71%, suggesting that our cross-sectional model explains a majority of the variability in worker skills across industries (though only about 40% of the variability of wages across workers). Column 6 adds a control for the industry premium ψ_j. Interestingly, the industry premium is significantly related to $\bar{\alpha}_j$ even after controlling for observed skills, implying that there is sorting on the unobserved component of worker skills as well as on the observed component.

In columns 7 and 8 we illustrate this further by decomposing $\bar{\alpha}_j$ into two components – the portion predicted by the index of worker skills from the ACS, and the remainder (i.e., the predicted value and the residual from the model in column 5). We regress each of these separately on ψ_j, and find that both are positively related to the industry premium, though the relationship is stronger for observed skills (R-squared = 0.27) than for unobserved skills (R-squared = 0.09).[33]

VI. Assessing the validity of the decomposition

A causal interpretation of the estimated firm effects from an AKM model estimated by OLS 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}. In a companion paper (Card et al., 2023) we present tests of the validity of the AKM model in LEHD data. We do not repeat those here, but present some evidence focused on the interpretation of the industry

[33] A stylized model that can explain the patterns in Figures 3A and 3B and the estimated coefficients in Table 4, is the following. There is a joint distribution of industry premiums, observed skills $(\bar{X}_j b)$ and unobserved skills $(u_j \equiv \bar{\alpha}_j - \bar{X}_j b)$ across industries, with $var[\psi_j] \approx var[\bar{X}_j b]$, $var[u_j] \approx 0.4\, var[\psi_j]$, $correl[\psi_j, \bar{X}_j b] \approx 0.6$, $correl[\psi_j, u_j] \approx 0.3$, and $correl[\bar{X}_j b, u_j] = 0$. Moreover the industry premiums estimated in a cross sectional model are the sum of the true premiums and unobserved worker skills (with no bias due to unobserved firm premiums). This model implies that 70% of skills are observed and that regressions of the premiums from a cross sectional model on ψ_j *and* $\bar{\alpha}_j$ have coefficients of 0.90 and 0.34, respectively.

averages of the firm effects.

a. An Event Study of Industry Movers

Panel A of Figure 4 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. As discussed above, this analysis is based on workers who were employed at the same firm for at least five quarters, then moved to a new firm in a new industry and were stably employed for the second firm for 5+ quarters. We divide industries into quartiles based on their estimated wage premiums and show eight series, corresponding to moves originating at firms in the top and bottom quartiles. The graph shows mean age-adjusted earnings in "event time" around the first quarter of employment at the new firm.

Three facts are apparent here. First, mean earnings of movers rise or fall based on the change in the estimated premiums for their origin and destination industries. Workers who move from jobs in industries in the lowest quartile of premiums to jobs in the 3rd or 4th quartile (labeled "1-3" and "1-4" in the figure) experience substantial earnings gains, while those who move from industries in the top quartiles of premiums to lower-wage industries (labeled "4-1" and "4-2") experience substantial losses. (We defer a discussion of the precise magnitudes of the earnings changes to the next subsection). By comparison, workers who move between jobs within the same quartile of premiums ("1-1" and "4-4" moves) experience relatively small average changes in earnings.

Second, most of the earnings changes happen at the move. There is no evidence of differential pre-trends for movers who end up moving up or down the industry ladder. However, there is some evidence of an adjustment process after a move: all of the lines slope up 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 (2023), where we find that movers to a new CZ take 2-3 quarters to adjust to their new earnings level.

Third, we see that a worker's initial wage level varies not just with the origin industry, 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 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. These patterns reflect assortative matching: 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 $\bar{\alpha}_j$ and $\psi_{j(f(i,t))}$ seen in Figure 3B.

The large earnings changes for some of the groups in panel A necessitate a zoomed-out scale, making it difficult to discern smaller differences across groups. In Panel B, we repeat the exercise but plot the mean of ε_{it}, the error term in the AKM decomposition, by event time quarter and mover-group. Ideally, there would be no systematic variation in mean residuals before or after a move. 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 is especially pronounced for movers from low-wage to high-wage industries, suggesting that these workers do not immediately see the full earnings gains associated with their new industry premium – and indeed even several quarters after the move their earnings residuals are somewhat lower than they were prior to the move. As we discuss below, this reflects a small but systematic change in average worker-firm match effects for between-industry movers.

Finally, Panel C of the Figure shows an event study of the mean hierarchy effects, $\hat{h}_{f(i,t)}$, for different mover groups. Recall that the hierarchy effect is the difference between the premium offered by the firm where the worker works and the mean average premium in the industry. Since our event study sample is limited to workers who were stably employed at their origin firm for five quarters before their move, and remain at their destination firm for five quarters after their move, the mean hierarchy effect is stable both before and after a move. However, it changes substantially at the move, with a clear tendency to rise when a worker is moving to a lower average premium industry and fall when a worker is moving to a higher average premium industry. The magnitudes of the changes are large: for example +14 log points for workers who move from 4th quartile industries to 1st quartile industries. As discussed in Section II.C, these changes create an omitted variable bias in an industry movers analysis that does not account for firm effects, leading to a substantial attenuation bias in the estimated industry premiums from such a design.

b. A more detailed analysis of earnings changes for industry movers

Figure 5 provides a more in-depth look at the components of wage changes around a switch in industries. We divide industries into 20 vingtiles based on their average pay premiums, and construct 400 cells corresponding to moves between origin and destination firms in each vingtile. We construct the mean changes in age-adjusted wages and in the components of wages 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 5 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. The slope is 0.49, implying that movers get about one-half of the change suggested by the change in industry premiums between their origin and destination industries.

Following equation (3), the change in the mean age-adjusted wage for a mover from vingtile v to v' consists of three components: (1) the difference in the mean of the industry pay premiums between vingtiles v and v'; (2) the change in the average hierarchy term (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, which includes as one subcomponent the change in worker-firm match effects. Panel B of Figure 5 shows the changes in the AKM residuals for the movers in each origin-destination pair. Consistent with the pattern in Figure 4B, we see that the residuals for movers up the industry premium ladder tend to be more negative, with a slope of -0.12. This is evidence against the assumption of exogenous mobility. If we extend the post-move observation to the 4th quarter after the move, after the recovery from the initial dip seen in Figure 4, the slope is somewhat smaller (-0.093). This remaining component derives primarily from the worker-firm match effect – when we construct a similar slope for the change in the average of ε_{it} within a job spell, it is -0.084.

Panel C shows the change in the hierarchy effects $h_{f(i,t)}$. Here we see a much more substantial downward slope, as expected from the pattern in Panel C of Figure 4. Workers who move to higher-premium industries systematically lose hierarchy effects, while those who move to lower-premium industries see increases in the hierarchy term. The slope is -0.39, indicating that changes in hierarchy effects offset a large fraction of the change in industry premiums for the average industry switcher, and account for most of the flattening of the slope of the scatter of points in panel A relative to the benchmark slope of 1. This pattern of hierarchy effects means that an industry mover design will yield substantially attenuated industry effects relative to the effects defined from a ground-up approach. Interestingly, we obtain extremely similar estimates of the slope of hierarchy effects when we re-create this analysis using moves between 2- or 3-digit industries as with the 4-digit industry moves used here.

As noted earlier, the hierarchy effects that we see in Panel C of Figure 5 are easily understood in the framework of a model where workers search for higher wages via a combination of moves up the industry ladder (to sectors with higher values of ψ_j), up the firm ladder within industries (to establishments with higher values of h_f). A similar logic suggests that workers may also trade off match effects, systematic deviations of a worker's pay at a job from the sum of additive worker and firm effects, against industry premiums. Indeed, we find that the negative slope in Panel B largely reflects changes in match effects – the slope of the change in the average of ε_{it} within a job spell with respect to the change in industry premiums is -0.084.

Panel D shows the change in earnings net of the hierarchy effect (that is, in $y_{it} - X_{it}\hat{\theta} - \hat{h}_{f(i,t)}$).

This is relatively close to the 45-degree line, with a slope of 0.88, meaning that the estimated industry effects do a good but not perfect job of predicting earnings changes for between-industry movers once the change in hierarchy is accounted for. Again, we note that if the post-move observation is shifted to 4 quarters after the move, the slope of the points in panel D would be a bit closer to 1 (0.91).

Our simple search model suggests that workers will tend to move up the *h* distribution as they gain experience. We explore this in Table 5. Here, we regress the hierarchy effect for worker *i* in quarter *t* on a quadratic in the number of quarters of experience that the worker has in their *current* industry. We estimate this separately for younger workers (aged 26 or less at the beginning of 2010) and 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 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 marginal returns to industry experience that are larger for younger than older workers. The estimated quadratics are concave and suggest that the within-industry experience effect peaks at around 18 – 20 quarters. 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. 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

How do the estimated industry wage premiums from our ground-up approach compare to those from a cross sectional model or an industry mover design? Figure 6 plots the estimated industry effects derived from our cross-sectional model (fit to the ACS) against the ground-up industry effects from the LEHD. (We use the same approach as in Table 4 to link industry premiums from the ACS to the NAICS industries in the LEHD). The analysis in Table 4 suggests that observed skill characteristics in the ACS capture about 70% of the skill differences summarized by $\bar{\alpha}_j$, but that the remaining unobserved skill characteristics tend to be higher in high-premium industries (column 8 of Table 2). This causes the scatter of points in Figure 6 to have (weighted) slope of 1.17 (standard error = 0.05). The implication is that a cross-sectional approach overstates the variation in industry effects by a factor of about 20%.[34]

Figure 7 plots the estimated premiums from a model fit to the LEHD with person and industry effects – i.e., an "industry mover design" – against the corresponding ground-up estimates. The two sets of effects are very highly correlated (ρ=0.96), but as expected given the pattern of changes in hierarchy effects for industry movers in panel C of Figure 5, the estimated slope coefficient is just 0.62. Surprisingly, perhaps, the attenuation bias in the distribution of industry effects from an industry mover design is twice as large in magnitude as the unobserved ability bias in the industry effects from a cross sectional model.

Table 6 compares the estimates of ψ_j from alternative cross-sectional and industry movers specifications to those from our preferred ground-up specification. In each column, we show the coefficient when we regress the alternative industry effects on our preferred estimates, weighting by the number of person-quarter observations in the industry in our LEHD estimation sample. 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 column 1).

In columns 2-4 we consider alternative cross-sectional models fit to the ACS. The specification in column 2 represents a "bare bones" human capital model, with a linear control for education, a quartic in potential experience, and dummies for 4 race/ethnicity groups, all fully interacted

[34] The (weighted) standard deviation of the estimated industry effects from a cross sectional approach is also larger than the (weighed) standard deviation of the ground-up estimates (0.17 versus 0.12) though we have made no attempt to correct these for estimation error. Our ACS sample has 11.6 million observations and for all but a few smaller industries the estimated industry effects are quite precise.

with gender. The specification in column 3 generalizes this by adding controls for field of study, dummies for education levels, and a variety of immigrant characteristics (again, all interacted with female). Finally, the specification in column 4 represents our benchmark ACS model (used in Figures 1B and 6 and the models in Table 2), which extends the specification in column 3 by adding dummies for CZ of residence. A comparison between columns 2, 3 and 4 shows that the standard deviations of the estimated industry effects are somewhat larger for the two simpler specifications (0.189 and 0.174 versus 0.169 for the benchmark cross sectional model); the regression coefficient of the estimated $\psi_j{}'s$ from the bare bones model on the ground-up $\psi_j's$ is also somewhat larger (1.24 versus 1.17), suggesting that the additional information in the field of study and immigrant variables reduces the bias from unobserved skill characteristics.

Columns 5-7 consider industry mover specifications like equation (3), estimated on the LEHD with person and industry fixed effects but no firm controls. 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. While the inclusion of person fixed effects removes unobserved ability bias, the industry effects from model (3) are subject to bias from omitted hierarchy effects, and consistently under-estimate the dispersion in industry pay premiums, with standard deviations about two-thirds as large as those from our preferred specification. We conclude that failure to account for the within-industry variation in firm pay premiums substantially attenuates the 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 construct separate average 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 that are specific to each education group. This yields separate industry premiums that reflect differences in the relative shares of the two education groups at higher- versus lower-premium firms within the same industry.

If the firm distributions of the two education groups are the same, the two premiums will be identical.[35] 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.

We note that in principle we could estimate separate AKM models for highly educated and less educated workers, and use these to construct an alternative set of industry effects for the two groups. The problem in our setting is that education is only observed for a relatively small share of workers in the LEHD (around 15%), leading to concerns about estimation errors. Evidence on the patterns of estimated pay premiums for different education groups in Portugal (Card, Cardoso, Heining and Kline, 2018, Figure 8) suggests that wage premiums are about the same for more and less educated workers at the same workplace. Relatedly, studies have found that workplace pay premiums for male and female workers are very similar in Germany, Italy, France, Portugal, and Brazil (Palladino, Roulet and Stabile, 2022). Given this evidence, and the limitations of the education information in the LEHD, we defer the estimation of separate AKM models by education to future work.[36]

Figure 8 explores the relationship between education-group-specific weighted averages of the establishment premiums from our AKM models, and the pooled averages described by equation

[35] Saez, Schoefer, and Seim (2019) find that a worker-specific tax cut spills over to the wages of ineligible workers at firms employing large shares of eligible workers, suggesting that rent sharing or other motives lead firms to adopt similar premiums across worker groups even when they are not similarly productive.

[36] Card et al. (2023) attempt to assess the validity of the constant-premium assumption by measuring across-firm variation in the return to education, net of the AKM person effects – that is, the variability of the firm-by-education match effect. To do this, we take residuals from the AKM model for those workers for whom we observe education, and regress them first on firm indicators and then on firm-by-education indicators. The adjusted R2 goes *down* in the latter specification, offering little evidence of substantial between-firm variation in the return to education.

(2). The group-specific industry effects are extremely highly correlated with the pooled average, with ρ = 0.98 for non-college workers and ρ = 0.99 for college workers. Evidently, despite the differences in the matching of lower and higher-education workers to workplaces within an industry, the average wage premiums in each industry are quite similar for the two groups, though the slopes in Figure 8 indicate slightly larger (more variable) 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 9 shows how the mean person effects for college and non-college workers are related to the pooled mean of the person effects in the same industry. Not surprisingly, the mean of the person effects is higher for college than for non-college workers, both overall and within industries. But 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 (2023). 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. We note in that paper that this differential sorting makes it appear as if the return to education is much higher in high-premium CZs, where the AKM decomposition indicates that this reflects a constant return to education combined with higher unobserved skill of college workers in those CZs. The same is true for industries: The observational return to education is higher in high-premium industries, on average, due to differential selection of high- and low-ability college graduates into those industries. This may reflect differences in the production technologies of the different industries and CZs, though Figure 8 indicates that any effects of these technologies on wages are shared equally by more- and less-educated workers.

VII. Local industry differentials

A number of studies inspired by KS have tried to use differences in the patterns of industry premiums across countries or over time to say something about the <u>explanation</u> for these differences (Rycx and Tojerow, 2007). Our setting provides a novel opportunity to pursue this agenda, focusing on differences in pay premiums and worker sorting across major CZs in the US.

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) $\beta_c^1 \equiv \frac{d\psi_{jc}}{d\psi_j}$, the slope of the relation between ψ_{jc} and ψ_j (the national industry premium);

ii) $\beta_c^2 \equiv \frac{d\bar{\alpha}_{jc}}{d\bar{\alpha}_j}$, the slope of the relation between $\bar{\alpha}_{jc}$ and $\bar{\alpha}_j$ (the national mean of skill); and

iii) $\beta_c^3 \equiv \frac{d\bar{\alpha}_{jc}}{d\psi_j}$, the slope of the relation between $\bar{\alpha}_{jc}$ and ψ_j.

Rows 1 and 2 of Table 7 summarize these slopes. The (weighted) national means of the three coefficients across our set of larger CZs and residual geographic areas are 0.88, 0.86, and 0.80, respectively. The standard deviations of the β_c coefficients across CZs, corrected for sampling error, are 0.08, 0.09, and 0.11, respectively.[37]

The CZ-level regressions that generate these coefficients have average R^2 coefficients of 0.75 for β_c^1, 0.83 for β_c^2, and 0.31 for β_c^3. The relatively high average R^2 for the models generating β_c^1 means that the industry wage premiums in a typical CZ are very highly correlated with their national analogues. Likewise, the 0.83 average R^2 for the models generating β_c^2 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 much lower R^2 coefficients for the models generating β_c^3 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.

[37] Recall that $\beta_c^1 < 1$ indicates that CZ *c* has industry premiums that are compressed relative to the national premiums, with a similar interpretation for β_c^2. That the means of β_c^1 and $\beta_c^2$are well below 1 is a minor version of a 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.

The lower rows of Table 7 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.[38] For each regression we show the coefficient and robust standard error, and the within-region R-squared – the explained portion of the variance of β_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 β_c^2 and β_c^3. As noted above (see footnote 16), above, this is what we would expect, since the two coefficients are essentially different measures of the same skill sorting phenomenon. They have a correlation of 0.86 with each other, and we have found that they have similar relationships with other CZ characteristics. Accordingly, in our subsequent analyses we focus on β_c^2; results for β_c^3 are similar when suitably scaled.

Somewhat unexpectedly, there is only a weak (marginally statistically significant) relationship between β_c^2 and β_c^1. 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. The relationship between β_c^1 and β_c^3 is a little stronger.

Table 8 relates β_c^1 and β_c^2 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, but 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

[38] Due to Census disclosure rules, the samples 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 β_c variables on dummies for the residual regions and 4 region indicators have R-squared (computed only over the larger CZs, excluding the composite areas) of 0.26 (β_c^1), 0.57 (β_c^2), and 0.60 (β_c^3).

β_c^3, which is more directly related to the Dauth et al. (2022) analysis but is not reported in the table.

The next rows relate the β_c's to the average pay premiums and average worker effects in the CZ. We see in row 23 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.[39] CZs with higher average pay premiums also have more sorting of high skilled workers to high-premium industries. Since wider 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 expected local average of worker skills:

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

An advantage of these measures is that our normalization choice amounts to an additive constant that does not vary across CZs, so it simply raises or lowers the values of the expected average skill indexes.

[39] 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.

The next rows of Table 8 show regressions of β_c^1 and β_c^2 on the two measures of expected local wage premiums and expected skill. We see a positive coefficients, with 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.

We explore this further in Appendix Table A3, 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 in general that both the expected component and local deviation of average pay premiums and average worker skills affect β_c^1 and β_c^2, 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 β_c^1. 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 8 explore two institutional features that might influence wage premia or worker sorting in a CZ: the local level of unionization and the local level of 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 β_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 β_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 β_c^1. It might also attract higher-skill workers to those industries, reducing β_c^2. 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. Contrary to expectations, the resulting minimum wage measures are *positively* related to both β_cs, significantly so for β_c^2. 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 8 are only 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 just how that structure generates the variation in industry premia and skill sorting examined here.

VII. Conclusions

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 firms in an industry. We show that neither cross-sectional estimates nor two-way fixed effects models with controls for workers and industries (i.e., an industry mover design) identify these effects. The former are biased by the unobserved heterogeneity in worker's skills, combined with the tendency for highly skilled workers to work in high-premium industries. The latter are biased by unobserved heterogeneity in the pay premiums offered by different employers in the same industry, coupled with a tendency for workers who switch industries to move between firms that have similar pay premiums (summing their industry-wide average premium and the firm-specific deviation from that average). 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 ground 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 4-digit NAICS industries of 0.12. Resource-related and capital-intensive industries tend to have 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, 2023), 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 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 similar 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 hierarchy bias that we have identified in the industry mover design is a more general phenomenon. It is not uncommon to use 2-way fixed effects models in which units on one side of the matching process are aggregated. This can lead to biases when there is unobserved heterogeneity within the units being aggregated, and matching depends in part on this heterogeneity. For example, studies 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, acting like the Roy-sorting bias that arises from unit-specific match effects. Further research is needed to understand the scope and magnitude of possible hierarchy biases, and distinguish them from Roy-type biases that are commonly presumed to drive two-sided matching models.

References

Abowd, John M., John Haltiwanger, Ron Jarmin, Julia Lane, Paul Lengermann, Kristin McCue, Kevin McKinney, and Kristin Sandusky. 2005. "The relation among human capital, productivity, and market value: Building up from micro evidence." In *Measuring capital in the new economy* (Carol Corrado, John Haltiwanger, and Dan Sichel, eds.), pp. 153-204. 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.

Abowd, John M., Paul Lengermann, and Kevin L. McKinney. 2003. "The measurement of human capital in the US economy." Longitudinal Employer-Household Dynamic Technical Paper no. TP-2002-09. Washington, DC: U.S. Census Bureau.

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.

Burdett, Kenneth, and Dale T. Mortensen. 1998. "Wage differentials, employer size, and unemployment." *International Economic Review* 39, no. 2: 257-273.

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. 2023. "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. 2018a. "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects." *The Quarterly Journal of Economics* 133, no. 3: 1107-1162.

Chetty, Raj, and Nathaniel Hendren. 2018b. "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. 2020. “The opportunity atlas: Mapping the childhood roots of social mobility.” Working paper. https://opportunityinsights.org/wp-content/uploads/2018/10/atlas_paper.pdf

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.

Gerard, Francois, Lorenzo Lagos, Edson Severnini, and David Card. 2021. Assortative matching or exclusionary hiring? The impact of employment and pay policies on racial wage differences in Brazil. *American Economic Review* 111, no. 10:3418–57.

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: 681-724.

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.

Lachowska, Marta, Alexandre Mas, Raffaele Saggio, and Stephen A. Woodbury. 2023. "Do firm effects drift? Evidence from Washington administrative data." *Journal of Econometrics* 233, no. 2: 375-395.

Lavetti, Kurt, and Ian M. Schmutte. 2023. "Gender differences in sorting on wages and risk." *Journal of Econometrics* 233, no. 2: 507-523.

Manning, Alan. 2003. *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton: Princeton University Press.

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.

Olley, G. Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment." *Econometrica* 64, no. 6: 1263-1297.

Palladino, Marco Guido, Alexandra Roulet, and Mark Stabile. 2021. "Understanding the role of firms in the gender wage gap over time, over the life cycle, and across worker types." CEPR Discussion Paper DP16671.

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

Saez, Emmanuel, Benjamin Schoefer, and David Seim. 2019 "Payroll taxes, firm behavior, and rent sharing: Evidence from a young workers' tax cut in Sweden." *American Economic Review* 109, no. 5: 1717-1763.

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.

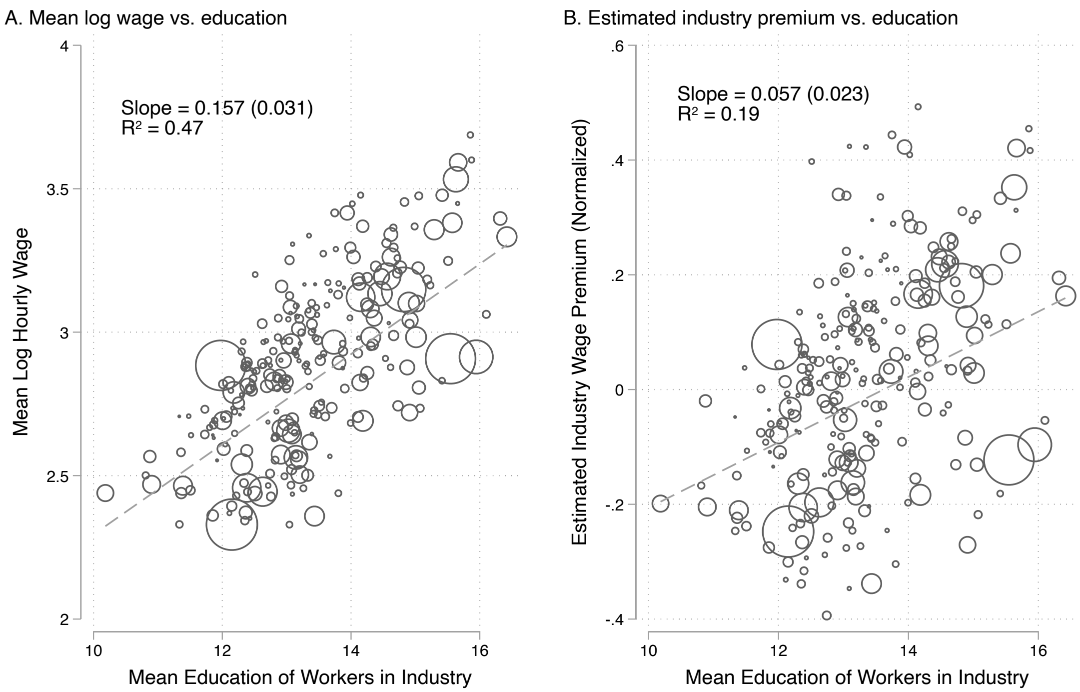
Vaghul, Kavya, and Ben Zipperer. 2021. "State and sub-state historical minimum wage data." https://github.com/benzipperer/historicalminwage. December 10; accessed August 15, 2022.

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.

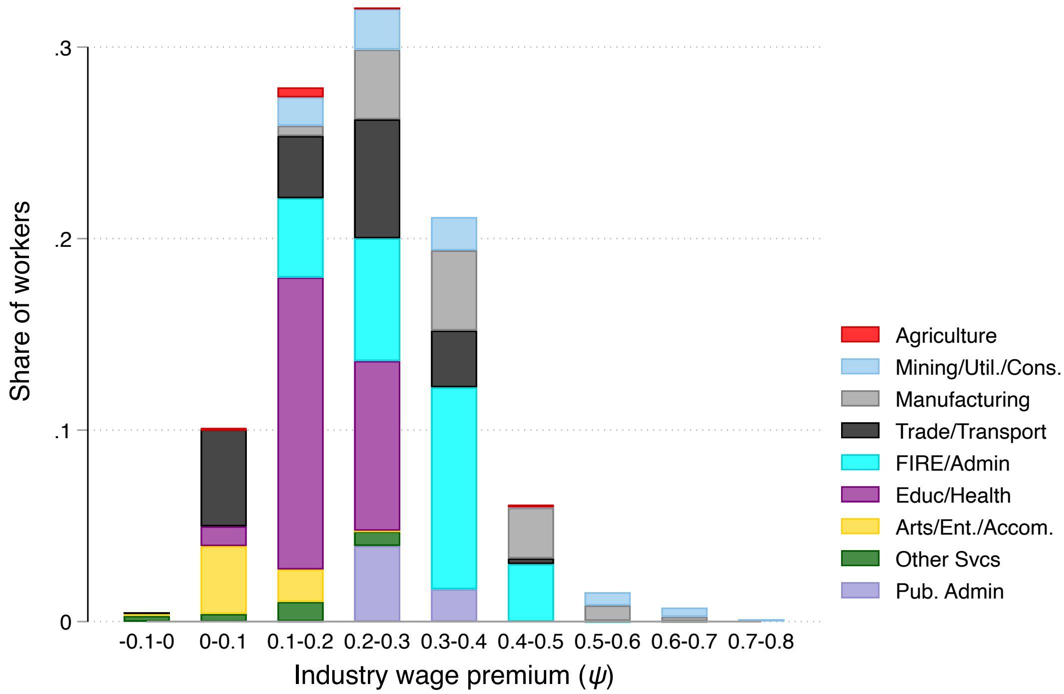
Yeh, Chen, Claudia Macaluso, and Brad Hershbein. 2022. "Monopsony in the US labor market." *American Economic Review* 112, no. 7: 2099-2138.

Figure 1. Average education of workers and average wages by industry, American Community Survey data, 2010-2018



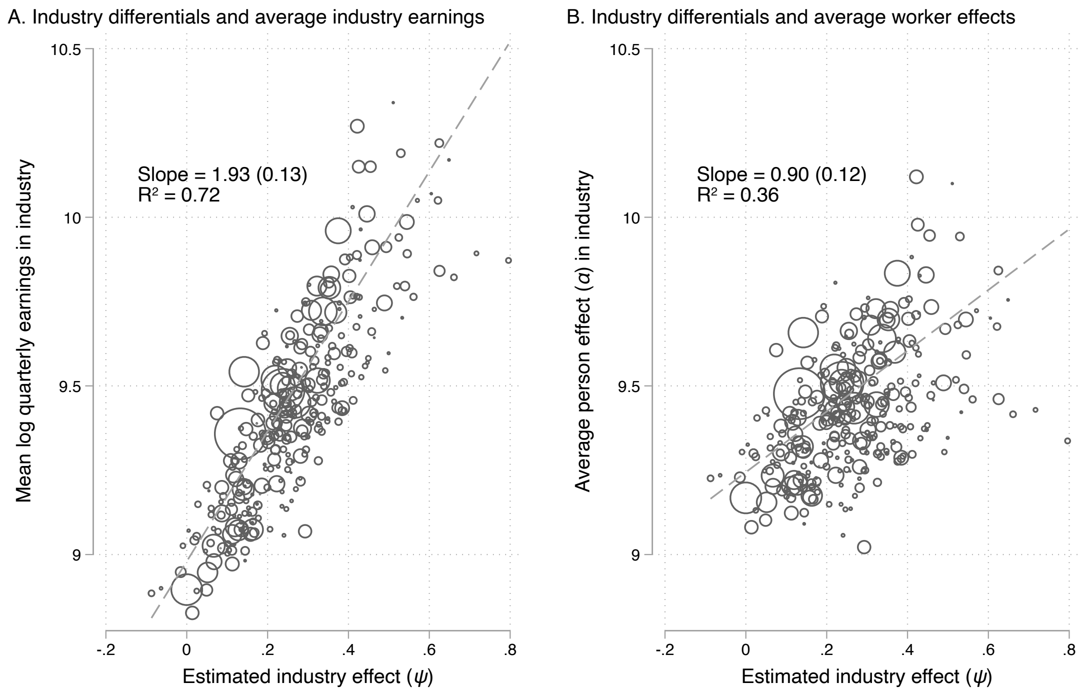
Notes: Samples are pooled 2010-2018 one-year public use samples from the American Community Survey. Individuals aged greater than 62 or with potential experience less than 2 are excluded. Each point corresponds to one of 262 industries that appear in the 2018 data; earlier data are crosswalked to these. 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



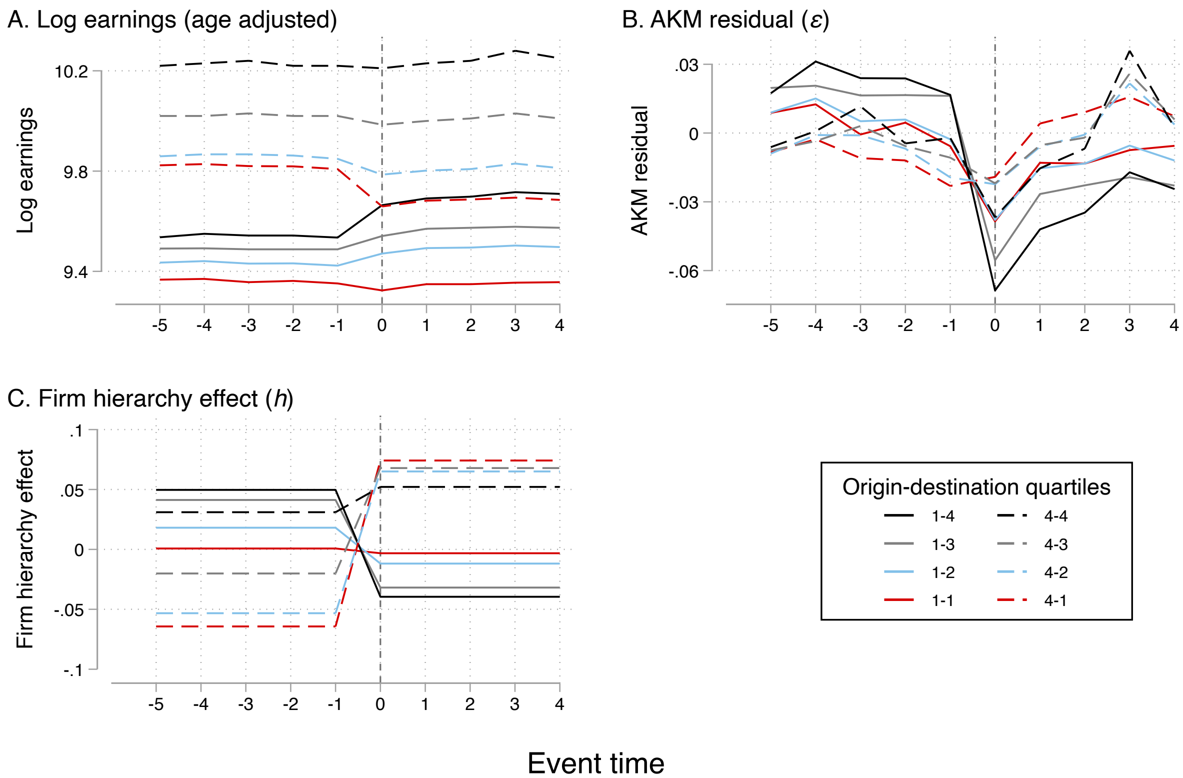
Notes: Figure shows the weighted histogram of estimated industry wage premiums, derived from the "ground-up" estimator described in the text. N=311 industries are weighted by the number of person-quarter observations. Colors represent the contributions coming from one-digit industry groupings.

Figure 3. Industry wage differentials, mean wages, and average worker effects



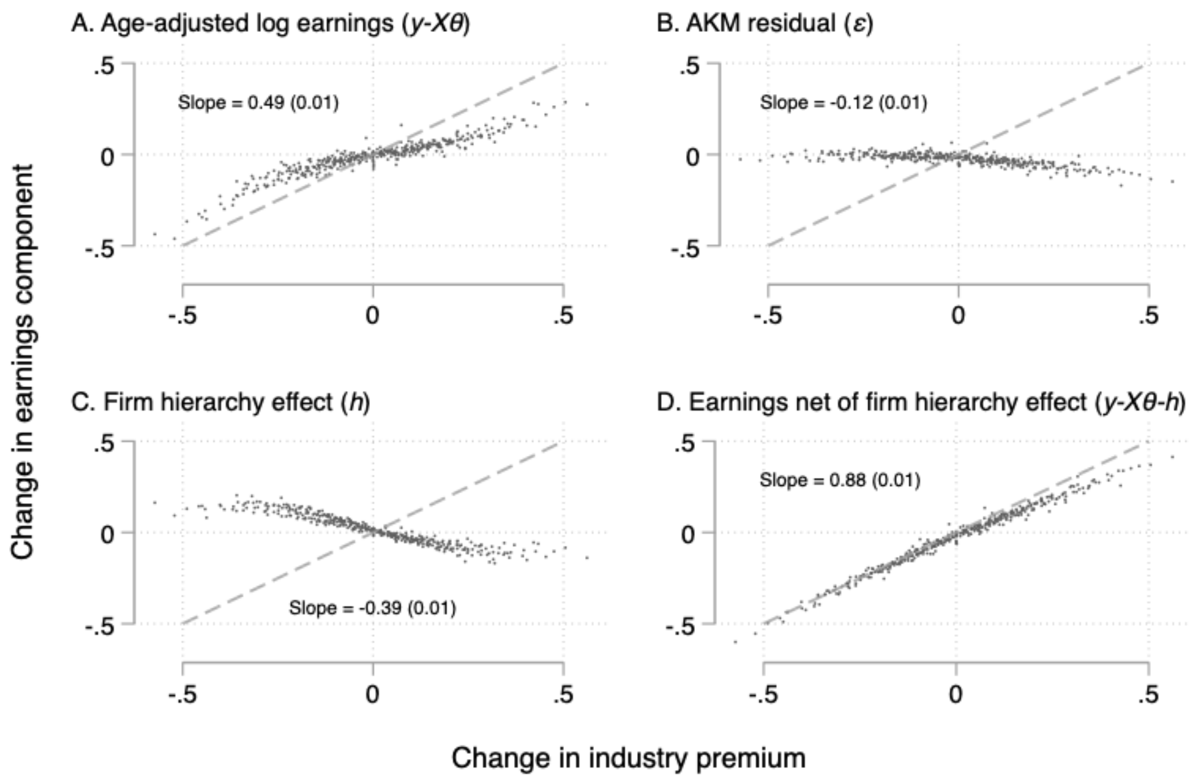
Notes: Figure shows mean log quarterly earnings and estimated industry differential and industry mean person effects from our “ground-up” estimator. N=311 industries are weighted by the number of person-quarter observations in our sample. Regression lines are weighted and robust standard errors is reported.

Figure 4. Event studies for earnings of industry movers



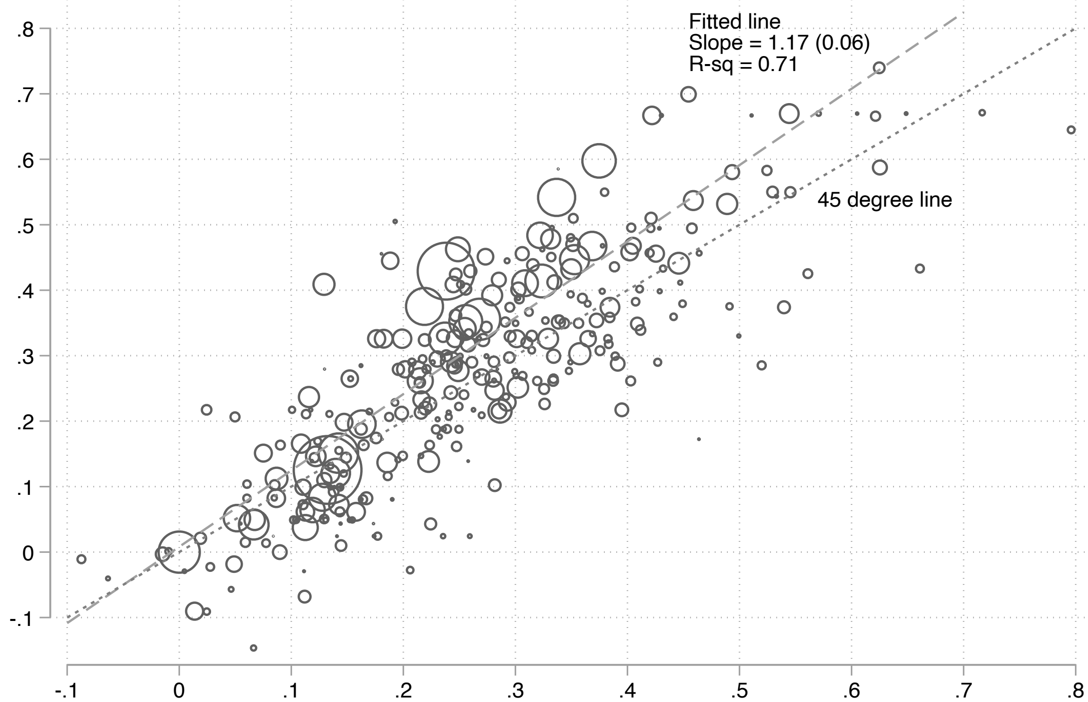
Notes: Figures show event-time means for workers who move between industries within CZs and originate in industries with estimated industry premiums in the top or bottom quartile. See text for definition of the AKM residual and the firm hierarchy effect.

Figure 5. Average earnings changes for industry movers



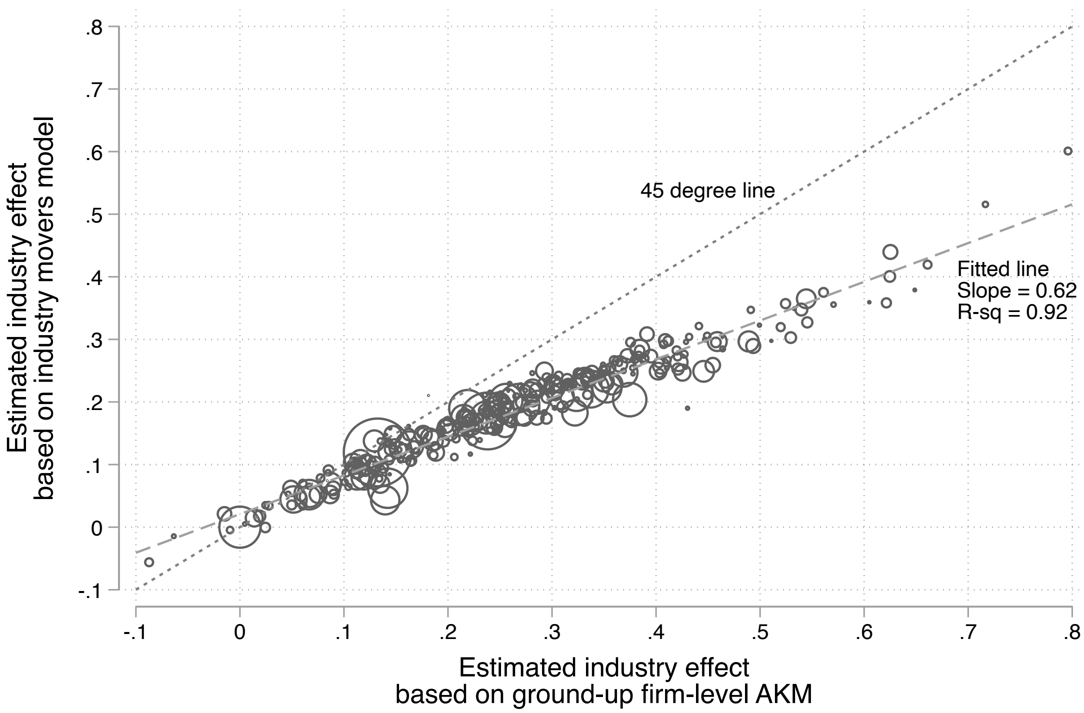
Notes: Within-CZ, between-industry movers are grouped into 400 cells based on vingtiles of their origin and destination industry premiums, from our “bottom-up” estimator. Plot shows, for each cell, the average change in each outcome between the final pre-move quarter and the first post-move quarter.

Figure 6. Comparing cross-sectional and AKM-based estimates of industry premia



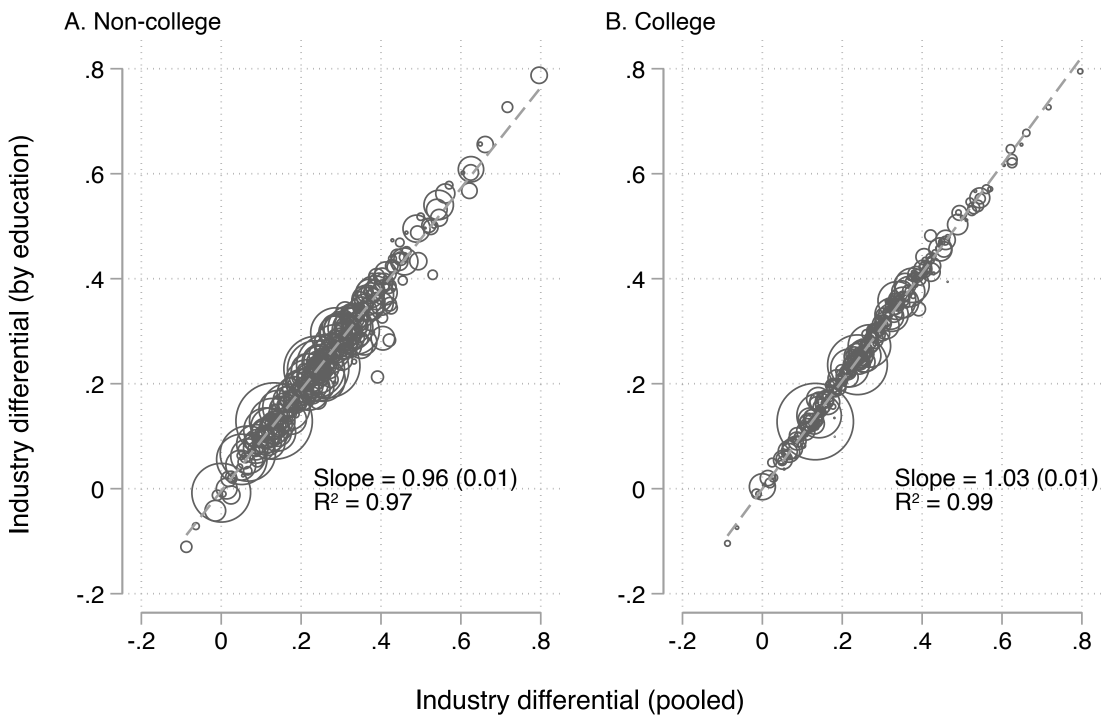
Notes: Estimated industry effects on the X-axis are from our ground-up, firm-level AKM estimator. Those on the Y-axis are from a cross-sectional model estimated on ACS data; see text for controls. Regression line is fit to the 311 industries and weighted by the number of person-quarter observations; robust standard error is reported.

Figure 7. Comparing industry movers and AKM-based estimates of industry premia



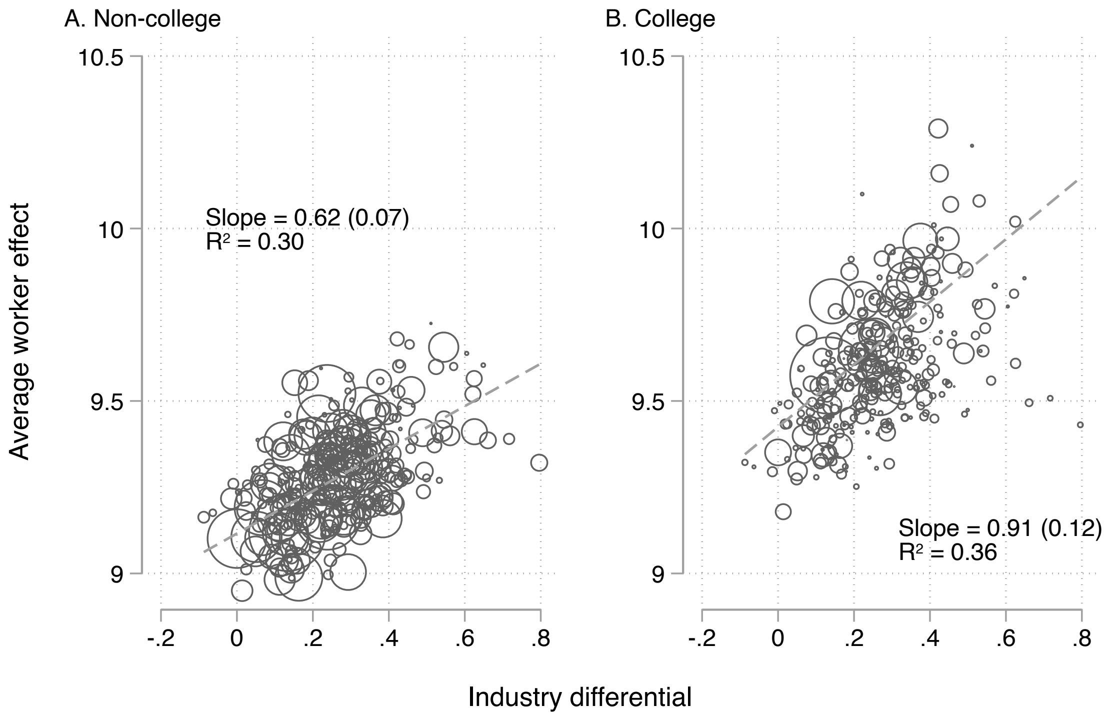
Notes: Estimated industry effects on the X-axis are from our ground-up, firm-level AKM estimator. Those on the Y-axis are from an industry movers model estimated on the LEHD data, with individual and industry fixed effects. Regression line is fit at the industry level and weighted by the number of person-quarter observations.

Figure 8. 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 9. Average worker effects by education and industry



Note: See notes to Figure 8. 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.

Table 1. Summary Statistics for LEHD Samples

	Full sample	Industry stayers	Industry switchers	Event study sample
	(1)	(2)	(3)	(4)
Mean quarterly earnings	15,510	16,050	14,630	16,350
(standard deviation)	(18,020)	(19,710)	(14,860)	(45,670)
Mean age	42	44	40	40
Fraction female	0.47	0.48	0.46	0.46
Fraction 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
Mean number of quarters observed	25.9	27.2	23.7	10.0
(standard deviation)	(7.7)	(7.8)	(7.0)	(0.3)
Number of person-quarter obs. (millions)	2,505	1,544	960.4	87.4
Number of unique people (millions)	111.7	65.7	46.1	8.7

Notes: See text for discussion of sample derivation. "Industry switchers" in column 3 are people observed in more than one industry within a single CZ. "Industry stayers" in column 2 may be observed in multiple CZ's, potentially in different industries in each, but are observed in only one industry per CZ. Event study sample is comprised of 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: Mean Wages, Estimated Wage Premiums, and Estimated Worker Skills for 2-digit Industries

NAICS	Industry Description	Based on Data from LEHD				Based on Data from ACS	
		Percent of Workforce	Mean Log Earnings	Mean Worker Skill (Mean Zero)	Mean Industry Premium	Mean Worker Skill (Mean Zero)	Mean Industry Premium
		(1)	(2)	(3)	(4)	(5)	(6)
11	Agriculture, Forestry, Fishing	0.6%	9.12	-0.24	0.16	-0.20	0.09
21	Mining, Quarrying, Oil and Gas	0.7%	9.91	0.07	0.62	-0.02	0.62
22	Utilities	0.9%	9.88	0.19	0.49	0.11	0.60
23	Construction	5.1%	9.48	0.01	0.25	-0.06	0.35
31	Manufacturing	12.0%	9.50	-0.06	0.35	0.02	0.37
42	Wholesale Trade	5.2%	9.58	0.08	0.30	0.03	0.35
44	Retail Trade	8.7%	9.14	-0.16	0.11	-0.14	0.13
48	Transportation and Warehousing	4.0%	9.40	-0.05	0.24	-0.01	0.31
51	Information	2.6%	9.77	0.19	0.39	0.12	0.43
52	Finance and Insurance	5.7%	9.69	0.17	0.32	0.12	0.52
53	Real Estate and Rental and Leasing	1.5%	9.39	-0.03	0.22	0.05	0.30
54	Professional, Scientific, Tech.	7.5%	9.80	0.27	0.33	0.26	0.48
55	Management of Companies	2.2%	9.72	0.18	0.34	0.21	0.57
56	Admin. and Support	4.6%	9.23	-0.15	0.18	-0.11	0.15
61	Educational Services	10.5%	9.40	0.06	0.13	0.15	0.16
62	Health Care and Social Assistance	14.6%	9.33	-0.06	0.19	0.02	0.30
71	Arts, Ent., Recreation	1.2%	9.25	-0.07	0.12	-0.09	0.14
72	Accommodation and Food Services	4.2%	8.95	-0.28	0.04	-0.27	0.04
81	Other Svcs. (except Public Admin)	2.4%	9.26	-0.08	0.14	-0.06	0.11
92	Public Administration	5.7%	9.47	-0.01	0.28	0.13	0.43
Weighted std. dev. across sectors		--	0.22	0.14	0.10	0.13	0.15

Notes: See text for descriptions of samples used in LEHD (columns 1-4) and ACS (columns 5-6). Entry in column 3 is average of estimated person effects from AKM models fit by CZ, normalized to have mean zero across all workers. Entry in column 4 is average wage premium, based on person-quarter weighted average of establishment premiums from AKM models fit by CZ, and normalized to have mean 0 in restaurants and eating places (NAICS 7225). Entry in column 5 is mean of estimated wage index ($X\theta$) from cross sectional regression model, normalized to have mean zero across all workers. Entry in column 6 is mean of estimated industry effects for industries in the 2-digit sector from cross sectional regression model, normalized to have mean of 0.04 in accommodation and food services.

Table 3: Mean Wages, Estimated Wage Premiums, and Estimated Worker Skills for 3-digit Manufacturing Industries

		Based on Data from LEHD				Based on Data from ACS	
NAICS	Industry Description	Percent of Workforce	Mean Log Earnings	Mean Worker Skill (Mean Zero)	Mean Industry Premium	Mean Worker Skill (Mean Zero)	Mean Industry Premium
		(1)	(2)	(3)	(4)	(5)	(6)
311-2	Food and kindred products	1.5%	9.28	-0.24	0.32	-0.13	0.27
313-4	Textile mill products	0.2%	9.14	-0.29	0.22	-0.14	0.23
315-6	Apparel and leather	0.1%	9.10	-0.23	0.15	-0.13	0.15
321	Wood Products	0.3%	9.20	-0.22	0.22	-0.11	0.21
322	Paper	0.4%	9.53	-0.09	0.43	0.00	0.44
323	Printing and Related	0.4%	9.28	-0.13	0.22	0.00	0.26
324	Petroleum and Coal	0.1%	10.05	0.22	0.62	0.10	0.69
325	Chemicals	0.9%	9.78	0.11	0.47	0.14	0.52
326	Plastics and Rubber	0.7%	9.32	-0.19	0.30	-0.06	0.30
327	Nonmetallic Minerals	0.4%	9.42	-0.12	0.34	-0.03	0.34
331	Primary Metals	0.4%	9.57	-0.12	0.48	-0.02	0.42
332	Fabricated Metals	1.4%	9.39	-0.09	0.28	-0.03	0.32
333	Machinery	1.1%	9.54	0.00	0.34	0.04	0.40
334	Computer and Electronic	1.2%	9.86	0.26	0.42	0.26	0.51
335	Electrical Equipment	0.4%	9.48	-0.06	0.34	0.04	0.38
336	Transportation Equipment	1.7%	9.65	0.02	0.43	0.06	0.46
337	Furniture and Related	0.3%	9.19	-0.21	0.19	-0.10	0.22
339	Miscellaneous	0.6%	9.42	-0.05	0.28	0.02	0.31
Weighted std. dev. across sectors		--	0.22	0.15	0.09	0.11	0.10

Notes: See notes to Table 2.

Table 4. Relationships Between Worker Skills and Industry Wage Premiums

Explanatory Variable:	Based on LEHD			Based on Linked LEHD and ACS				
	$\bar{\alpha}_j$	ψ_j	$\bar{\alpha}_j$	$\bar{\alpha}_j$	$\bar{\alpha}_j$	$\bar{\alpha}_j$	$E[\bar{\alpha}_j \mid \bar{X}_j\hat{\theta}]$	$\bar{\alpha}_j - E[\bar{\alpha}_j \mid \bar{X}_j\hat{\theta}]$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean earnings in industry $(\bar{y}_j)$	0.62 (0.02)	0.37 (0.02)						
Industry premium (ψ_j)			0.90 (0.12)	0.90 (0.12)		0.33 (0.10)	0.66 (0.15)	0.24 (0.08)
Skill index $(\bar{X}_j\hat{\theta})$					1.04 (0.07)	0.90 (0.08)		
Num. observations	311	311	311	310	310	310	310	310
R-squared	0.87	0.72	0.36	0.36	0.71	0.75	0.27	0.09

Note: Each column is a separate regression model, estimated across 4-digit industries, with dependent variable in column heading. In columns 4-8 the sample excludes one industry (Central Banking) that cannot be linked to ACS. Skill index is constructed from ACS data and merged to 4-digit industry data from LEHD. All models are fit by weighted OLS, using the number of person-quarter observations in the industry in our LEHD sample as a weight. Robust standard errors in parentheses.

Table 5. 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, squared	-0.0033	-0.0029	-0.0016	-0.0016
	(0.0006)	(0.0003)	(0.0003)	(0.0002)
Fixed effects for worker, CZ, industry, time	N	Y	N	Y
Person-quarter observations (millions)	89.8	89.8	421.8	421.8
Adjusted R-squared	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: Each column is a separate regression. Dependent variable in all columns is the hierarchy effect for a given worker in a given quarter (the difference between the workplace pay premium and the industry average pay premium). Young workers in columns 1-2 are those who were not yet 27 at the beginning of 2010; older workers are all others in our main sample. Industry experience is the cumulative number of quarters from 2010-Q1 to date that the worker has been observed in their current industry (including previous spells). Standard errors are clustered at the industry level.

Table 6. Comparisons of Industry Effects from Alternative Models

	Preferred model	Cross-sectional models (fit to ACS)			Industry Mover models (fit to LEHD)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Alternative model controls for:							
Person fixed effects					X	X	X
Year effects and age/experience		X	X	X	X	X	X
Education, gender, and race/ethicity		X	X	X			
Field of study and immigration			X	X			
CZ fixed effects				X		X	
Industry-by-CZ fixed effects							X
Standard deviation of industry effects	0.122	0.189	0.174	0.169	0.079	0.079	0.082
Regression of alternative model	1.00	1.24	1.17	1.17	0.62	0.62	0.66
estimates on preferred model		(0.07)	(0.06)	(0.06)	(0.02)	(0.02)	(0.01)
R^2 (adjusted)		0.645	0.677	0.710	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. Alternative models in columns 2-4 are fit to ACS, and include 222 NAICS codes (some combining 4 digit codes). These are then merged to 310 4-digit NAICS codes. Alternative models in columns 5-7 are fit to LEHD with person effects and 311 4-digit NAICS dummies, as well as year and age effects. Robust standard errors in parentheses.

Table 7. Summaries of CZ-specific β Coefficients

	β_c^1	β_c^2	β_c^3
	(1)	(2)	(3)
Mean	0.880	0.863	0.799
Standard deviation	0.082	0.097	0.134
Sampling-error adjusted SD	0.076	0.094	0.115
Regressions of β in column heading on β indicated in row:			
Models for β_c^1			
Coefficient		0.36	0.62
(Standard error)		(0.21)	(0.20)
R^2 (within-region)		0.14	0.26
Models for β_c^2			
Coefficient	0.43		1.14
(Standard error)	(0.22)		(0.14)
R^2 (within-region)	0.14		0.74
Models for β_c^3			
Coefficient	0.42	0.65	
(Standard error)	(0.09)	(0.06)	
R^2 (within-region)	0.26	0.74	

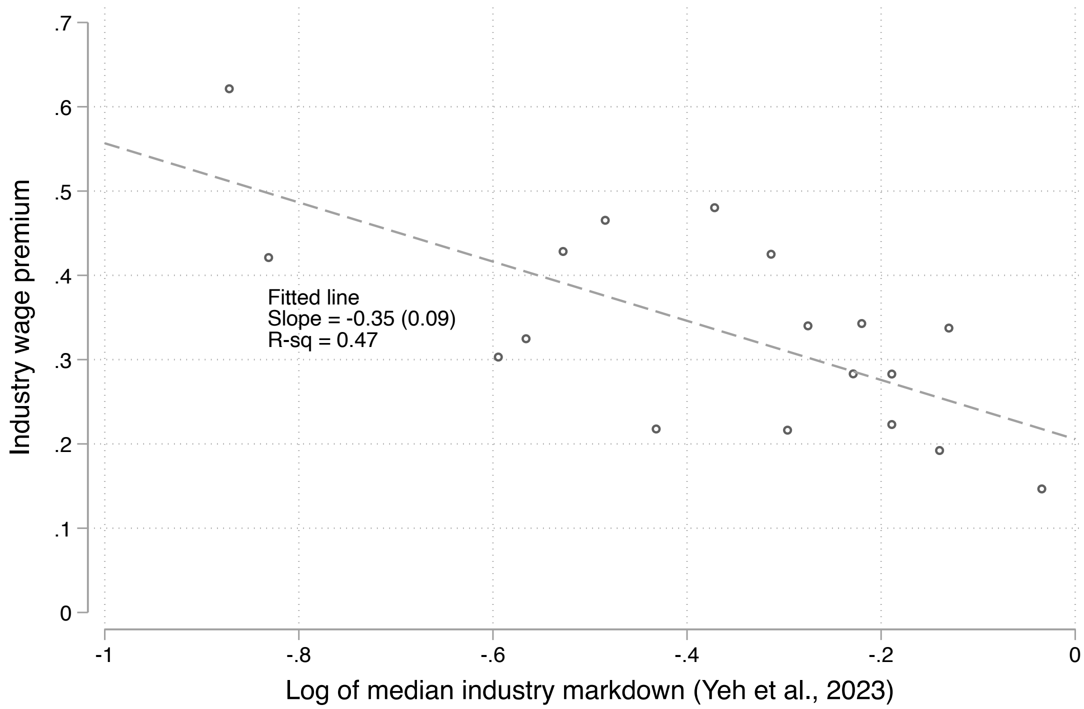
Note: β cofficients are slopes of a CZ's estimated industry effects or industry average person effects with respect to the national industry effect or industry average person effect -- see text. Summary statistics are weighted by the number of person-quarter observations in each CZ. Regressions relate one of the β coefficients for a CZ to another β coefficient for the same CZ, and control for indicators for four regions and for each of approximately 10 composite CZ's. N≈60. Reported R-squared statistics indicate the share of the within-CZ variation in the β explained, among the approximately 50 non-composite CZs. Regressions are weighted by the number of person-quarter observations in the CZ; and standard errors are heteroskedasticity-robust.

Table 8. Bivariate Relationships Between Beta Coefficients and CZ Characteristics

	Mean	Dependent Variable:			
		β_c^1		β_c^2	
	[Std Dev.]	Coefficient	R^2	Coefficient	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 pay premium 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 expected average worker skills, given industry composition:					
Expected mean pay premium 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 additional controls 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 Figure A1. Relationship between markdowns and industry average pay premiums across 3-digit manufacturing industries



Note: Each dot represents one of 18 3-digit manufacturing industries reported in Table 3. X-axis represents log of median industry wages divided by marginal revenue products, as reported by Yeh et al. (2023, Table 1); more negative values represent larger markdowns. Y-axis represents mean of industry premiums from ground-up approach (Table 3, column 3). Fitted regression line is unweighted, with robust standard errors.

Table A1: Summary of AKM Estimation Results

	Variance / covariance	Std. Dev.	Correlation	Variance share
	(1)	(2)	(3)	(4)
Panel A: Average of CZ-level variances/covariances computed from CZ-level AKM models				
Log quarterly earnings	0.406	0.637		1.000
Estimated person, establishment effects and covariates				
Person effects	0.274	0.523		0.675
Establishment effects	0.044	0.210		0.108
Covariate index	0.019	0.137		0.046
Residual	0.049	0.222		0.121
Covariances				
(person effects, establishment effects)	0.014		0.132	0.071
(person effects, covariate index)	-0.005		-0.070	-0.025
(establishment effects, covariate index)	0.001		0.023	0.003
Panel B: Variances/covariances from national AKM model (Card, Rothstein, and Yi 2023)				
Log quarterly earnings	0.426	0.653		1.000
Estimated person, establishment effects and covariates				
Person effects	0.274	0.524		0.644
Establishment effects	0.050	0.223		0.117
Covariate index	0.018	0.135		0.043
Residual	0.054	0.232		0.127
Covariances				
(person effects, establishment effects)	0.021		0.182	0.100
(person effects, covariate index)	-0.007		-0.100	-0.033
(establishment effects, covariate index)	0.001		0.124	0.003

Note: Table presents summaries of estimated AKM models. Panel A summarizes results from the primary models used in this paper, fit separately to each of approximately 60 CZs or regional aggregates. Panel B summarizes results from an alternative model that is fit to the entire U.S., as in Card, Rothstein, and Yi (2023). Total sample size is approximately 2.505 billion person-quarters.

Appendix Table A-2. Industry Wage Premiums, Mean Worker Effects, & Average Pay, 4-digit industries

NAICS	Industry	Mean log earnings	Mean worker effect (normalized to mean 0)	Industry premium	Percent of workforce
		(1)	(2)	(3)	(4)
	Across industry mean	9.434	9.456	0.236	
	Across industry SD	0.278	0.185	0.122	
1111	Oilseed and Grain Farming	9.216	-0.15	0.152	0.03%
1112	Vegetable and Melon Farming	9.120	-0.21	0.155	0.04%
1113	Fruit and Tree Nut Farming	9.013	-0.27	0.104	0.06%
1114	Greenhouse, Nursery, and Floriculture	9.032	-0.26	0.102	0.08%
1119	Other Crop Farming	9.113	-0.22	0.131	0.03%
1121	Cattle Ranching and Farming	9.108	-0.27	0.177	0.08%
1122	Hog and Pig Farming	9.159	-0.32	0.259	0.02%
1123	Poultry and Egg Production	9.143	-0.31	0.236	0.03%
1124	Sheep and Goat Farming	9.019	-0.25	0.084	0.00%
1125	Aquaculture	9.185	-0.15	0.141	0.00%
1129	Other Animal Production	9.150	-0.23	0.174	0.01%
1131	Timber Tract Operations	9.555	0.08	0.284	0.00%
1132	Forest Nurseries and Forest Products	9.176	-0.16	0.130	0.00%
1133	Logging	9.254	-0.15	0.195	0.04%
1141	Fishing	9.643	0.00	0.464	0.00%
1142	Hunting and Trapping	9.179	-0.19	0.181	0.00%
1151	Support Activities for Crop Production	9.098	-0.26	0.164	0.11%
1152	Support Activities for Animal Production	9.178	-0.18	0.162	0.02%
1153	Support Activities for Forestry	9.330	-0.06	0.190	0.01%
2111	Oil and Gas Extraction	10.220	0.39	0.625	0.18%
2121	Coal Mining	9.872	-0.12	0.796	0.07%
2122	Metal Ore Mining	9.893	-0.03	0.717	0.05%
2123	Nonmetallic Mineral Mining & Quarrying	9.536	-0.08	0.407	0.09%
2131	Support Activities for Mining	9.841	0.00	0.625	0.30%
2211	Electric Power Gener., Trans. and Distn.	9.986	0.24	0.544	0.55%
2212	Natural Gas Distribution	9.940	0.22	0.525	0.13%
2213	Water, Sewage and Other Systems	9.575	0.06	0.316	0.21%
2361	Residential Building Construction	9.399	0.01	0.176	0.46%
2362	Nonresidential Building Construction	9.652	0.12	0.329	0.62%
2371	Utility System Construction	9.596	0.02	0.365	0.37%
2372	Land Subdivision	9.631	0.17	0.267	0.03%
2373	Highway, Street, and Bridge Construction	9.504	0.00	0.301	0.45%
2379	Other Heavy and Civil Eng. Construction	9.649	0.06	0.383	0.09%
2381	Fndtn., Struct., & Bldg. Exterior Contractors	9.347	-0.06	0.199	0.53%
2382	Building Equipment Contractors	9.503	0.06	0.237	1.64%

Appendix Table A-2 (continued)

NAICS		Mean log earnings	Mean worker effect (normalized to mean 0)	Industry premium	Percent of workforce
		(1)	(2)	(3)	(4)
2383	Building Finishing Contractors	9.324	-0.07	0.183	0.49%
2389	Other Specialty Trade Contractors	9.409	-0.05	0.246	0.45%
3111	Animal Food Manufacturing	9.454	-0.12	0.365	0.05%
3112	Grain and Oilseed Milling	9.582	-0.07	0.449	0.07%
3113	Sugar and Confectionery Product Manuf.	9.395	-0.15	0.348	0.06%
3114	Fruit & Veg. Prsrvng. & Spec. Food Manuf.	9.278	-0.24	0.326	0.16%
3115	Dairy Product Manufacturing	9.457	-0.16	0.412	0.14%
3116	Animal Slaughtering and Processing	9.069	-0.43	0.293	0.44%
3117	Seafood Product Preparation and Pkging.	9.226	-0.23	0.281	0.02%
3118	Bakeries and Tortilla Manufacturing	9.217	-0.26	0.282	0.21%
3119	Other Food Manufacturing	9.373	-0.16	0.326	0.17%
3121	Beverage Manufacturing	9.452	-0.06	0.310	0.18%
3122	Tobacco Manufacturing	9.702	-0.03	0.533	0.01%
3131	Fiber, Yarn, and Thread Mills	9.057	-0.40	0.241	0.03%
3132	Fabric Mills	9.194	-0.28	0.270	0.05%
3133	Textile & Fabric Finishing & Coating Mills	9.219	-0.21	0.233	0.03%
3141	Textile Furnishings Mills	9.135	-0.32	0.246	0.05%
3149	Other Textile Product Mills	9.074	-0.26	0.135	0.05%
3151	Apparel Knitting Mills	8.982	-0.37	0.144	0.01%
3152	Cut and Sew Apparel Manufacturing	9.122	-0.19	0.143	0.07%
3159	Apparel Access. & Other Apparel Manuf.	9.072	-0.25	0.138	0.01%
3161	Leather and Hide Tanning and Finishing	9.243	-0.21	0.258	0.00%
3162	Footwear Manufacturing	9.094	-0.26	0.162	0.01%
3169	Other Leather & Allied Product Manuf.	9.030	-0.28	0.118	0.01%
3211	Sawmills and Wood Preservation	9.252	-0.20	0.250	0.08%
3212	Veneer, Plywood, & Engineered Wood	9.297	-0.21	0.292	0.07%
3219	Other Wood Product Manufacturing	9.146	-0.24	0.176	0.17%
3221	Pulp, Paper, and Paperboard Mills	9.764	0.00	0.561	0.12%
3222	Converted Paper Product Manufacturing	9.435	-0.14	0.372	0.29%
3231	Printing and Related Support Activities	9.282	-0.13	0.216	0.42%
3241	Petroleum and Coal Products Manuf.	10.050	0.22	0.621	0.13%
3251	Basic Chemical Manufacturing	9.892	0.14	0.545	0.16%
3252	Resin, Synth. Rubber, & Artif. Fibers	9.796	0.06	0.520	0.10%
3253	Pesticide, Fertilizer, & Other Ag. Chemical	9.727	0.05	0.464	0.04%
3254	Pharmaceutical & Medicine Manuf.	9.912	0.21	0.493	0.31%
3255	Paint, Coating, and Adhesive Manuf.	9.568	0.02	0.349	0.06%
3256	Soap, Cleaning, & Toilet Prep. Manuf.	9.499	-0.04	0.339	0.10%
3259	Other Chemical Product and Preparation	9.555	-0.02	0.380	0.09%

Appendix Table A-2 (continued)

NAICS		Mean log earnings	Mean worker effect (normalized to mean 0)	Industry premium	Percent of workforce
		(1)	(2)	(3)	(4)
3261	Plastics Product Manufacturing	9.292	-0.20	0.281	0.53%
3262	Rubber Product Manufacturing	9.427	-0.17	0.384	0.14%
3271	Clay Product and Refractory Manufacturing	9.349	-0.15	0.299	0.04%
3272	Glass and Glass Product Manufacturing	9.427	-0.16	0.389	0.08%
3273	Cement & Concrete Product Manufacturing	9.420	-0.10	0.315	0.17%
3274	Lime and Gypsum Product Manufacturing	9.597	-0.11	0.499	0.02%
3279	Other Nonmetallic Mineral Manuf.	9.414	-0.11	0.315	0.07%
3311	Iron and Steel Mills and Ferroalloy Manuf.	9.822	-0.04	0.661	0.10%
3312	Steel Product Manuf. from Purchased Steel	9.526	-0.11	0.432	0.06%
3313	Alumina and Aluminum	9.546	-0.15	0.491	0.06%
3314	Nonferrous Metal (except Aluminum)	9.559	-0.08	0.441	0.07%
3315	Foundries	9.394	-0.18	0.376	0.12%
3321	Forging and Stamping	9.392	-0.11	0.306	0.10%
3322	Cutlery and Handtool Manufacturing	9.394	-0.09	0.301	0.04%
3323	Architectural and Structural Metals Manuf.	9.349	-0.11	0.258	0.33%
3324	Boiler, Tank, & Shipping Container Manuf.	9.508	-0.08	0.383	0.09%
3325	Hardware Manufacturing	9.358	-0.12	0.274	0.02%
3326	Spring and Wire Product Manufacturing	9.296	-0.15	0.251	0.04%
3327	Mach. Shops; Turned Prod.; & Screw Manuf.	9.383	-0.06	0.246	0.34%
3328	Coating, Engraving, Heat Treating, etc	9.275	-0.18	0.254	0.12%
3329	Other Fabricated Metal Product Manuf.	9.458	-0.08	0.334	0.27%
3331	Agric., Const., & Mining Machinery Manuf.	9.598	-0.02	0.409	0.25%
3332	Industrial Machinery Manufacturing	9.641	0.12	0.332	0.11%
3333	Commercial and Svc. Ind. Machinery Manuf.	9.557	0.05	0.312	0.09%
3334	HVAC Equipment Manufacturing	9.351	-0.15	0.291	0.13%
3335	Metalworking Machinery Manufacturing	9.475	0.00	0.274	0.18%
3336	Engine, Turbine, & Power Equip. Manuf.	9.675	0.05	0.420	0.10%
3339	Other General Purpose Machinery Manuf.	9.536	0.00	0.338	0.27%
3341	Computer and Peripheral Equip. Manuf.	10.190	0.49	0.530	0.18%
3342	Communications Equipment Manufacturing	9.881	0.30	0.403	0.11%
3343	Audio and Video Equipment Manufacturing	9.692	0.17	0.333	0.02%
3344	Semicond. & Electronic Component Manuf.	9.763	0.18	0.405	0.40%
3345	Control Instruments Manufacturing	9.826	0.24	0.402	0.44%
3346	Magnetic and Optical Media	9.728	0.17	0.378	0.02%
3351	Electric Lighting Equipment Manufacturing	9.401	-0.10	0.300	0.05%
3352	Household Appliance Manufacturing	9.386	-0.15	0.327	0.07%
3353	Electrical Equipment Manufacturing	9.538	0.00	0.344	0.15%
3359	Other Electr. Equip. & Component Manuf.	9.481	-0.07	0.356	0.14%

Appendix Table A-2 (continued)

NAICS		Mean log earnings	Mean worker effect (normalized to mean 0)	Industry premium	Percent of workforce
		(1)	(2)	(3)	(4)
3361	Motor Vehicle Manufacturing	9.796	0.03	0.540	0.22%
3362	Motor Vehicle Body and Trailer Manuf.	9.311	-0.21	0.295	0.12%
3363	Motor Vehicle Parts Manufacturing	9.434	-0.17	0.384	0.55%
3364	Aerospace Product and Parts Manuf.	9.911	0.28	0.459	0.56%
3365	Railroad Rolling Stock Manufacturing	9.569	-0.09	0.447	0.03%
3366	Ship and Boat Building	9.511	-0.08	0.388	0.13%
3369	Other Transportation Equipment Manuf.	9.495	-0.07	0.369	0.03%
3371	Furniture & Kitchen Cabinet Manuf.	9.127	-0.25	0.162	0.20%
3372	Office Furniture (including Fixtures) Manuf.	9.312	-0.13	0.239	0.10%
3379	Other Furniture Related Product Manuf.	9.186	-0.25	0.236	0.03%
3391	Medical Equipment and Supplies Manuf.	9.518	-0.01	0.335	0.31%
3399	Other Miscellaneous Manufacturing	9.311	-0.11	0.220	0.25%
4231	Motor Vehicle & Parts Wholesalers	9.390	-0.03	0.223	0.28%
4232	Furniture & Home Furnishing Wholesalers	9.423	0.01	0.217	0.09%
4233	Lumber & Other Const. Materials Whlsl.	9.424	-0.02	0.238	0.19%
4234	Prof. & Comm. Equip. & Supplies Whlsl.	9.788	0.24	0.350	0.63%
4235	Metal & Mineral (except Petroleum) Whlsl.	9.491	-0.01	0.304	0.12%
4236	Appliances & Electrical Wholesalers	9.667	0.17	0.303	0.32%
4237	Hardware,Plumbing & Heating Wholesalers	9.477	0.04	0.236	0.23%
4238	Mach., Equip., & Supplies Wholesalers	9.547	0.06	0.279	0.63%
4239	Misc. Durable Goods Wholesalers	9.352	-0.06	0.216	0.24%
4241	Paper and Paper Product Wholesalers	9.482	0.03	0.258	0.11%
4242	Drugs and Druggists' Sundries Wholesalers	9.888	0.25	0.421	0.19%
4243	Apparel, Piece Goods, Notions Wholesalers	9.447	0.06	0.214	0.12%
4244	Grocery and Related Product Wholesalers	9.416	-0.09	0.302	0.66%
4245	Farm Product Raw Material Wholesalers	9.410	-0.05	0.266	0.05%
4246	Chemical and Allied Products Wholesalers	9.671	0.11	0.352	0.12%
4247	Petroleum & Products Wholesalers	9.584	0.03	0.349	0.08%
4248	Beer, Wine, & Distilled Alcohol Wholesalers	9.509	0.01	0.295	0.19%
4249	Misc. Nondurable Goods Wholesalers	9.348	-0.05	0.199	0.26%
4251	Wholesale Elec. Mkts. & Agents & Brokers	9.831	0.27	0.358	0.70%
4411	Automobile Dealers	9.448	0.03	0.215	0.99%
4412	Other Motor Vehicle Dealers	9.260	-0.06	0.113	0.11%
4413	Automotive Parts, Access., & Tire Stores	9.134	-0.17	0.111	0.37%
4421	Furniture Stores	9.209	-0.13	0.149	0.16%
4422	Home Furnishings Stores	9.278	-0.04	0.121	0.13%
4431	Electronics and Appliance Stores	9.354	-0.05	0.230	0.36%
4441	Building Material and Supplies Dealers	9.122	-0.15	0.087	0.75%

Appendix Table A-2 (continued)

NAICS		Mean log earnings	Mean worker effect (normalized to mean 0)	Industry premium	Percent of workforce
		(1)	(2)	(3)	(4)
4442	Lawn and Garden Equip. & Supplies Stores	9.118	-0.14	0.060	0.09%
4451	Grocery Stores	9.026	-0.22	0.067	1.38%
4452	Specialty Food Stores	9.068	-0.19	0.078	0.09%
4453	Beer, Wine, and Liquor Stores	9.026	-0.15	-0.009	0.07%
4461	Health and Personal Care Stores	9.236	-0.06	0.116	0.63%
4471	Gasoline Stations	8.895	-0.35	0.049	0.37%
4481	Clothing Stores	9.168	-0.12	0.130	0.32%
4482	Shoe Stores	9.159	-0.15	0.147	0.06%
4483	Jewelry, Luggage, & Leather Goods Stores	9.282	-0.04	0.143	0.08%
4511	Sporting Goods & Musical Inst. Stores	9.042	-0.16	0.019	0.20%
4512	Book Stores and News Dealers	9.010	-0.21	0.046	0.04%
4521	Department Stores	8.979	-0.25	0.067	0.65%
4529	General Merch Stores & Supercenters	8.947	-0.30	0.052	1.09%
4531	Florists	8.900	-0.22	-0.064	0.02%
4532	Office Supplies, Stationery, and Gift Stores	9.154	-0.14	0.111	0.13%
4533	Used Merchandise Stores	8.892	-0.33	0.025	0.06%
4539	Other Miscellaneous Store Retailers	9.117	-0.16	0.085	0.15%
4541	Electronic Shopping and Mail-Order Houses	9.454	0.02	0.242	0.26%
4542	Vending Machine Operators	9.096	-0.24	0.135	0.03%
4543	Direct Selling Establishments	9.359	-0.07	0.229	0.11%
4811	Scheduled Air Transportation	9.627	0.25	0.188	0.42%
4812	Nonscheduled Air Transportation	9.745	0.25	0.293	0.04%
4821	Rail Transportation	9.485	-0.01	0.338	0.00%
4831	Ocean & Great Lakes Water Trans.	9.746	0.18	0.378	0.04%
4832	Inland Water Transportation	9.706	0.07	0.429	0.03%
4841	General Freight Trucking	9.375	-0.12	0.286	0.83%
4842	Specialized Freight Trucking	9.367	-0.13	0.285	0.34%
4851	Urban Transit Systems	9.602	0.03	0.395	0.25%
4852	Interurban and Rural Bus Transportation	9.276	-0.18	0.263	0.02%
4853	Taxi and Limousine Service	9.077	-0.18	0.066	0.04%
4854	School and Employee Bus Transportation	9.054	-0.17	0.024	0.14%
4855	Charter Bus Industry	9.115	-0.18	0.117	0.02%
4859	Other Transit & Ground Pass. Trans.	9.003	-0.28	0.101	0.06%
4861	Pipeline Transportation of Crude Oil	10.170	0.30	0.649	0.01%
4862	Pipeline Transportation of Natural Gas	10.050	0.27	0.571	0.03%
4869	Other Pipeline Transportation	10.070	0.24	0.605	0.01%
4871	Scenic & Sightseeing Transportation, Land	9.203	-0.12	0.144	0.01%
4872	Scenic & Sightseeing Transportation, Water	9.189	-0.04	0.055	0.01%

Appendix Table A-2 (continued)

NAICS		Mean log earnings	Mean worker effect (normalized to mean 0)	Industry premium	Percent of workforce
		(1)	(2)	(3)	(4)
4879	Scenic & Sightseeing Transportation, Other	9.389	0.01	0.173	0.00%
4881	Support Activities for Air Transportation	9.421	-0.04	0.261	0.18%
4882	Support Activities for Rail Transportation	9.403	-0.16	0.349	0.03%
4883	Support Activities for Water Transportation	9.763	0.14	0.427	0.08%
4884	Support Activities for Road Transportation	9.254	-0.16	0.208	0.08%
4885	Freight Transportation Arrangement	9.448	0.00	0.248	0.18%
4889	Other Support Activities for Transportation	9.261	-0.17	0.226	0.02%
4911	Postal Service	9.138	-0.25	0.181	0.00%
4921	Couriers and Express Delivery Services	9.472	0.11	0.152	0.42%
4922	Local Messengers and Local Delivery	9.138	-0.22	0.153	0.04%
4931	Warehousing and Storage	9.210	-0.22	0.223	0.68%
5111	Newspaper, Periodical, Book Publishers	9.518	0.07	0.270	0.37%
5112	Software Publishers	10.150	0.49	0.455	0.33%
5121	Motion Picture and Video Industries	9.749	0.27	0.295	0.17%
5122	Sound Recording Industries	9.617	0.19	0.248	0.01%
5151	Radio and Television Broadcasting	9.650	0.20	0.256	0.19%
5152	Cable and Other Subscription Programming	9.766	0.17	0.411	0.07%
5171	Wired Telecommunications Carriers	9.746	0.05	0.489	0.64%
5172	Wireless Telecommunications Carriers	9.661	0.02	0.458	0.15%
5174	Satellite Telecommunications	9.873	0.26	0.429	0.01%
5179	Other Telecommunications	9.765	0.16	0.421	0.09%
5182	Data Processing, Hosting, and Related Svcs.	9.806	0.25	0.351	0.28%
5191	Other Information Services	9.875	0.29	0.391	0.29%
5211	Monetary Authorities-Central Bank	10.030	0.43	0.410	0.02%
5221	Depository Credit Intermediation	9.493	0.04	0.256	1.71%
5222	Nondepository Credit Intermediation	9.663	0.12	0.332	0.57%
5223	Activities Related to Credit Intermediation	9.602	0.09	0.306	0.27%
5231	Secur. & Commod. Intermed. & Brokerage	10.270	0.66	0.422	0.48%
5232	Securities and Commodity Exchanges	10.340	0.64	0.511	0.01%
5239	Other Financial Investment Activities	10.150	0.52	0.425	0.41%
5241	Insurance Carriers	9.718	0.14	0.369	1.28%
5242	Agencies, Brokerages, & Other Ins. Activ.	9.553	0.10	0.249	0.89%
5251	Insurance and Employee Benefit Funds	9.619	0.14	0.324	0.03%
5259	Other Investment Pools and Funds	9.964	0.37	0.430	0.02%
5311	Lessors of Real Estate	9.333	-0.06	0.202	0.42%
5312	Offices of Real Estate Agents and Brokers	9.478	0.08	0.196	0.20%
5313	Activities Related to Real Estate	9.390	-0.02	0.213	0.48%
5321	Automotive Equipment Rental and Leasing	9.348	-0.09	0.248	0.14%

Appendix Table A-2 (continued)

NAICS		Mean log earnings	Mean worker effect (normalized to mean 0)	Industry premium	Percent of workforce
		(1)	(2)	(3)	(4)
5322	Consumer Goods Rental	9.199	-0.20	0.200	0.11%
5323	General Rental Centers	9.361	-0.06	0.216	0.03%
5324	Commercial Equipment Rental and Leasing	9.618	0.04	0.360	0.12%
5331	Lessors of Nonfin. Intangibles	9.800	0.30	0.303	0.02%
5411	Legal Services	9.724	0.22	0.309	1.04%
5412	Accting., Tax Prep., Bookkeeping, Payroll	9.648	0.21	0.255	0.76%
5413	Arch., Engin., and Related Services	9.791	0.24	0.353	1.33%
5414	Specialized Design Services	9.570	0.16	0.221	0.10%
5415	Computer Systems Design & Related Svcs.	9.960	0.38	0.375	1.76%
5416	Mgmt., Scientific, & Technical Consulting	9.796	0.27	0.322	1.02%
5417	Scientific Research & Development Svcs.	10.010	0.37	0.446	0.68%
5418	Advertising, PR, & Related Svcs.	9.707	0.26	0.273	0.38%
5419	Other Prof., Scientific, & Technical Svcs.	9.372	0.03	0.147	0.45%
5511	Mgmt. of Companies and Enterprises	9.721	0.18	0.337	2.16%
5611	Office Administrative Services	9.613	0.13	0.280	0.39%
5612	Facilities Support Services	9.364	-0.17	0.334	0.11%
5613	Employment Services	9.193	-0.14	0.140	1.28%
5614	Business Support Services	9.204	-0.18	0.186	0.60%
5615	Travel Arrangement & Reservation Services	9.409	-0.01	0.215	0.17%
5616	Investigation and Security Services	9.074	-0.24	0.135	0.52%
5617	Services to Buildings and Dwellings	9.060	-0.25	0.113	0.98%
5619	Other Support Services	9.385	-0.04	0.224	0.19%
5621	Waste Collection	9.410	-0.14	0.334	0.15%
5622	Waste Treatment and Disposal	9.612	0.01	0.403	0.13%
5629	Remediation & Other Waste Mgmt. Svcs.	9.465	-0.06	0.320	0.11%
6111	Elementary and Secondary Schools	9.358	0.02	0.132	7.22%
6112	Junior Colleges	9.419	0.15	0.075	0.45%
6113	Colleges, Univ., & Professional Schools	9.542	0.20	0.142	2.44%
6114	Bus. Schls. & Computer & Mgmt. Training	9.571	0.14	0.241	0.06%
6115	Technical and Trade Schools	9.359	-0.02	0.187	0.10%
6116	Other Schools and Instruction	9.206	-0.03	0.050	0.13%
6117	Educational Support Services	9.464	0.08	0.186	0.10%
6211	Offices of Physicians	9.519	0.10	0.219	2.15%
6212	Offices of Dentists	9.278	-0.05	0.129	0.70%
6213	Offices of Other Health Practitioners	9.278	-0.04	0.109	0.52%
6214	Outpatient Care Centers	9.438	-0.02	0.249	0.67%
6215	Medical and Diagnostic Laboratories	9.430	-0.04	0.271	0.23%
6216	Home Health Care Services	9.200	-0.15	0.143	0.60%

Appendix Table A-2 (continued)

NAICS		Mean log earnings	Mean worker effect (normalized to mean 0)	Industry premium	Percent of workforce
		(1)	(2)	(3)	(4)
6219	Other Ambulatory Health Care Services	9.311	-0.10	0.219	0.22%
6221	General Medical and Surgical Hospitals	9.487	0.05	0.238	5.08%
6222	Psychiatric and Substance Abuse Hospitals	9.341	-0.10	0.247	0.21%
6223	Specialty (except Psychiatric) Hospitals	9.536	0.08	0.260	0.23%
6231	Nursing Care Facilities	9.074	-0.28	0.163	1.19%
6232	Residential Facilities	9.060	-0.29	0.158	0.49%
6233	Assisted Living Facilities for the Elderly	8.972	-0.33	0.113	0.49%
6239	Other Residential Care Facilities	9.073	-0.26	0.143	0.13%
6241	Individual and Family Services	9.081	-0.24	0.119	0.94%
6242	Comm. Food, Housing, Other Relief Svcs.	9.183	-0.14	0.129	0.11%
6243	Vocational Rehabilitation Services	9.012	-0.29	0.112	0.22%
6244	Child Day Care Services	8.827	-0.38	0.014	0.45%
7111	Performing Arts Companies	9.374	0.06	0.134	0.06%
7112	Spectator Sports	9.656	0.28	0.193	0.07%
7113	Promoters of Perf. Arts, Sports, & Similar	9.482	0.12	0.170	0.05%
7114	Agents/Mgrs. for Artists, Athletes, etc.	9.724	0.35	0.222	0.02%
7115	Indep. Artists, Writers, & Performers	9.576	0.15	0.231	0.03%
7121	Museums, Historical Sites, and Similar	9.323	0.00	0.138	0.15%
7131	Amusement Parks and Arcades	9.137	-0.11	0.060	0.09%
7132	Gambling Industries	9.065	-0.29	0.167	0.22%
7139	Other Amusement & Recreation Industries	9.199	-0.08	0.087	0.48%
7211	Traveler Accommodation	9.072	-0.25	0.128	1.15%
7212	RV Parks and Recreational Camps	9.071	-0.13	0.005	0.02%
7213	Rooming Houses, Dormitories	9.027	-0.27	0.111	0.01%
7223	Special Food Services	9.019	-0.27	0.090	0.29%
7224	Drinking Places (Alcoholic Beverages)	8.885	-0.23	-0.087	0.10%
7225	Restaurants and Other Eating Places	8.896	-0.29	0.000	2.66%
8111	Automotive Repair and Maintenance	9.224	-0.10	0.122	0.60%
8112	Electronic/Precision Equip. Repair & Maint.	9.458	0.01	0.250	0.09%
8113	Commercial Equip. Repair and Maint.	9.494	0.01	0.281	0.17%
8114	Personal & Hhld. Goods Repair and Maint.	9.156	-0.13	0.085	0.04%
8121	Personal Care Services	8.948	-0.23	-0.015	0.29%
8122	Death Care Services	9.312	0.00	0.124	0.07%
8123	Drycleaning and Laundry Services	9.011	-0.33	0.144	0.18%
8129	Other Personal Services	9.034	-0.21	0.059	0.13%
8131	Religious Organizations	9.149	-0.05	0.028	0.09%
8132	Grantmaking and Giving Services	9.577	0.16	0.224	0.12%
8133	Social Advocacy Organizations	9.351	0.00	0.165	0.16%

Appendix Table A-2 (continued)

NAICS		Mean log earnings	Mean worker effect (normalized to mean 0)	Industry premium	Percent of workforce
		(1)	(2)	(3)	(4)
8134	Civic and Social Organizations	9.173	-0.10	0.090	0.13%
8139	Bus., Prof., Labor, Political, and Similar	9.624	0.15	0.285	0.31%
8141	Private Households	9.075	-0.31	0.206	0.07%
9211	Exec., Leg., & Other Govt. Support	9.451	-0.02	0.268	2.66%
9221	Justice, Public Order, and Safety Activities	9.516	-0.01	0.324	1.69%
9231	Admin. of Human Resource Programs	9.428	-0.01	0.243	0.63%
9241	Admin. of Environmental Quality Programs	9.515	0.07	0.247	0.21%
9251	Admin. of Housing, Comm. Devt.	9.433	-0.01	0.251	0.08%
9261	Administration of Economic Programs	9.475	0.04	0.245	0.38%
9281	National Security and International Affairs	9.267	-0.13	0.193	0.02%

Notes: Industry names are abbreviated for space. See notes to Table 2.

Table A-3. Testing the Role of Industry Composition Effects

	β_c^1		β_c^2	
	(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.