Location, Location, Location

David Card
UC Berkeley

Jesse Rothstein
UC Berkeley

Moises Yi
US Census Bureau

August 2021

ABSTRACT

We use longitudinal data from the LEHD to study the causal effect of location on earnings. We specify a model for log earnings that includes worker effects and fixed effects for different commuting zones (CZs) fully interacted with industry, allowing us to capture potential impacts of local specialization. Building on recent work on firm-specific wage setting, we show that a simple additive model provides a good approximation to observed changes in log earnings when people move across CZ's and/or industries, though it takes a couple of quarters for migrants to fully realize the gains of a move. We also show that the earnings premiums for different CZ-industry pairs are nearly separable in industry and CZ, with statistically significant but very small interaction effects. Consistent with recent research from France, Spain and Germany, we find that two thirds of the variation in observed wage premiums for working in different CZs is attributable to skill-based sorting. Using separately estimated models for high and low education workers, we find that the locational premiums for the two groups are very similar. The degree of assortative matching across CZs is much larger for college-educated workers, however, leading to a positive correlation between measured returns to skill and CZ average wages or CZ size that is almost entirely due to sorting on unobserved skills within the college workforce.

*We thank Enrico Moretti, Pat Kline, Alice Wu and seminar participants in the Berkeley Labor Lunch for many helpful comments and suggestions. We are also grateful to Richard Jin for outstanding research assistance. All results in this paper derived from the LEHD have been reviewed to ensure that no confidential information is disclosed (DRB clearance numbers CBDRB-FY21-122, CBDRB-FY21-149, CBDRB-FY21-288). Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau.

There are large, persistent differences in earnings across cities and regions of many countries, including the U.S. (Moretti, 2011). Larger cities tend to have higher average earnings and higher returns to education, but there is also substantial variation among cities of similar size. The source of these differences is a perennial puzzle in economic geography. One explanation is that they reflect the non-random sorting of people to places (e.g., Behrens et al., 2014): High-wage cities are simply the places where people with greater earnings capacity choose to live. A second is that the observed *place effects* derive from the local concentration of productive industries (e.g., Ellison and Glaeser, 1997; Rosenthal and Strange, 2004). A third explanation is that geographic pay differences reflect the causal effects of the places themselves, arising from endogenous factors like population density (Ciccone and Hall, 1996) or the stock of human capital (Duranton and Puga, 2004; Glaeser and Gottlieb, 2009; Moretti, 2011; Diamond, 2016), or from exogenous factors like geography or climate.

The various explanations have sharply different implications for the impacts of inter-regional mobility. A human capital explanation implies that mobility per se has no effect on earnings. In contrast, explanations based on industry composition or place-based productivity imply that workers can increase their earnings by moving. They also have distinct implications for policy. An industry-based explanation implies that regions can benefit from attracting clusters of high-wage industries, providing a rationale for tax breaks or subsidies (e.g., Greenstone, Hornbeck, and Moretti 2010). Likewise, endogenous externalities suggest a role for policies to increase the local population or attract highly skilled workers (Moretti 2004a,b). Externalities arising from fixed factors like climate or geography, on the other hand, leave less room for policy.

Despite several decades of research there is surprisingly little consensus on the relative importance of these three explanations, or even on the fundamental question of whether movers experience systematic wage changes that are correlated with conventionally estimated place effects. A seminal study by Glaeser and Maré (2001) found a mixed pattern of evidence on the effects of moving

into or out of metropolitan areas, depending on the data set and direction of the move.[1] Subsequent studies using much larger administrative data sets (Combes et al., 2008; de la Roca and Puga, 2017; Dauth et al., 2018) have found stronger evidence of place effects in non-U.S. settings. Other studies have argued that the impacts of place vary widely across skill groups (e.g., Gould, 2007).

In this paper we use data from the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program to follow individuals as they move across commuting zones (CZs) and estimate the causal effects of location.[2] We build on recent methodological developments in the estimation of firm-specific pay premiums (e.g., Abowd, Kramarz, Margolis 1999; Card, Heining, and Kline 2013), adapting these to the study of inter-regional wage differences. Generalizing the specification of Combes et al. (2008), we model log wages as a function of permanent individual effects, time-varying person characteristics, and a fully interacted set of CZ × industry effects. This framework allows the wage effect of a given place to vary in an unrestricted fashion across industries, capturing returns to specialization and other forms of non-separability that arise in spillover-type models.

In a first methodological contribution we use event study-style analyses and comparisons of earnings changes for movers in various origin-destination cells to show that worker mobility is approximately exogenous with respect to transitory earnings fluctuations, and with respect to any idiosyncratic gains (or "match effects") from specific CZ's or industries. These findings parallel recent evidence for firm-to-firm mobility (see Card et al. 2018 for a survey) and allow us to obtain unbiased estimates of the CZ × industry effects using a standard two-way fixed effect framework.

Our second contribution is to characterize the role of industry in driving CZ wage differences. We decompose place-by-industry effects into a combination of place effects, industry effects, and an

[1] Appendix Figure 1 presents a visual summary of Glaeser and Maré's main results. They used the Panel Study of Income Dynamics and the National Longitudinal Study of Youth, both of which have relatively small samples in most cities.

[2] CZs are meant to capture regions within which workers commute, and unlike metropolitan statistical areas (MSAs), they cover the entire country, including both urban and rural areas.

interaction (or match) effect. Importantly, we find that places and industries have nearly separable effects on log earnings. This holds true using relatively broad industry categories, or much finer 4-digit categories. This additive structure implies that employers in high-wage places pay a roughly constant local wage premium over and above the national premium (or discount) for their industry, making it straightforward to define meaningful ***CZ effects***.[3] We also quantify the impacts of local concentrations of high-wage industries: We find that industry clustering contributes only a small amount to between-CZ earnings variation.

Our third contribution is to use our estimated earnings models to decompose average earnings premiums across CZ's. We show that about two-thirds of the variation in raw earnings differences across CZ's is attributable to differences in the average person effects (i.e., human capital) of a CZ's workers, while about one third is attributable to CZ earnings premiums. As has been found in France (Combes et al., 2008), Spain (de la Roca and Puga, 2017), and Germany (Dauth et al., 2018), we find that higher earnings-capacity workers are more likely to live in high wage CZ's, an assortative matching pattern that magnifies the inequality in average earnings differences across CZ's.

Our fourth contribution is to explore differences in the effects of different CZ's on the earnings of higher and lower skilled workers, and on the observed return to skill across places. In cross-sectional models, larger and higher-wage cities appear to have higher returns to education, but these models do not account for ability differences. We estimate separate AKM-style models for people with more and less education (at least some college vs. at most high school) and compare the CZ × industry effects for the two skill groups. We find very similar effects for high and low educated workers. Importantly, however, we find much greater assortative matching of higher-skill workers to high-wage places (consistent with Diamond, 2016). This differential sorting based on unobserved skills within the college-

[3] Conceptually the problem of defining a single CZ effect in the presence of CZ-industry interactions is similar to the problem of defining an average treatment effect in the presence of heterogeneous treatment effects.

educated workforce explains nearly all of the higher apparent return to college in larger and higher-wage cities.

Fifth, we summarize the implications of our results by comparing the elasticities of earnings with respect to market size derived from simple cross-sectional models and from our two-way fixed effects models. Larger markets have higher wages. We find that the contribution of skill sorting to the market size elasticity is substantially downward biased in cross-sectional models by the failure to account for unobserved worker skills, which are captured by the person effects in our main models but are treated as part of the residual in standard cross-sectional models. We also find that larger cities have more dispersion in skills (Eeckhout et al., 2014), and a greater degree of within-CZ assortative matching between high skilled workers and high-return industries (Dauth et al., 2018).

Finally, we conclude by examining housing cost differences across CZs. We find that larger and higher-earnings CZs have much higher housing costs than smaller or lower-earnings CZs, enough so to more than completely offset their larger effects on nominal earnings. Thus, movements to larger or to higher earnings locations mean reductions in real income.

Our work is related to three main literatures. The first is a set of studies that, following Glaeser and Mare (2001), use longitudinal data to separate the effects of place from the non-random sorting of workers.[4] Most prominent are recent papers using tax records for France, Spain, and Germany. Combes et al. (2008) estimate earnings models for France that include fixed effects for workers, employment areas (EAs, comparable to commuting zones), and industry. They show that nearly one-half of the variance in mean log wages across EA's is attributable to worker characteristics. De la Roca and Puga estimate models for Spain that include fixed effects for workers, urban areas (UAs), and measures of cumulative work experience in larger UAs. Like Combes et al. (2008), they find that the effect of UA size

[4] A related set of papers use a similar strategy but derive from the intergenerational mobility literature. See, e.g., Chetty and Hendren (2018a,b) and Finkelstein, Gentzkow, and Williams (2016, 2021).

on mean wages is reduced by about one-half when they include worker effects. They also find that past experience in larger UAs yields persistent wage advantages. Finally, Dauth et al. (2018) estimate models for Germany that include fixed effects for workers and *establishments*. They then decompose mean wages in different travel-to-work areas, finding that about 40% of the variance in mean log wages across areas is attributable to the characteristics of workers, while another 40% is attributable to assortative matching of higher-skilled workers to areas with high-paying firms.

Second, we relate to the large literature in urban economics on market size elasticities and the returns to agglomeration (Rosenthal and Strange, 2004; Baum-Snow and Pavan, 2012; Behrens et al., 2014; Eekhout et al., 2014). A related literature considers the impact of high-wage employers or industries on local economic development (Greenstone, Hornbeck, and Moretti 2010; Kline and Moretti 2014) or worker location choices (Diamond, 2016). We contribute to this literature by showing that the U.S. exhibits substantial sorting of higher-skilled workers to larger, higher-wage CZ's, but that, contrary to an important thrust of the existing literature, industry-based agglomerations play a very modest role in explaining local wage variation.

Finally, our work is also related to the literature beginning with Abowd, Kramarz and Margolis (1999) that examines firm-specific pay differences. Studies in this vein focus on *firms* or *plants* as the unit of analysis and, with the notable exception of Dauth et al. (2018), have ignored the role of geography. We show that there is an important geographic component in wage setting that adds a roughly constant pay premium for workers in all industries.

II. Stylized facts about geographic earnings premiums in the U.S.

To set the stage for our analysis, in this section we use cross-sectional data from the American Community Survey (ACS) to establish some stylized facts about differences in earnings across commuting zones (CZ's) in the United States. In subsequent sections we then use LEHD panel data and

more complex methods to probe the sources of these differences. To preview our main findings from the ACS, many of which have been noted in past papers, we show that there is wide variation in mean earnings across commuting zones; that only a small share of this variation is attributable to differences in observed human capital; that larger (more populous) places have only modestly higher observed human capital but tend to have substantially higher earnings; and that college workers' earnings vary more across places than do high school workers, yielding larger observed returns to education in higher-wage places and in larger places.

Our ACS sample is formed by pooling public-use microdata for 2010-2018. We focus on individuals aged 18-62 with at least one year of potential experience. We impute weeks worked last year based on the intervals reported in the ACS, and construct an estimate of annual hours based on this imputation and reported hours per week. We then construct an hourly wage for people with positive annual hours and positive wage and salary earnings for the previous year, Winsorizing at $5 and $500.[5]

We assign local labor markets based on the 1990 CZ definitions developed by Tolbert and Sizer (1996). Commuting zones are intended to approximate integrated labor market areas, with each CZ comprised of one or more complete counties. There are 741 CZs in the United States. The lowest level of geography in the ACS is the public use micro area (PUMA), which is contained within counties in larger urban areas but can include parts of 10 or more counties in sparse areas. We use the fractions of people in each PUMA who lived in each county in the 2000 Census (for the 2010 and 2011 ACS) or the 2010 Census (for the 2012+ ACS) to probabilistically allocate respondents to counties and CZ's. We drop CZ's with small populations for which our allocation procedures find few observations, yielding a total of 688 commuting zones. We note that the size distribution of CZ's is highly skewed: 58.2% of all workers in the ACS are in the largest 50 CZ's, 86.3% are in the largest 200, and 99.7% are in the 688 CZs that we use.

[5] Our main sample has 11.7 million workers, providing relative large samples for even modest-sized CZ's, e.g., around 10,000 observations for the CZ's ranked at roughly 200th in size (e.g., Binghamton, NY; Morgantown, WV; and Byron, TX).

Table 1 presents some CZ-level summary statistics from this sample, weighted by CZ population. We show statistics for workers in all 688 CZ's in columns 1-2 and for the largest 50 CZ's in columns 3-4. To get a sense of how the highlighted characteristics vary across CZ's, in columns 5-7 we show the regression coefficient, standard error, and R-squared from a simple (weighted) regression of the CZ-mean of the characteristic on the log of the weighted count of workers in the CZ, which we use as a measure of "market size" throughout the remainder of the paper. For example, as shown in row 1, the mean of log hourly wages in our ACS sample is 2.863 (about $17.50), and the standard deviation across CZ's is 0.141. The $\pm 2\sigma$ "Lester range" is therefore 56 log points or about 75%. The coefficient of a regression of mean log hourly wages on log size is 0.069 with a standard error of 0.003; log size explains 51.9% of the cross-CZ variance in log wages. Log size is also highly correlated with the local share of workers with at least a BA degree, with the fraction of immigrants in the CZ, and (negatively) with the share of white non-Hispanic workers.

To make some initial progress in documenting how the cross-CZ variation in wages is related to "skills", we estimate a regression model that relates the log hourly wage of worker *i* in commuting zone *c*, y_{ic}, to a set of individual controls X_{ic} (including education, experience, gender, race/ethnicity, and region of origin for immigrants), [6] and CZ fixed effects ψ_c:

$$y_{ic} = \psi_c + X_{ic}\beta + u_{ic} \qquad (1)$$

Since the mean log wage in a CZ is the sum of the estimated CZ effect $(\hat{\psi}_c)$ and the mean predicted wage $(\bar{X}_c\beta)$, the model provides a simple decomposition of the mean into observed factors and unobserved factors. We emphasize that $\hat{\psi}_c$, the CZ wage residual, includes both unobserved skills and the local wage premium. The CZ-level means of these components are shown in Panel b of Table 1. The

[6] We classify immigrants into 4 broad regions of origin: Europe plus Canada, Australia, and New Zealand; Central and South America including Mexico, East and South Asia (including India, China, Vietnam etc.) and the rest of the world. We include these interacted with female gender.

standard deviation of $\hat{\psi}_c$ across CZs is 0.109, about double the standard deviation of the CZ mean predicted wage $\bar{X}_c\beta$.

Figure 1 graphs these two components against the CZ mean log wage. (In the figure we normalize both the observed and unobserved wage components to have mean zero across CZs for comparability). We note the relatively wide range of the x-axis, which extends from about $12.20/hour in the lowest-wage CZ's (e.g., Knox County MO, with a population of under 5,000) to about $24.40/hour in the highest-wage CZ's (San Francisco, San Jose, and Washington DC).

The blue dots in the figure, representing observed wage determinants, slope gently upward, confirming that higher-wage CZ's have workers with higher observed skills. The red dots rise much more quickly, implying that much of the variation in mean wages is due to unobserved factors. The slopes of the two fitted series (estimated using regressions that weight by CZ size) are 0.27 and 0.73, respectively, indicating that observed skill factors account for only about a quarter of the across-CZ variation in mean wages.[7]

Figure 2 explores how observed and unobserved factors vary with CZ labor market size. For visual clarity we drop 19 very small CZs with log sizes under 8 (about 3,000 workers). Mean observed and unobserved wage components are nearly uncorrelated with log size for CZ's with log size < 12 (about 160,000 workers). Among larger CZ's, however, both components rise with size. In weighted models the larger CZ's dominate, leading to an elasticity of the observed component of wages with respect to workforce size of 0.013 and an elasticity of the unobserved component of wages of 0.055 (see Panel b in Table 1). Thus, about one-quarter of the overall 0.068 elasticity of earnings with respect to CZ size is attributable to observed worker characteristics and about three-quarters to unobserved factors.

[7] The two slopes add to 1, and since $var[\bar{y}_c] = cov[\bar{y}_c, \bar{X}_c\beta + \hat{\psi}_c]$ they can be interpreted as shares of the variance of $\bar{y}_c$ that is attributable to observed and unobserved factors, respectively.

Figure 3 explores how the return to college varies across CZs. We fit equation (1) separately for ACS respondents with high school diplomas or less and for those with bachelor's degrees or more, and plot the two sets of CZ fixed effects $\hat{\psi}_c$ against the mean wage in the CZ. Here, we see that the unobserved component of wages rises faster across CZ's for college graduates than for high school workers, implying larger returns to education in high-wage places. (Both series are normalized to mean zero, so while the figure shows how the return to education varies across CZs, the level of the return cannot be inferred.)

Appendix Figure 2 shows the relationship of observed and unobserved wage components with the mean fraction of college-educated (BA+) workers in each CZ. Both wage components are strongly positively correlated with the college share. While the correlation with the observed component is largely mechanical (since college education is powerful predictor of wages), the correlation with the unobserved component is not. Instead, it reflects the fact that locational wage *premiums* in the U.S. are very strongly correlated with the local share of college-educated workers (see Moretti 2004a,b).

A large literature has developed around the importance of local industry structure in determining the relative success of different places (e.g., Ellison and Glaeser, 1997). To get a sense of the importance of industry structure for wages, we fit an alternative version of equation (1) including dummies for 20 major industries.[8] We then use this alternative model to redefine predicted skills to include any component associated with industry. As shown in Panel c of Table 1 this alternative skill measure has a slightly larger variation across CZ's than the one without industry controls (std. dev. = 0.054 vs. 0.050) and is slightly more positively correlated with log size (regression coefficient = 0.015 vs. 0.013). We also extracted just the industry component of this skill index, creating a "mean industry composition effect" for each CZ. This has a rather small standard deviation across CZ's (0.012) and only a

[8] There are 24 major industries in the NAICS, identified by the first 2 digits of the code. For our ACS analysis we pool the three subgroups of manufacturing into one sector and combine the Management of Companies and Industries sector (NAICS 7570) with Professional, Scientific and Technical industries.

slight relationship with log size (coefficient = 0.002), providing initial evidence that variation in industry structure is not a large factor in CZ-average wages. Panel d of Table 1 repeats the exercise, this time using much more detailed "4-digit" industries, of which there are 295 in the ACS This only slightly increases the variation in the skill index across CZs and its association with CZ size: The standard deviation of this detailed skill index across CZs (shown in the table) is 0.017; its regression coefficient on CZ size is 0.004.

A key limitation of our ACS analyses is that our skill measure reflects only observed characteristics. Some of the variation in the "unobserved" component wages surely reflects differences in skills that are not captured in X_{ic} but are observed by employers and fully rewarded in the labor market. To explore this, we turn to LEHD panel data that allow us to follow workers as they move across CZs and separate permanent earnings differences among workers from causal place effects. We start with a brief overview of the LEHD data before turning to the modeling framework that we will use in the remainder of the paper.

III. Longitudinal Earnings Data from LEHD

Longitudinal Employer-Household Dynamics (LEHD) data are derived from quarterly earnings reports provided by employers to state unemployment insurance (UI) agencies, which are then assembled by the Census Bureau into a national data set. (In early years the LEHD had less than national coverage, but during the period we study it included all states.) The core data set includes total wages paid by a given employer to a given worker in a quarter and a few characteristics of each worker and establishment, including location and industry.[9] The underlying UI tax system covers about 95% of

[9] Technically an "employer" in the LEHD is an establishment (identified by a firm's State Employer Identification Number, SEIN, and a firm's reporting unit number, SEINUNIT). Many worker characteristics in the LEHD are derived from the Census Bureau's Person Characteristics File – see Abowd et al. (2009).

private sector employment in the U.S., as well as state and local government employees, but excludes federal government employees, members of the armed services, and self-employed workers.

To construct our sample, we begin with the universe of all person-employer-quarters (PEQ's) for workers age 22-62 in the LEHD from 2010Q1 to 2018Q2. We exclude PEQ's with earnings less than $3,800 (roughly the earnings from a full-time job at the federal minimum wage), and those for workers with multiple employers in the quarter. To ensure we capture only full-quarter earnings, we also exclude observations in transitional quarters – the first quarter of a "job spell" of continuing PEQ's for the same establishment and worker, where the employment relationship may have begun midway through the quarter, and the last quarter of such a spell, where the relationship may have ended mid-quarter. We also drop workers who are observed for less than 8 quarters in the 2010Q1-2018Q2 period, and those who have a PEQ with unknown industry and/or location of the establishment. We then divide the resulting sample into 30 mutually exclusive 3.33% subsamples. We pool the first 3 of these as our main estimation sample, which we refer to as the *A sample*. We pool the next 3 to form the *B sample*, which we use together with the A sample to estimate measurement-error-corrected second moments. In some exercises below we use one or more of the remaining subsamples to form instrumental variables.

We use the industry code and location of the establishment in a PEQ to assign workers to an industry and CZ in each quarter. We use 735 of the 741 CZ's defined by Tolbert and Sizer (1996), omitting a handful of very small CZs. [10] In cases where we combine information from the LEHD and the ACS, we restrict our analysis to the 688 CZs that can be identified in the ACS. The correlation between CZ-level mean log quarterly earnings estimated from the LEHD and mean log hourly wages estimated from our 2010-2018 ACS sample is 0.94.[11]

[10] Specifically, we exclude six CZs that are not in the largest connected set in one or more of the 33 random 3 1/3% samples.

[11] This is a weighed correlation, using the number of workers in our ACS sample as a weight. The unweighted correlation is 0.73.

For our primary analyses, we use 24 "2-digit" industries based on the first two digits of the NAICS code.[12] This results in roughly 18,000 CZ-industry cells. Since many CZ's are small and some industries are also relatively small, not all CZ-industry cells are populated. In some analyses, we limit attention to the (roughly) 300 CZ's that have workers in each industry. We also explore some specifications using more detailed 4-digit NAICS industries, with 312 unique codes. For these analyses, we limit attention to the 50 largest CZs.

For our event study analysis, we define a move as occurring when a worker switches from one CZ-industry cell to another; in some analyses, we only consider moves across CZs, dropping moves between industries in a single CZ. We allow for up to six quarters of non-employment between the last one used in the original cell and the first one used in the new cell.[13] For tractability purposes, we further restrict our sample to workers who are observed switching CZ-industry cells once, and who had stable jobs in the same CZ-industry cell for at least 5 consecutive quarters before and after the switch.

There are some limitations in the LEHD data. As is true with most linked employer-employee administrative data sets, it is not possible to distinguish between unemployment and non-participation. We use only quarters in which an individual is employed. We also observe only quarterly earnings, not the number of hours worked within the quarter nor even the start and end date of an employment spell. Our exclusion of observations with quarterly earnings below $3,800, of quarters with multiple jobs, and of transitional quarters should eliminate many but not all part-time and partial quarter employment spells.

Table 2 presents summary statistics for our analytical samples. We show some basic worker characteristics, the fractions observed in 1, 2 or 3+ CZ's in the 8-year period, the fraction observed in 1,

[12] Under this classification, construction comprises 1 industry, manufacturing comprises 3 industries, and hotels, restaurants and cultural/recreation facilities comprise 1 industry.

[13] Transitional quarters are considered non-employment when computing the transitional gap. Thus, we allow workers to be out of the labor force or unemployed for up to four consecutive quarters between the last appearance in the old cell and the first one in the new cell.

2 or 3+ industries, and the mean estimated person effect from our models discussed below, which represents a simple summary of overall worker "skill" (on a log scale). The first column presents results for the full sample, while column 2 shows our estimation sample, limited on age and quarterly earnings. Workers in the estimation sample have similar mean earnings, age, fraction female and fraction foreign born as the broader LEHD population but are somewhat less mobile across CZ's and industries. Columns 3-6 divide our estimation sample based on whether people are observed in multiple CZs and/or industries. A key difference that emerges here is that people who change CZ's but stay in the same industry are positively selected, while those who stay in the same CZ but change industry are negatively selected.

Column 7-10 pertain to our event study sample, in which people have exactly one move between CZ-industry cells in the data. This move can involve only a change in industries (column 8), only a change of CZ's (column 9) or a change in both (column 10). Again we see that industry-changers who stay in the same CZ have lower mean earnings and estimated person effects, while those who change CZ but stay in the same industry have higher mean earnings and higher person effects.

An initial look at the impacts of mobility

Using the event study sample described in Table 2 we begin with an initial descriptive event study of the earnings changes associated with moves between CZ's with higher and lower average wages. We construct an adjusted earnings measure by regressing quarterly earnings on time effects and a polynomial in age, and taking the residual. Then, following Card, Heining, and Kline (2013) (hereafter CHK), we define origin and destination groups by dividing CZs into four quartiles based on average earnings, yielding 16 origin and destination groups. Figure 4 plots the means of adjusted wages by

quarter relative to the move for 8 of these groups, restricting attention to movers leaving CZs in the top and bottom quartiles (2 origin quartiles × 4 destination quartiles = 8 groups).[14]

Even these simple means show some well-behaved patterns. In particular, earnings are quite flat before a move, with no sign that workers move after experiencing downward or upward shocks. Earnings are not as flat following moves; they tend to grow between quarter +1 and quarter +2, suggesting some changing features early in a new job spell (e.g., ramping up of hours or a probationary wage period).[15] This growth is of similar magnitude across all origin-destination combinations, and in many groups it continues after quarter +2, perhaps indicating workers climbing a new job ladder. The change in earnings from the origin to destination is consistent with a causal effect of CZs: Those who move to higher-wage CZs tend to see earnings increases, while those who move to lower-wage CZs tend to see declines. However, the identity of the destination CZ also predicts pre-move earnings for workers from the same origin CZ's (and vice versa): Those from origin quartile 1 who will move to quartile 4 earn more before the move than those who will move to low-quartile CZs. A simple explanation for this pattern is sorting based on individual heterogeneity, with higher-skill workers from the same origin CZ's tending to move to higher-pay CZs.

IV. Methods

We develop our methodology in four stages, corresponding to the main steps of our analysis. First, we present our two-way fixed effects model and discuss some important specification concerns. Second, we present our approach for analyzing the role of industry differences in CZ-average wages.

[14] Recall that we exclude transitional quarters and allow for several quarters of non-employment between them. Thus, there may be as many as 6 quarters between the last observation in the origin CZ (labeled -1) and the first one in the destination (labeled +1).

[15] Glaeser and Maré (2001) emphasized the possibility of an adjustment process for movers. As can be seen in Appendix Figure 1, their NLSY sample shows that it takes 1 year to get to the new level of earnings for people who move into a metro area. In the LEHD data it seems that the adjustment process is at most 2 quarters long.

Next, we discuss our methods for assessing the additivity assumption of our model. Finally, we summarize our approach for examining city size effects.

Two-way fixed effects model

Extending the approaches of Combes et al. (2008) and de le Roca and Puga (2017), we adopt the two-way fixed effects model of Abowd, Kramarz, and Margolis (1999) to the locational context by assuming that log earnings can be summarized by an additive model with worker fixed effects, time-varying worker characteristics, and a set of fully interacted CZ-by-industry effects. Specifically, letting y_{it} represent the log of observed earnings of worker i in quarter t, and letting $cj(i,t)$ represent the CZ-by-industry cell, we consider models of the class:

$$y_{it} = \alpha_i + \psi_{cj(i,t)} + X_{it}\beta + \epsilon_{it}. \quad (2)$$

Here, α_i is a fixed effect for worker i that captures factors like education as well as unmeasured skills like motivation that are rewarded in the labor market, X_{it} is a vector of time-varying characteristics (age and calendar time effects), and $\psi_{cj(i,t)}$ is an additive wage premium or discount for jobs in the CZ and industry combination $cj(i,t)$.[16] The error term ϵ_{it} captures all other factors, including transitory worker-specific earnings shocks, transitory industry- or CZ-specific shocks, and any person-specific match effect associated with working in the specific CZ and industry combination.

The key assumption needed to ensure that ordinary least squares (OLS) applied to equation (2) yields unbiased estimates of the CZ-industry premiums is that the sequence of $\epsilon_{it}{}'s$ is orthogonal to the sequence of CZ-industry choices made by worker i -- a so-called "exogenous mobility" assumption. As noted in the worker-firm context by CHK, there are two main threats to this assumption. The first is that mobility is correlated with transitory earnings shocks – as could happen if workers who experience negative (positive) shocks to the value of their skills tend to move to lower- (higher-) paying CZ's. Such

[16]The person effects and CZ-industry effects have to be normalized. For expositional convenience it is useful to assume that the CZ-industry-effect for a low-wage industry in a small rural CZ is 0, so $\psi_{cj(i,t)}$ represents the premium relative to that.

threats would be revealed in Figure 4 by "Ashenfelter dips" (or blips) prior to a move (Ashenfelter, 1978). Importantly, we see little evidence of such patterns.

A second key threat is that mobility across industries and/or CZ's is driven in part by idiosyncratic match effects – as is often assumed in Roy (1951)-style sectoral choice models. Under exogenous mobility, model (2) implies that the expected change in earnings for a worker who moves from CZ/industry pair (c, j) to CZ/industry pair (d, k) is $\psi_{d,k} - \psi_{c,j}$, while the expected change in earnings for a worker who moves in the opposite direction is $\psi_{c,j} - \psi_{d,k}$ -- i.e., equal in magnitude but opposite in sign. In contrast, if there are idiosyncratic match effects that partly drive mobility, then this symmetry prediction will fail. It is even possible (and is actually assumed in some Roy type models) that movers in both directions experience earnings gains. We present more analysis of symmetry in gains and losses below. However, we can already see some support in the patterns in Figure 4. In particular, we see that the gains for movers from quartile 1 CZ's to quartile 4 CZ's are (roughly) equal and opposite to the losses for movers from quartile 4 CZ's to quartile 1 CZ's.

Equation (2) provides a simple framework for decomposing the variance of individual earnings into components attributable to the person fixed effects, the time-varying person characteristics, the CZ-industry premiums, and their covariances. Specifically (omitting subscripts):

$$V(y) = V(\alpha) + V(\psi) + V(X\beta) + 2cov(\alpha, \psi) + 2cov(\alpha, X\beta) + 2cov(\psi, X\beta) + V(\epsilon) \qquad (3)$$

While the terms on the right-hand side of (3) can be estimated by a simple "plug-in" procedure, past studies of firm wage setting have emphasized that estimation errors in α and ψ will lead to biases (see e.g., Andrews et al. 2008; Kline et al., 2020). Our sample contains approximately 18,000 city-industry combinations, many fewer than the number of firms in a typical AKM implementation. As a result, our parameter estimates are relatively precise, suggesting that sampling error biases will be small.

Nevertheless, to address any concerns we use covariances between our two non-overlapping 10% subsamples, A and B, to-recover unbiased estimates of the variances and covariances in (3).[17] Specifically, we estimate equation (2) on subsample A, obtaining estimates $\{\hat{\alpha}_i^A, \hat{\beta}^A, \hat{\psi}_{cj}^A\}$, and again on subsample B, obtaining estimates $\{\hat{\alpha}_i^B, \hat{\beta}^B, \hat{\psi}_{cj}^B\}$. The estimates of $\hat{\beta}$ and $\hat{\psi}$ can be compared directly across samples, but the person effects pertain to different individuals in the two samples. Thus, we construct new estimates $\hat{\alpha}_i^{A(B)}$ of the person effects for the workers in subsample A using the parameter estimates $\hat{\beta}^B$ and $\hat{\psi}_{cj(i,t)}^B$ from sample B:

$$\hat{\alpha}_i^{A(B)} = \frac{1}{T_i}\sum_t y_{it} - \hat{\psi}_{cj(i,t)}^B - X_{it}\hat{\beta}^B ,$$

where T_i is the number of observations for person *i.* We then use cross-products of combinations of the parameters estimated in the two subsamples to estimate the terms in (3):

$$\begin{gathered} \widehat{V(\alpha)} = cov\left(\hat{\alpha}^A, \hat{\alpha}^{A(B)}\right);\ \widehat{V(\psi)} = cov\left(\hat{\psi}^A, \hat{\psi}^B\right) \\ \widehat{V(X\beta)} = cov\left(X\hat{\beta}^A, X\hat{\beta}^B\right);\ \widehat{cov(\alpha,\psi)} = cov\left(\hat{\alpha}^A, \hat{\psi}^B\right) \\ \widehat{cov(\alpha,X\beta)} = cov\left(\hat{\alpha}^A, X\hat{\beta}^B\right);\ \widehat{cov(\psi,X\beta)} = cov\left(\hat{\psi}^A, X\hat{\beta}^B\right) \end{gathered} \tag{4}$$

These provide unbiased estimates of the terms in the decomposition (3).

Specification issues

A key assumption of equation (2) – and one that has raised concerns in the worker-firm context – is that log earnings are additively separable in person effects and CZ-industry effects. Under this assumption, workers who move from one CZ to another and stay in the same industry should experience the same earnings gains (or losses) regardless of their skill levels. We take several approaches to evaluate this.

[17] Kline, Saggio and Sølvsten (2020) present a method for estimating unbiased quadratic forms of the parameters of a high-dimensional model when it is not possible to use separate subsamples, as we do here.

First, we examine the residuals from equation (2), looking for evidence that the mean residuals for high or low skilled workers (classified by the estimated value of α_i) are larger in magnitude for jobs in high or low premium CZ-industry cells (classified by the estimated value of ψ_{cj}).

Second, we examine average wage changes for workers moving from one CZ-industry cell to another, and compare these with the prediction from the model. As noted, the mean predicted change in earnings for a worker who moves from CZ/industry pair (c, j) to CZ/industry pair (d, k) is $\Delta_{dk,cj} = \psi_{d,k} - \psi_{c,j}$. We construct the average observed change for all movers, and for movers with relatively high or low values of their estimated $\alpha's$, and compare these to the predicted changes from the model, $\widehat{\Delta}_{dk,cj} = \hat{\psi}_{dk} - \hat{\psi}_{cj}$. We note that sampling errors on the $\widehat{\Delta}'s$ may bias these comparisons. To sidestep this problem, we return to our split sample procedure, instrumenting $\widehat{\Delta}$ in sample A with an estimate based on the mean estimate of $\widehat{\Delta}$ from models fit to the 27 3.33% samples not included in Sample A.

Third, as discussed below, we estimate model (2) separately for more- and less-educated workers, using subsets of the LEHD data set that have education information from linked ACS responses. We verify that the two estimates of ψ_{cj} are very highly correlated across education samples.

Decomposing city and industry effects

Once we establish that the two-way fixed effects model (1) provides an adequate description of the data, we then move on to investigate the structure of the CZ-industry effects. We are primarily interested in the across-CZ variation in ψ_{cj}, and in the degree to which this variation is driven by industry effects. We try to answer two types of key questions: (1) Is the wage gap between New York City (NYC) and New Orleans primarily due to the presence of a larger finance sector in NYC? (2) Is the wage gap for workers in finance particularly high in NYC, or is there a NYC wage advantage that is widely shared across industries?

We present two sets of analyses below to explore the role importance of industries. First, we investigate the separation of ψ_{jc} into pure CZ effects, pure industry effects, and potential "match" effects. We begin with a simple regression of $\hat{\psi}_{cj}$ on CZ dummies and industry dummies:

$$\hat{\psi}_{cj} = \theta_c + \mu_j + e_{cj} \quad (7)$$

The R-squared of this model provides an indication of how large are the deviations from a simple additively separable specification without match effects. To derive a more formal test, we estimate (7) using estimates of $\hat{\psi}_{cj}$ from our B sample. We then take the residuals from this model and include them as an additional regressors in an expanded model for $\hat{\psi}_{cj}$ from our A sample. We use the mean squared error (MSE) from this expanded model relative to the MSE from a model that excludes the residuals from the B sample to estimate the size of the match effects in the true CZ-industry premiums. We also estimate specifications that include the share of CZ c's employment that is in industry j, s_{cj}, as a regressor in (7). The coefficient on this variable measures the extent to which match effects are related to industry concentration.

To explore the potential relationship between industry employment share and match effects, we use a Oaxaca-style decomposition (Oaxaca 1973; Blinder 1973) of the CZ averages of ψ_{cj}. The average wage premium earned by workers in CZ *c* relative to jobs in the omitted industry and CZ is

$$\Psi_c \equiv \sum_j s_{cj}\psi_{cj} \,. \quad (5)$$

Let w_c represent the share of total national employment in CZ c, and let $\bar{s}_j$ represent the share of national employment in industry *j*, $\bar{s}_j = \sum_c w_c s_{cj}$, with $\bar{\psi}_j$ defined analogously as $\bar{\psi}_j = \sum_c w_c \psi_{cj}$.[18] Then, the city average wage premium can be decomposed as:

[18] Note that we use a fixed set of city weights to construct the national average wage premiums for each industry, rather than using industry-specific city weights. This choice simplifies the interpretation of the interaction term in our decomposition. Recall that we limit our analysis to CZs with employment in each industry, so these fixed-weight averages are well defined.

$$\Psi_c = \sum_j \bar{s}_j \bar{\psi}_j + \sum_j \bar{s}_j (\psi_{cj} - \bar{\psi}_j) + \sum_j (s_{jc} - \bar{s}_j)\bar{\psi}_j + \sum_j (s_{jc} - \bar{s}_j)(\psi_{jc} - \bar{\psi}_j). \quad (6)$$

The first term in (6) represents the average industry premium earned by workers in a representative CZ with national-average industry shares, and is constant across CZ's. The second term is a share-weighted average of the gap between the industry premium earned in CZ *c* and the corresponding national average industry premium, $(\psi_{cj} - \bar{\psi}_j)$. This measures the average locational wage premium for the CZ. The third term represents the excess share of workers in different industries $(s_{jc} - \bar{s}_j)$, weighted by the national average wage premium in each industry. This measures the compositional effect arising from the non-random allocation of industries to places. Finally, the fourth term can be interpreted as a *return to clustering* effect arising from a correlation between the excess share of an industry in a given CZ and any excess wage premium earned by workers in that industry.

To illustrate, consider again the example of New York and the role of the finance sector. Term 2 would be larger if the NYC pay premium for finance is higher than the average finance premium in all CZ's. Term 3 would be larger if NYC has a relatively large share of workers in finance, which is highly paid everywhere. Finally, term 4 would be larger if NYC workers are more likely to work in finance than workers in other cities with smaller local wage premiums for the finance sector.

We calculate the variance of the average city wage premium Ψ_c across CZs and the relative contributions of the 3 key terms in (6). We note that the fourth term will be close to zero if there are no match effects (so $\psi_{ic} = \theta_c + \mu_j$) or if the excess share of a CZ's workers in a given industry is uncorrelated with the match effect. Thus, the magnitude of the fourth term allows us to assess the importance of local industry specialization that is correlated with local pay premiums, while the third term captures local specialization in industries that tend to pay high wages in all CZ's.

Geographic variation in the return to education

Next, we examine differences in the returns to education across places. The starting point for this analysis is the observation (documented in Figure 3) that the gap in wages between more- and less-educated workers tends to be bigger in cities with higher average wages (or in larger cities). Equation (2) allows us to disentangle the sources of the pattern: Is it due to the fact that CZ-specific wage premiums are different for higher- and lower-educated workers, and the premium for highly educated workers is more sensitive to local characteristics? Is it attributable to differences in the industry structure of cities with more highly educated workers? Or is it due to differences in the unobserved skills of more- and less-educated workers in different cities, which can be captured by the person effects in an AKM-style framework but remain unobserved in simple cross-sectional wage models?

We begin by estimating (2) separately for high- and low-education workers, defining the former as those with at least some college education, and the latter as those with only high school or less. Denoting education by e, the AKM model for education group e (e=H, L) is:

$$y_{iet} = \alpha_i + \psi^e_{cj(i,t)} + X_{it}\beta^e + \epsilon_{iet}. \quad (8)$$

The estimated CZ-industry effects $\psi^e_{cj(i,t)}$ are highly correlated across education levels. We construct the averages of α, *X*, and *y* for each education group in each CZ-industry, $\bar{\alpha}_{ce}$, $\bar{X}_{ce}$, and $\bar{y}_{ce}$. Last, we define the share of group-*e* workers in CZ c who work in industry j as s_{cje} (with $\sum_j s_{cje} = 1$ for each c and e).

We are interested in the components of the education gap in CZ c, $\bar{y}_{cH} - \bar{y}_{cL}$. This can be written as the sum of three components:

$$\bar{y}_{cH} - \bar{y}_{cL} = (\bar{\alpha}_{cH} - \bar{\alpha}_{cL}) + (\bar{X}_{cH}\beta^H - \bar{X}_{cL}\beta^L) + \sum_j \left(s_{jcH}\psi_{jcH} - s_{jcL}\psi_{jcL}\right). \quad (9)$$

The two key terms are the first, reflecting the gap in average human capital between high- and low-education workers in the CZ, and the third, reflecting differences in CZ-industry effects.

We again use an Oaxaca-style decomposition to further decompose the third term:

$$\sum_j \left(s_{jcH}\psi_{jcH} - s_{jcL}\psi_{jcL}\right)$$
$$= \sum_j s_{jc}\left(\psi_{jcH} - \psi_{jcL}\right) + \sum_j \left(s_{jcH} - s_{jcL}\right)\psi_{jc}$$
$$+ \sum_j \left(s_{jcH} - s_{jc}\right)\left(\psi_{jcH} - \psi_{jc}\right) - \left(s_{jcL} - s_{jc}\right)\left(\psi_{jcL} - \psi_{jc}\right). \quad (10)$$

Where ψ_{jc} is the premium for a model estimated pooling the two education groups. Here, the first term reflects an average return to skill in the CZ; the second reflects the relative shares of the two groups in higher premium industries; and the third is a variant of the interaction effect in equation (6) with components reflecting the *relative* clustering of high- versus low-education workers in industries with a higher or lower local industry premium for that group.

Combining (9) and (10) yields a decomposition of the return to skill in CZ c into five components, one of which we expect to be quite small. We use this for two purposes. First, we decompose the across-CZ variation in $\bar{y}_{cH} - \bar{y}_{cL}$ into the variances of each of the five components (plus covariances). Second, we explore which of these components account for the association of the CZ return to skill with the CZ average wage level seen in Figure 3, by regressing each component separately on CZ mean wages. We also repeat this exercise for the association with CZ size. Insofar as the wage gap in a CZ reflects true differences in the return to skill, we expect the first term in (10) to play an important role. To the extent that the wage gap is driven by differential assortativeness of high- and low-education workers to different places, however, we expect the first term in (9) to play a predominant role.

Decomposing the Effects of CZ Size

As a final exercise we follow a long tradition in urban economics and explore the sources of the tendency for average wages to rise with city size.[19] In our model mean wages in a CZ decompose as

$$\bar{y}_c = \bar{\alpha}_c + \Psi_c + \bar{X}_c\beta \quad (11)$$

[19] For example, Behrens, Duranton and Robert-Nicoud (2014) develop a model in which this tendency is driven by a combination of worker selection, firm selection, and externalities. An alternative perspective is developed by Eeckhout et al. (2014), who argue that mean skill is approximately constant across cities.

We can further use equation (6) to decompose Ψ_c into a constant plus three terms that vary across CZ's:

$$\bar{y}_c = \bar{\alpha}_c + L_c + C_c + I_c + \bar{X}_c\beta \qquad (12)$$

where $L_c = \sum_j \bar{s}_j(\psi_{cj} - \bar{\psi}_j)$ is the average locational wage premium for CZ c, $C_c = \sum_j(s_{jc} - \bar{s}_j)\bar{\psi}_j$ is is the compositional effect for CZ c, and I_c is is the final interaction term in (6). This means that if we fit a simple regression model like

$$\bar{y}_c = \delta_0 + \delta_1 \log(Size_c) + \xi_c \,,$$

the coefficient δ_1 can be decomposed into the sum of partial correlations of log size with (i) the mean worker effect in the CZ; (ii) the average locational wage premium for the CZ; (iii) the compositional effect for the CZ, and (iv) the interaction effect. We present estimates of these components in our final set of tables. As mentioned, we also decompose the effect of log size on the mean wage gap between high- and low-education workers using a parallel approach.

V. Results

The basic decomposition

We applied the two-way fixed effects model (2) to our main analytic sample (sample A) and to the alternative sample (sample B). Table 3 presents the terms of the implied variance decomposition given by equation (3). We estimate these terms using the cross-sample technique described in equation (4). Simple plug-in estimates, however, are nearly identical, as shown in Appendix Table 1.

Column 1 of Table 3 presents the standard deviations of the estimated person effects, the estimated CZ-industry effects, the estimated covariate index $X_{it}\hat{\beta}$, and the residual terms, at the person-quarter level. We also show the correlations among the 3 terms. Column 2 of the table shows the implied variance shares of the components. In our analytical samples the standard deviation of log quarterly earnings is 0.654, and the associated variance is 0.428. About 74% of this variance is attributable to the person effects, 2.2% to the CZ-industry effects, and 5.3% to the time-varying

covariates (age and time effects). Another 5.4% is attributed to the positive covariance between the person effects and the CZ-industry effects, while the covariances involving $X_{it}\hat{\beta}$ are negligible, as expected. Finally, about 14% of the overall variance in log quarterly earnings is unexplained (implying an R-squared coefficient for the model as whole of about 86%).

The variance shares from our two-way fixed effects model are not too different from the shares that have been estimated in previous two-way fixed effects models of worker and firm pay components (e.g., CHK), though the estimated contribution of the CZ-industry effects in our application is a little lower than is typically estimated for the firm effects in the worker-firm literature, while the residual component is larger. This is not too surprising given that the identity of the specific employer in a specific industry and location can matter for pay-setting (e.g., between larger and smaller hotels in the same city), and such firm-specific components are rolled into the residual in our model.

Columns 3-6 of Table 3 explore the same decomposition, but applying it to CZ-by-industry mean earnings (columns 3-4) and to CZ mean earnings (columns 5-6). The importance of individual effects falls when the data are aggregated, but they still account for 53% of the CZ-industry variance of earnings and 47% of the across-CZ variance. In contrast the variance contribution of the estimated CZ-industry effects rises from 2.2% at the individual level to 12.5% at the CZ-industry mean level and to 19.6% at the CZ-mean level. Even more remarkably, the share of the variance attributable to the covariance between person effects and CZ-industry effects rises from 5.4% at the individual level to 30.8% at the CZ-industry mean level and to 34.7% at the CZ-mean level, implying strong sorting of high-person-effect workers to high-Ψ_c CZs. The variance shares of CZ average wages in column 6 are not too different from the variance shares of local labor market mean wages reported by Dauth et al (2018) for West Germany in

the 2008-2014 period, though Dauth et al. start with an AKM model with worker and establishment effects and aggregate that model to the labor market area.[20]

To help visualize the implications of our model, Figure 5 presents two simple maps. The upper panel highlights CZ's with mean log wages in each of three terciles, as estimated from our ACS sample.[21] The lower panel shows the classification of the same CZ's into terciles of the estimated average "CZ effect" – i.e., $\widehat{\Psi}_c$. For reference, Appendix Table 2 shows the characteristics of CZ's in each tercile, as well as those with very high (top 10) and very low (bottom 10) estimated CZ effects.[22]

The maps illustrate two key points. First, as expected, high wage CZ's tend to be on the coasts, though there are also some high wage CZ's in more rural but resource-rich areas (like Western North Dakota). Second, CZ's in the top and bottom terciles of average wages tend to have a CZ effect in the same range. This reflects both the mechanical fact that a higher CZ effect will raise average wages, and the fact that CZ's with a higher average pay premium tend to attract workers with higher person effects. In fact, the estimates in column 5 of Table 3 imply that a CZ with a 1 percentage point higher CZ effect has average wages that are 1.88 percentage points higher, since (ignoring the covariates):

$$\frac{\partial E[\bar{y}_c|\Psi_c]}{\partial \Psi_c} = 1 + \rho(\bar{\alpha}_c, \Psi_c)\frac{\sigma(\bar{\alpha}_c)}{\sigma(\Psi_c)} = 1 + 0.56 \times \frac{0.100}{0.064} = 1.88$$

There are nevertheless some interesting exceptions to this pattern. Florida has mostly middle or upper tercile average wages but bottom tercile CZ effects. There are also some resource-intensive CZ's (on the

[20] Dauth et al. (2018, Table 2, column 3) report that the shares of the variance of market-average wages explained by person effects, establishment effects, and the covariance of person and establishment effects are 39.8%, 23.6% and 41.7%, respectively.

[21] For disclosure reasons we cannot report classifications based on estimates from the LEHD for small CZ's. The maps show only the largest third of CZs, containing nearly 90% of the population.

[22] The top 10 CZs are Anchorage, AK; Bakersfield, CA; Bismarck, ND; Hobbs, NM; Minot, ND; New York, NY; Odessa, TX; San Francisco, CA; San Jose, CA; and Seattle, WA. The bottom 10 CZs are Columbia, MO; Florence, SC; Gainesville, FL; Hattiesburg, MS; Monett, MO; Ocala, FL; Springfield, MO; St. George, UT; Traverse, MI; and West Plains, MO.

Texas Gulf Coast and in the Permian Basin, for example) that have relatively high CZ effects but only average earnings levels.

Table 4 provides a further summary of the role of variation in worker skills and CZ-industry effects at the CZ level. In columns 1 and 2 of the table we regress the mean of the estimated person effects and the mean of the estimated CZ-industry effects for all PEQ's in CZ c on mean log earnings in the CZ. The coefficients, 0.62 and 0.38, respectively, indicate that over 60% of the variation in mean CZ earnings is attributable to differences in worker characteristics, α, while something less than 40% is attributable to local wage premiums, ψ. In columns 3 and 4 we report the slope coefficients from Figure 1, based on the decomposition of log wages in the ACS into observable and unobservable factors. Observed skills (captured by education, experience, and place of origin for immigrants) account for only 27% of the variation in mean CZ wages, yielding a substantially downward-biased estimate of the importance of worker differences across CZ's relative to the contribution of the person effects from an AKM-style model.

Validating the Specification

As noted above, for OLS to provide unbiased estimates of the coefficients in our two-way fixed effects model we need exogenous mobility (EM) – moves across CZ's and industries have to be uncorrelated with the error term ϵ_{it} in (2). CHK presented several tests for EM in their analysis of workers and firms in Germany and found although EM can be rejected, departures from the patterns predicted by EM are small. In this section we apply some of the same tests and reach essentially the same conclusion.

Figure 6 presents average earnings by quarter relative to the date of a move, separately for those who move across different categories of CZs. It closely resembles Figure 4, but categorizes CZs based on the wage premium $\bar{\psi}_c$ rather than unadjusted mean earnings. We limit attention to the event

study sample described in Table 2, excluding movers who change industry but stay in the same CZs.[23] As in Figure 4, we show only the earnings profiles for movers from the top and bottom quartiles, 4 and 1, respectively.

Under exogenous mobility, the expected change in earnings from period -1 to period 1 for a person who moves from a CZ in quartile q_k to a CZ in quartile q_h is $E[\bar{\psi}_c|c \in q_h]$-$E[\bar{\psi}_c|c \in q_k]$ (ignoring changes in industry and differences in average premiums within the origin and destination quartiles). Thus, it should be symmetric – a move from quartile 4 to 1 should yield roughly as large a drop in earnings as the increase from a move from 1 to 4.[24] The figure indicates that this prediction largely holds in the data. Moreover, we see again that earnings are flat prior to a move and adjust relatively quickly following a move – within two quarters, the movers' earnings have largely stabilized.

Appendix Figure 3 shows the wage residuals from model (2) for the same groups of movers. Focusing on the residuals allows us to greatly expand the scale of the y-axis. For movers who leave quartile 1, we tend to see negative residuals in the first post-move quarter, indicating that they have not achieved the full earnings gain the model predicts. But by the second quarter after a move, the residuals are much closer to 0, and within a year the mean residuals are all less than 1% in magnitude. In contrast, movers who leave quartile 4 have positive residuals in every quarter after the move, suggesting that some movers from upper quartile CZ-industry cells are moving to cells with lower average pay but receiving a positive match effect, partially offsetting the losses implied by the estimated

[23] Recall that in our event study sample we allow up to a year between the last (non-transitional) quarter of work in the old CZ and the first (non-transitional) quarter of work in the new CZ.

[24] Violations of symmetry are possible even with exogenous mobility because CZs are grouped into quartiles. Those who leave quartile 4 for quartile 1 may come from different CZs within the quartile than those who move to quartile 4. Below, we explore symmetry more carefully with much smaller groups of CZs. Violations of symmetry can also occur because many CZ movers also change industry, and the industry changes may not be symmetric.

CZ-industry effects in our model.[25] Nevertheless, the magnitude of the residual components is small, amounting to only a few percentage points. Overall, the model seems to fit reasonably well.

Figure 7 presents another approach to validating the model. Here, we divide CZs into vingtiles based on their mean earnings premiums ($\bar{\psi}_c$) and plot the average change in log earnings for movers in each of the 400 possible origin and destination cells against the predicted change based on $\bar{\psi}_c$ in the origin and destination vingtiles. Since the CZ effects are estimated with some error we use an IV approach to predict the change in mean $\bar{\psi}_c$ for people in our A sample between the origin and destination cells using mean CZ effects estimated using the averages of the CZ-industry effects from models fit to each of the 27 other 3.33% samples.[26] Across all 400 transitions, the (instrumented) slope is almost exactly one (0.996, standard error=0.0132), suggesting that the two-way fixed effects model does a good job of capturing the average effect of moves. There is some evidence that the earnings losses are smaller than predicted for people with predicted earnings changes in the -5 to -15% range, and that the gains are smaller than predicted for movers with predicted earnings changes in the 10-20% range, but overall the model fits the data relatively well. Figure 8 explores this lack of symmetry further, by comparing the average changes for "upward" movers to the average changes for the symmetric "downward" movers. Earnings fall by a bit less for the downward movers than would be expected from the increase for the upward movers, especially when the move is between CZs with fairly similar $\bar{\psi}_c$'s, but the symmetry violations are modest.

We also conducted the analysis in Figure 7 separately for workers in the top and bottom terciles of the distribution of α_i. The slope across the origin-destination cells for movers in bottom tercile is 0.980 (standard error = 0.015), while the slope for movers in the top tercile is 0.978 (s.e. = 0.026). The

[25] Such a pattern could also arise if younger people tend to work in high-premium CZ's for a few years, then move to lower-premium CZ's, but receive a reward for their "big city" experience. Such rewards are documented by de la Roca and Puga (2016) in Spain. Di Addario et al. (2021) develop a methodology for estimating persistent effects of past jobs in an AKM framework. The extension of their methodology to our setting is a topic of future research.
[26] Results are nearly indistinguishable if we use the B sample to construct the instrument.

fact that both of these are very close to 1 suggests that the CZ effects estimated in our pooled model are highly predictive of the earnings gains and losses for workers with lower or higher skills, supporting the additive structure of our basic model.

Finally, in Appendix Figure 4 we plot mean residuals by decile of the person and CZ-industry effect distributions. As has been found in previous applications of the AKM model, we find positive average residuals at the bottom corner of the distribution, corresponding to the lowest individual and CZ-industry effects, and negative average residuals at the top corner. But both are quite small (absolute value < .005) indicating that the model violations are typically small and unsystematic.

Overall, our investigation indicates that the simple AKM model, with its implicit assumptions of additive separability and exogenous mobility, fits the data quite well, though there are some violations of the exogenous mobility assumption, potentially reflecting persistent match effects, that could be incorporated in future work.

The role of industry

Equation (2) includes unrestricted CZ-by-industry effects, ψ_{jc}. We find large variation across commuting zones in the weighted average of these effects. In this section, we explore the role of industry in the ψ_{jc} effects. Of particular interest are whether industry and place effects are additively separable, or whether there are important place-by-industry match effects, and whether agglomeration of certain industries in particular places contributes importantly to across-place differences in earnings even without geographic differences in pay within industries.

We begin with equation (7), estimating a simple analysis of variance decomposition of the $\hat{\psi}_{jc}$ estimates. Columns 1-3 of Table 5 report models with CZ effects, industry effects, and both, respectively. Over two-fifths of the variation in $\hat{\psi}_{jc}$ is across CZs, and a slightly larger share is across industries. Because CZ and industry are somewhat correlated with each other due to industry agglomerations, these need not be additive. However, column 3 shows that they are very nearly so: A

specification with both CZ and industry effects explains 88.4% of the variance, just 0.6 percentage points less than the sum of that explained in columns 1 and 2. This is our first indication that industry agglomerations do not play a large role in between-CZ earnings differences.

Based on the R-squared of the model in column 3, match effects – differences in industry effects across CZs, or alternatively differences in CZ effects across industries – account for as much as 12% of the variation of ψ_{jc}. But this calculation potentially overstates the importance of match effects since sampling errors in $\hat{\psi}_{jc}$ are included as part of the residual of the model in column (3). To explore this, we take advantage of the fact that we estimate the AKM decomposition on multiple non-overlapping samples of the full LEHD data. The estimates in columns 1-3 are based on sample A. We re-estimate column 3 in sample B and compute the residual, $\hat{e}^B_{jc}$, then include that residual as an additional explanatory variable in sample A. The coefficient on the residual estimates the correlation between $\hat{e}^A_{jc}$ and $\hat{e}^B_{jc}$, or, alternatively, the reliability of estimated match effects.

The results are presented in column 4. The coefficient on $\hat{e}^B_{jc}$ is 0.88. This is quite high, implying that there is relatively little measurement error in our estimates of the match effects and that the variance component due to sampling error is very small. We estimate that the variance of the true match effect accounts for 10% of the variance of ψ_{jc}.[27] In other words, while not a large share of the overall variance, match effects are statistically distinguishable and reliably estimated.

Having established that estimated match effects are not statistical artifacts, we next explore whether they reflect industry agglomerations. In the final column of Table 5, we return to the specification from column 3 but add a control for the size of the CZ-industry cell, measured as a share of

[27] The 0.878 reliability of the estimated match effects implies that true match effects account for about 88% of the residual component of CZ-industry effects after controlling for CZ dummies and Industry dummies. The R-squared of 0.884 in column 3 of Table 5 implies that the residual component is about 12% of the variance in estimated CZ-industry effects. Since the standard deviation of the estimated CZ industry effects is 0.097 (Table 3, column 3), the standard deviation of the true match effects is approximately 0.031, accounting for about 10% of the variance in the true (sampling error corrected) CZ-industry effects.

the CZ labor market. We find that this is strongly positively related to the match effect: A one percentage point increase in the cell's employment share is associated with a 2.4 percent increase in earnings in the cell, on top of pure CZ and pure industry effects. However, the goodness of fit increases only slightly here from column 3; a comparison of the R-squared coefficients in columns 3-5 implies that size explains only about 10% of the variance of match effects. Evidently, while there is variation in match effects, it is very weakly related to the industry's CZ employment share.

To explore this further, we turn to the formal decomposition of CZ-mean effects, Ψ_c, presented in equation (6). We are interested in how much of the variation in Ψ_c is due to "pure" locational wage effects, relative to employment in the same industries in other locations; to agglomeration in industries that are high wage everywhere; and to match effects.

Results are presented in Table 6. Surprisingly, pure locational wage premia account for slightly more than 100% of the variation in Ψ_c. Agglomeration in industries with different national wage profiles explains 1.5%, but because the share-weighted national industry effect in a CZ is negatively correlated with the pure locational component – high wage industries tend to locate in areas with low location wage effects – the covariance is negative and more than fully offsets the small positive variance contribution of agglomeration. Finally, specialization of CZs in industries where the CZ has a comparative pay advantage – a match effect – explains an additional 0.8% of the overall variation in Ψ_c, a negligible amount. Thus, while match effects and agglomeration are statistically significant, overall they each play negligible roles in explaining across-CZ differences in Ψ_c.

A potential concern with our analysis here is that we may not be accounting for industry in enough detail. Our estimates are based on "2-digit" industry coding with 24 sectors. This is too coarse to capture many well-known industry agglomerations, particularly in specific manufacturing subsectors (e.g., furniture in North Carolina or steel in Pittsburgh). It is computationally infeasible to include a full set of industry-by-commuting zone effects with very detailed industries. Instead, we explore this by

limiting attention to the 50 largest commuting zones, with 58% of the workers in our sample. We re-estimate the AKM model in this sample, using four-digit industries. Results are presented for the largest cities with two-digit industries in column 2 of Table 6, and with four-digit industries in column 3. We find that CZ average person effects ($\bar{\alpha}_c$) are extremely highly correlated ($\rho = 0.999$) between these two sets of estimates, indicating that the apportionment of CZ earnings differences into person and place-by-industry effects is highly robust to industry coding. A comparison of the variance decompositions in columns 2-3 of Table 6 for the 50 largest CZ's using 2-digit or 4-digit industries confirms this robustness.

It appears that the overwhelming component of the variance of Ψ_c across commuting zones reflects differences in the true causal effect of places that are common across industries, rather than a confounding effect arising from differences in industry structure. We noted in the discussion of Table 3, however, that places that offer higher wage premiums tend to have higher-α workers. Given this, there is a subtle feature of high-Ψ_c places that may interact with local industry structure to affect average wage outcomes. Specifically, Dauth et al. (2019) find evidence that high skilled workers (indexed by their person effect in an AKM model) are more likely to work at establishments paying higher wage premiums in larger cities in Germany. In our setting, a similar phenomenon could affect the mapping between workers and <u>industries</u> in high-Ψ_c places.

To explore this, we construct a measure of the degree of assortativeness of workers into high-wage industries at the CZ level by estimating the correlation of $\overline{\alpha}_{jc}$ with the <u>average</u> industry pay premium $\overline{\psi}_j$ separately for each CZ. We see some evidence that the degree of sorting is higher in high-pay cities: for CZ's in the bottom quintile of values of Ψ_c the correlation is 0.57 while in CZ's in the top quintile the correlation is 0.70. We return to this issue below in discussing the effects of CZ **size**. There, we see a much stronger pattern of increasing assortativeness in larger CZ's, consistent with the results of Dauth et al. (2018).

A final potential concern is that we do not account for occupational differences in our analysis. The LEHD does not include occupation measures. But we don't believe it is necessary to control for occupation in any case, for three reasons. First, occupation is largely a characteristic of the worker, capturing the skills that the worker brings to his or her job search, so is likely to be absorbed by our permanent worker earnings control α. Second, particularly in lower-skill industries, occupations are highly correlated with detailed industry – e.g., waiters and busboys are found overwhelmingly in the restaurant and food service industry, school teachers are found in elementary and secondary schools, auto mechanics are found in automotive maintenance and repair. Our analysis in Table 6 showed that our main conclusions are highly robust to using detailed (4 digit) industry effects to define the pay premiums in equation (2). Last, insofar as workers gain or lose access to occupations when they move locations, as when a teacher's aide moves to a place with lower credentialing requirements and is able to obtain a position as a classroom teacher, the consequent wage changes are reasonably considered to be part of the causal place effect.

Returns to education

Next we turn to an exploration of differences in the return to education across commuting zones. In order to implement the analysis described in equations (9) and (10), we have to re-estimate our two way fixed effects models separately on samples of low and high educated workers. We start with our full analysis sample (the sample from which we draw the A and B subsamples), finding the subset that have been interviewed in the ACS at some point in the 2001-2017 period, were age 30 or older when interviewed, and provided their education. This subsample represents about 14% of the full analysis sample, somewhat larger than our main 10% A sample. We then separate out two groups: people with no more than 12 years of completed schooling (the low education group) and people with at least some college (the high education group).

Appendix Table 3 provides a summary of the estimation results for two-way fixed effects models fit to these two samples. For ease of comparison we use the same format as Table 3, but show two panels, one for each education group. The key features of the estimation results are broadly similar to those for the overall sample. In particular, at the individual level the variance of the person effects accounts for around three quarters of the overall variance of earnings, regardless of sample, while CZ industry effects and their covariance with the person effects account for another 7-10%. At the more aggregated level, however, an interesting pattern emerges. For lower-educated workers, CZ-industry effects account for a noticeably larger share of the variance of mean earnings across CZ-industry cells (20%) and CZ-average cells (38%) than for higher-educated workers (11% and 14%, respectively). In contrast, variation in the person effects of a CZ's residents account for a smaller share of the variation in CZ average earnings differences for lower-educated workers (33%) than for higher-educated workers (53%). There is also a notable difference in the degree of assortative matching across CZ's: the correlation between mean person effects and the average CZ wage premium is 0.42 for lower-educated workers but 0.65 for higher-educated workers.

With this background we turn to equations (9) and (10), which decompose the CZ-level earnings gap between high- and low-education workers, $\bar{y}_{cH} - \bar{y}_{cL}$. Equation (9) includes three components: The difference in person effects between high- and low-education workers in the CZ, differences in covariates (age and time), and the difference in CZ-aggregated wage effects. Equation (10) further decomposes this last term into three: the CZ average excess wage premium, the CZ's industry composition, and the interaction between these.

Columns 1 and 2 of Table 7 present regressions of the CZ-level wage gap and each of its components on the mean log wage in the CZ (as measured from the ACS). Consistent with the pattern shown in Figure 3 (which is based entirely on ACS data), the coefficient of a regression of the education wage gap in a CZ on the mean log wage is 0.644 – higher-wage CZs have much larger education wage

gaps. When we decompose the source of this divergence using equation (9), we find that over 90% (0.604/0.644) is attributable to differences in person effects. That is, high-wage cities have high apparent returns to education not because more educated workers would gain more by moving there but because of assortative matching: the relative skills of the more educated workers who live there are higher. Less than one-fifteenth of the total (0.042/0.644) is attributable to differences in CZ-aggregated wage effects – that is, to higher returns to education (averaged across industries) in high-wage cities. More than 100% of this component reflects industry sorting: In high wage cities, more educated workers are more concentrated in industries that pay high wages everywhere relative to less educated workers. There is no indication that high wage cities have higher within-industry wage premia for more educated workers (i.e., $\psi_{jcH} - \psi_{jcL}$); rather, the contribution of this component, the first term of (10), is negative, indicating that within industries less educated workers receive *more* of a premium for working in high-wage cities than their more educated coworkers.

Columns 3 and 4 of Table 7 use the same five-part decomposition, this time exploring the overall variation in the CZ-level return to skill rather than just the portion associated with mean CZ earnings. We see a generally similar pattern here. Differences in the person effects of more- and less-educated workers explain 78% of the variation in the apparent return to education, while differences in true returns explain only 7.5%. Differences in industry composition explain 2.1% of the variance, but the covariance between industry composition and person effects makes a large variance contribution (19.5%).

The CZ size gradient

As a final investigation, we explore the wage premium associated with larger CZs. Table 8 presents a series of bivariate regressions of different CZ characteristics on the log size of the CZ's workforce. These characteristics are measured from ACS data in columns 1-2 and from the LEHD in columns 3-4. The first rows show that the elasticity of earnings with respect to CZ size is between 0.07 and 0.08 depending on

the exact earnings measure. Rows 4 and 5 decompose earnings into a skill component (predicted earnings based on education, experience, etc., in the ACS; $\bar{\alpha}_c$ in the LEHD) and a CZ effect. Here, the stories are quite different: The ACS analysis indicates that skill is only weakly related to CZ size, explaining only 17% of the size effect on earnings, while the LEHD indicates a much stronger skill-size gradient, explaining nearly two-thirds of the total. Evidently a large part of the higher earnings in large CZs reflects more skilled workers within observed skill categories. The LEHD sample also indicates that the standard deviation of worker skill is larger in large cities, where in the ACS this relationship is much weaker.

In rows 10-13 we use the decomposition of equation (6) to dig further into the sources of the relationship between the city-average wage premium, Ψ_c, and log CZ size. Here we see that the CZ-specific premium component ($\sum_j \bar{s}_j(\psi_{cj} - \bar{\psi}_j)$) drives the entire effect. Once again, industry composition effects and interactions between local industry-specific premiums and the local share of workers in the industry are unimportant.

Finally, Row 14 returns to the question of whether the degree of sorting between high wage workers and high wage industries varies across CZ's. Classifying CZ's by size we see a significant positive relationship: a 100 log point increase in CZ size is associated with a rise of +0.0612 in the correlation between the mean person effects in an industry and the average premium paid by that industry nationally (i.e., $\rho(\bar{\alpha}_{jc}, \bar{\psi}_j)$). This confirms Dauth et al.'s (2018) findings in Germany regarding the role of market size in enhancing sorting.

Finally, in Table 9 we examine the relationships between log size and the components of the return to education. Panels a and b examine non-college and college workers separately, while Panel c examines the gap between the two. Comparing the first row of Panel a with that of Panel b, we see that the elasticity of earnings with respect to city size is much larger for more educated workers regardless of

data source, though the size elasticities for both education groups are somewhat larger using our LEHD samples than using the ACS.

Next, in row 2 of each panel we examine the part of earnings that is explained by observed skills in the ACS (mainly education, experience, and immigrant source country) or by person effects in the LEHD. In the ACS, the observed skill component for lower-educated workers is negatively correlated with log size, whereas in the LEHD, mean person effects for low educated workers are positively correlated with size. Among higher-educated workers the discrepancy is even larger: here the ACS data point to only a small size elasticity of worker skills, but the more comprehensive skill measure in the LEHD (based on mean person effects) is quite strongly correlated with CZ size.

The opposite pattern emerges for the CZ wage effects. The CZ wage premiums that we estimate from the ACS are strongly correlated with log size, and explain most of the correlation of earnings with size. But the CZ premiums we estimate from the LEHD, which are purged of the confounding effect of unobserved worker skills, are much less strongly correlated with size.

Lastly, in panel c of Table 9, we examine the determinants of the size effect on the measured wage gap between high- and low-educated workers. We see that 88% of the size effect (0.0540/0.0612) derives from the higher size elasticity of skill for highly educated workers than less educated workers. All of the remaining 12% is attributable to the differences in relative industry composition for the two education groups, though there is a small interaction effect that accounts for about 2% (=0.0015/0.0612) of the overall elasticity of the return to education with respect to CZ, offset by a small negative contribution from differences in earnings effects within industries.

VI. Earnings and the cost of living

Thus far, we have considered decompositions of nominal earnings, unadjusted for local differences in the cost of living. But there is well-known, dramatic variation in housing costs between

places, and many of the coastal cities with the highest housing costs also have the biggest causal effects on earnings (Figure 5). Housing is not the only cost that varies across cities; using a more comprehensive measure, Diamond and Moretti (2021) find that cost of living varies dramatically across places, identifying as high-cost places many of the coastal cities (e.g., San Francisco, New York, Boston) that we find have high place effects on nominal earnings. A natural question is how the causal earnings effects that we identify relate to differences in local costs – does moving to a higher-earnings-effect city mean an increase in real earnings, or are the increased nominal earnings offset by higher costs?

Diamond and Moretti (2021; hereafter *DM*) find that the real consumption of college graduates does not vary with city size, as earnings differences move roughly one-to-one with prices, but that non-college workers' earnings fail to keep up with rising prices in larger cities, leading to lower real consumption. This is a cross-sectional analysis, so a part of the earnings differences across CZs that they identify reflects differences in unobserved skill. Our results suggest that the larger cities tend to have more skilled workers, particularly among college workers, and that differences in causal earnings effects are much smaller. Specifically, DM find that the elasticity of prices with respect to log city size for college graduates is 0.040.[28] We find that the elasticity of average earnings of college workers is 0.102, but that the elasticity of the causal earnings effect is just 0.029. Combining this with DM's estimate, the implication is that the causal effect of moving from a smaller, low-cost, low-wage city to a larger, high-cost, high-wage city is to *reduce* real earnings, even for college graduates.

To investigate this more directly, we use rents and housing costs information from the ACS to explore how housing costs vary with CZ size and mean wages, two dimensions that we explored earlier (in tables 8 and 4, respectively). Appendix Table 4 presents regressions of four CZ-level housing

[28] We obtain this as the difference in the expenditure and consumption elasticities in DM's Figure 13a. DM's college graduates group does not align exactly with our some college or more group, but DM obtain similar price elasticities for their middle- and lower-skill groups, and we find even smaller causal effects on nominal earnings, so the result holds for lower-skill workers as well.

measures (mean log home values for homeowners and mean log rents for renters, each unadjusted and then adjusted for housing characteristics[29]) on CZ size and, separately, on CZ mean log earnings. We find that the elasticity of housing costs with respect to log size is around 0.2 or larger, while the elasticity with respect to mean log wages ranges from 1.8-3.0, depending on the measure. By comparison, Table 8 indicates that the elasticity of the CZ earnings effect with respect to log size is around 0.03, one-seventh or less of the corresponding housing cost elasticity. Table 4 shows that the elasticity of the earnings effect with respect to mean log earnings in the CZ is 0.38, less than one-quarter that of the housing cost elasticity. Because housing typically represents at least one-third of a household's budget, these estimates imply (consistent with Diamond and Moretti's results) that housing costs consume more than 100% of the nominal earnings gain that a typical worker obtains from moving to a larger or higher-earnings CZ.

An interesting question, beyond the scope of this paper, is how this can be sustained. Causal effects on nominal earnings seem to require productivity differences across places. But why would workers prefer to live in high-productivity cities if they will need to give up all of the earnings advantage of those cities in higher housing costs? One potential explanation is that large cities have even greater consumption amenity advantages than they do productivity advantages (Albouy 2011; Albouy, Cho, and Shappo 2021).

VII. Concluding Remarks

We have used an earnings model with a combination of individual worker effects and additive premiums for different commuting zone/industry combinations to address longstanding questions about

[29] The quality adjustment models include controls for type of unit (with five classes of apartment building size), number of bedrooms and of total rooms, year of construction, indicators for utilities included in rent (for rents only), and an indicator for whether the owner has a mortgage (homeowner model only). The model for rents has an R2 of 0.34, while that for home values has R2=0.50.

the impact of place on labor market outcomes in the U.S. This class of two-way fixed effect models, originated by Abowd, Kramarz and Margolis (1999), has proven very useful in answering questions about the role of firms in wage determination (Card et al., 2018). While versions of this approach have been used in France (Combes et al., 2008), Spain (de la Roca and Puga, 2017), and Germany (Dauth et al., 2018), ours is the first to apply it to the U.S., using data from the Census Bureau's LEHD program.

Building on the AKM-related literature, we first show that a model with additive premiums for different commuting zone (CZ) and industry combinations provides a relatively good summary of the main patterns in the data, and can be estimated by simple OLS methods without too much concern for biases arising from either the strategic timing of moves or idiosyncratic match effects in earnings that drive mobility decisions.

A key advantage of our specification, which generalizes models with only worker and place effects (or with additive worker, place, and industry effects) is that we can carefully assess the role of industry in mediating observed place effects in average earnings. Such effects can arise in two main ways. First, there can be a pure compositional effect if some CZ's have a higher fraction of high-wage industries. More subtly, there can be an interaction effect if the earnings premiums for different industries vary across places and employment is locally concentrated in industries with a larger local premium. We find very small roles for either mechanism. In fact, CZ × industry pay premiums are approximately separable: only about 10% of the variation in these premiums is due to CZ-specific premiums. Moreover, we find almost no evidence of interaction effects arising from the concentration of employment in sectors with a local pay premium. Importantly, these conclusions are highly robust to the definition of industry. Comparing models with only 24 industries with models with close to 300, we reach nearly identical conclusions.

As in the AKM-related literature, we measure worker skills by the worker's fixed effect in the earnings model. We find that this measure of skill varies far more widely across CZ's than a more

traditional measure based on observable characteristics like education, age and gender. Consistent with work in France, Spain and Germany we find that the main explanation for high wage places is the presence of high wage people. Indeed, about 50% of the variation in mean CZ earnings is due to differences in average person effects across places; another 35% is due to the positive covariance between person effects and mean CZ wage premiums. The positive assortative matching between high wage workers and high-wage places magnifies overall earnings inequality in the market as a whole, and represents an important feature of locational equilibrium.

We find two interesting sources of variation in the degree of assortative matching. First, college-educated workers are more highly sorted to high-pay premium CZ's than their less educated counterparts. Nearly all of the higher "return to college" that is observed in higher-wage (or larger) places is attributable to the presence of college workers with the highest unobserved skills in those places. The differences in sorting are consistent with Diamond (2016), though her model ignores unobserved skills and treats earnings differences as causal. Second, we find that the sorting within CZ's of high wage workers into high wage industries is enhanced in larger places. This confirms a similar finding by Dauth et al. (2018) in Germany, and illustrates a potential benefit of increased market size for the overall productivity of the economy.

References

Ashenfelter, Orley. "Estimating the effect of training programs on earnings." *The Review of Economics and Statistics* 60, no. 1 (1978): 47-57.

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

Abowd, J. M., B. E. Stephens, L. Vilhuber, F. Andersson, K. L. McKinney, M. Roemer, and S. Woodcock. "The LEHD infrastructure files and the creation of the Quarterly Workforce Indicators." In: T. Dunne, J. B. Jensen, and M. J. Roberts (Eds.), *Producer dynamics: New evidence from micro data*, vol. 68 (2009): 149-230. Chicago, IL: University of Chicago Press.

Di Addario, Sabrina L., Patrick M. Kline, Raffaele Saggio, and Mikkel Sølvsten. "It ain't where you're from, it's where you're at: Hiring origins, firm heterogeneity, and wages." National Bureau of Economic Research working paper no. w28917, 2021.

Albouy, David. "Are big cities bad places to live? Estimating quality of life across metropolitan areas." National Bureau of Economic Research working paper no. w14472, revised 2011.

Albouy, David, Heepyung Cho, and Mariya Shappo. "Immigration and the pursuit of amenities." *Journal of Regional Science* 61, no. 1 (2021): 5– 29.

Andrews, Martyn J., Len Gill, Thorsten Schank, and Richard Upward. "High wage workers and low wage firms: Negative assortative matching or limited mobility bias?" *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171, no. 3 (2008): 673-697.

Baum-Snow, Nathaniel, and Ronni Pavan. "Understanding the city size wage gap." *The Review of Economic Studies* 79, no. 1 (2012): 88-127.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. "Productive cities: Sorting, selection, and agglomeration." *Journal of Political Economy* 122, no. 3 (2014): 507-553.

Blinder, Alan S. "Wage discrimination: Reduced form and structural estimates." *Journal of Human Resources* 8, no. 4 (1973): 436-455.

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

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

Chetty, Raj, and Nathaniel Hendren. "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects." *The Quarterly Journal of Economics* 133, no. 3 (2018a): 1107-1162.

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

Ciccone, Antonio, and Robert E. Hall. "Productivity and the density of economic activity." *The American Economic Review* 86, no. 1 (1996): 54-70.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. "Spatial wage disparities: Sorting matters!" *Journal of Urban Economics* 63, no. 2 (2008): 723-742.

Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Suedekum. "Matching in cities." National Bureau of Economic Research working paper no. w25227, 2018.

Diamond, Rebecca. "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000." *American Economic Review* 106, no. 3 (2016): 479-524.

Diamond, Rebecca, and Enrico Moretti. "Where is standard of living the highest? Local prices and the geography of consumption." Working paper, July 2021.

Duranton, Giles, Puga, Diego. Micro-foundations of urban agglomeration economies. In: J. Vernon Henderson and Jacques-François Thisse (Eds.), *Handbook of Urban and Regional Economics*, vol. 4 (2004): 2016-2117. North-Holland, Amsterdam.

Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny. "Spatial sorting." *Journal of Political Economy* 122, no. 3 (2014): 554-620.

Ellison, Glenn, and Edward L. Glaeser. "Geographic concentration in US manufacturing industries: a dartboard approach." *Journal of Political Economy* 105, no. 5 (1997): 889-927.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. "Sources of geographic variation in health care: Evidence from patient migration." *The quarterly journal of economics* 131, no. 4 (2016): 1681-1726.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. "Place-based drivers of mortality: Evidence from migration." *American Economic Review* 111, no. 8 (2021): 2697-2735.

Glaeser, Edward L., and Joshua D. Gottlieb. "The wealth of cities: Agglomeration economies and spatial equilibrium in the United States." *Journal of Economic Literature* 47, no. 4 (2009): 983-1028.

Glaeser, Edward L., Maré, David C. "Cities and skills." *Journal of Labor Economics* 19 no. 2 (2001): 316–342.

Gould, Eric D. "Cities, workers, and wages: A structural analysis of the urban wage premium." *The Review of Economic Studies* 74, no. 2 (2007): 477-506.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. "Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings." *Journal of Political Economy* 118, no. 3 (2010): 536-598.

Kline, Patrick, and Enrico Moretti. "Local economic development, agglomeration economies, and the big push: 100 years of evidence from the Tennessee Valley Authority." *The Quarterly Journal of Economics* 129, no. 1 (2014): 275-331.

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

Moretti, Enrico. "Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data", *Journal of Econometrics* 121, no. 1-2 (2004a): 175-212.

Moretti, Enrico. "Human capital externalities in cities." In: J. Vernon Henderson and Jacques-François Thisse (Eds.), *Handbook of Urban and Regional Economics*, vol. 4 (2004b): 2243-2291. North-Holland, Amsterdam.

Moretti, Enrico. "Local labor markets." In Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics*, vol. 4B. (2011): 1237-1313. North-Holland, Amsterdam**.**

Oaxaca, Ronald. "Male-female wage differentials in urban labor markets." *International Economic Review* 14, no. 3 (1973): 693-709.

Roca, Jorge de La, and Diego Puga. "Learning by working in big cities." *The Review of Economic Studies* 84, no. 1 (2017): 106-142.

Rosenthal, Stuart S., and William C. Strange. "Evidence on the nature and sources of agglomeration economies." In: J. Vernon Henderson and Jacques-François Thisse (Eds.), *Handbook of Urban and Regional Economics*, vol. 4 (2004): 2119-2171. North-Holland, Amsterdam.

Tolbert, Charles M., and Molly Sizer. "U.S. commuting zones and labor market areas: A 1990 update." U.S. Department of Agriculture, Economic Research Service, staff paper number AGES-9614 (1996).

Table 1: Descriptive Statistics for Commuting Zones: 2010-2018 American Community Survey

	All CZ's		Largest 50 CZ's		Univariate Regression on Log(Workforce Size)		
	Mean (1)	(Std. Dev) (2)	Mean (3)	(Std. Dev) (4)	Coeff. (5)	(Std. Err) (6)	R-sq. (7)
a. Mean Worker Characteristics							
Mean Log Hourly Wage	2.863	(0.141)	2.944	(0.113)	0.069	(0.003)	0.519
Mean Share BA+	0.324	(0.083)	0.371	(0.065)	0.039	(0.002)	0.485
Mean Years Education	13.615	(0.461)	13.798	(0.417)	0.142	(0.011)	0.209
Mean Share Immigrants	0.167	(0.123)	0.231	(0.117)	0.060	(0.002)	0.523
Mean Share Black Non-Hisp.	0.119	(0.094)	0.129	(0.076)	0.011	(0.002)	0.028
Mean Share Asian Non-Hisp.	0.056	(0.056)	0.080	(0.054)	0.023	(0.001)	0.360
Mean Share Hispanic	0.168	(0.152)	0.209	(0.142)	0.044	(0.004)	0.180
Mean Share White Non-Hisp.	0.658	(0.189)	0.582	(0.164)	-0.077	(0.004)	0.364
b. Decomposition of mean wages from regression model [a/]							
CZ wage residual	0.000	(0.109)	0.062	(0.087)	0.055	(0.002)	0.572
CZ predicted wage	2.863	(0.050)	2.882	(0.048)	0.013	(0.001)	0.152
c. Add 2-digit Industry Effects to wage model							
CZ wage residual	0.000	(0.105)	0.060	(0.084)	0.054	(0.002)	0.573
CZ predicted wage	2.863	(0.054)	2.884	(0.052)	0.015	(0.001)	0.165
d. Add 4-digit Industry Effects to wage model							
CZ wage residual	0.000	(0.102)	0.057	(0.083)	0.052	(0.002)	0.566
CZ predicted wage	2.863	(0.057)	2.887	(0.053)	0.017	(0.001)	0.185
Mean (4-digit) Ind. Composition Effect	0.186	(0.017)	0.193	(0.016)	0.004	(0.000)	0.142
Share of Total Sample	1.000		0.583				

Notes: Sample in columns 1-2 includes 11,733,554 observations in 688 Commuting Zones in the American Community Survey (ACS), 2010-18 (excluding Alaska). All statistics are weighted using ACS population weights. Regression models in columns 5-7 are fit to CZ-means of row variable using log(CZ workforce size) as explanatory variable, and weighting by sum of ACS weights for workers in CZ.

[a/] See text for description of regression model. Model includes controls for education, gender, potential experience, country of birth.

Table 2: Characteristics of Samples Derived from Longitudinal Employer-Household Dynamics (LEHD) Data Base

			Estimation Sample:				Event Study Sample			
	All LEHD Obs. (1)	Estimation Sample (2)	CZ & Ind. Stayers (3)	CZ Stayer Ind. Mover (4)	CZ Mover Ind. Stayer (5)	CZ & Ind. Mover (6)	All (7)	CZ Stayer Ind. Mover (8)	CZ Mover Ind. Stayer (9)	CZ & Ind. Mover (10)
Share Qtly Earnings ≥ $3800	0.837	1	1	1	1	1	1	1	1	1
Mean Earnings (if ≥ $3800)	16,050	16,600	16,890	15,210	18,500	16,300	17,420	16,200	19,090	17,750
Mean Age	41.11	42.51	44.82	40.11	40.96	38.79	39.95	39.82	40.65	39.44
Fraction Female	0.492	0.472	0.497	0.465	0.476	0.405	0.438	0.453	0.458	0.391
Fraction Foreign Born	0.167	0.163	0.167	0.175	0.148	0.143	0.144	0.155	0.135	0.135
Number of CZ's during Sample Period:										
1 CZ	0.631	0.731	1	1	0	0	0.454	1	0	0
2 CZ's	0.231	0.201	0	0	0.825	0.712	0.546	0	1	1
3+ CZ's	0.138	0.068	0	0	0.175	0.288	0	0	0	0
Number of 2-digit Industries during Sample Period:										
1 Industry	0.495	0.630	1	0	1	0	0.282	0	1	0
2 Industries	0.263	0.257	0	0.775	0	0.616	0.719	1	0	1
3+ Industries	0.242	0.113	0	0.225	0	0.385	0	0	0	0
Mean # Quarters Observed	21.39	22.49	24.12	21.22	22.43	19.76	10	10	10	10
Mean Est. Person Eff.		9.419	9.434	9.338	9.526	9.405	9.492	9.432	9.579	9.503
(std. dev.)		(0.561)	(0.568)	(0.540)	(0.568)	(0.548)	(0.542)	(0.525)	(0.550)	(0.550)
Num. P-Q obs (millions)	368.70	255.80	140.00	46.94	21.05	47.80	12.30	5.58	3.46	3.25
Num. Persons (millions)	17.24	11.37	5.80	2.21	0.94	2.42	1.23	0.56	0.35	0.33

Notes: Sample includes person-quarter (PQ) observations for indivduals age 22-62 with at least 8 quarters of employment in the LEHD 2010Q1 to 2018Q2. Quarterly observations for individuals with multiple employers are excluded, as are the first and last (transitional) quarters of any spell with the same employer, and quarters for which industry or location information is missing. Estimation sample in column 2 drops quarters with <$3800 in earnings.

Table 3: Summary of Estimated Two-Way Fixed Effects Models

	Person-quarter level		CZ-industry level		CZ level	
	Std. Dev. or Correlation	Var. Share	Std. Dev. or Correlation	Var. Share	Std. Dev. or Correlation	Var. Share
	(1)	(2)	(3)	(4)	(5)	(6)
Log earnings or mean log earnings	0.654	1.000	0.275	1.000	0.145	1.000
Variance components (std. deviations in odd-number columns, variance shares in even-numbered columns)						
Person effects	0.561	0.736	0.199	0.526	0.100	0.472
CZ-industry effects	0.097	0.022	0.097	0.125	0.064	0.196
Covariate index (Xβ)	0.150	0.053	0.015	0.003	0.005	0.001
Residual	0.243	0.138	0.000	0.000	0.000	0.000
Covariance components (correlations in odd-number columns, variance shares in even-numbered columns)						
Person/CZ-industry	0.211	0.054	0.593	0.308	0.563	0.347
Person/Covariates	-0.010	-0.004	0.387	0.032	-0.029	-0.001
CZ-industy/Covariates	0.026	0.002	0.256	0.010	-0.297	-0.009

Notes: Table shows variance decompositions based on equation (3). Columns 1-2 pertain to the variance of individual quarterly earnings. Columns 3-4 pertain to the variance of mean earnings by CZ-industry cell. Columns 5-6 pertain to the variance of mean earnings by CZ. Entries in columns 1-3-5 for "variance components" are standard deviations of earnings component indicated in row heading; for "covariance components" are the estimated correlations of the indicated variance components. Entries in columns 2-4-6 are variance shares explained by variance or covariance components. Variance components are estimated using two-sample method -- see text.

Table 4: Cross-CZ Covariation of Mean Measured Skills and Mean CZ Effects with Mean Earnings/Wages

	LEHD		ACS	
	Mean Person Effects	Mean CZ/Industry Effects	Mean Skill Index	CZ Effect
	(1)	(2)	(3)	(4)
Estimated coefficient on CZ mean log earnings/wages	0.621	0.382	0.268	0.732
(standard error)	(0.013)	(0.013)	(0.016)	(0.016)

Notes: Table shows estimated regression coefficients from weighted OLS regression of mean CZ-level variable described in column heading on mean log CZ earnings (columns 1-2) or mean log CZ hourly wages (columns 3-4). Dependent variables in columns 1-2 are CZ averages of estimated person effects and CZ-industry effects from two-way fixed effects model estimated on LEHD data. Dependent variables in columns 3-4 are CZ averages of estimated covariate index and estimated CZ effects from linear regression model estimated on ACS data.

Table 5: Estimated Models for Estimated CZ-Industry Effects

			Models with CZ and Industry Effects		
	CZ Effects only	Industry Effects only	no other controls	plus residual from CZ+industry model for Sample B	plus local emp. share of industry
	(1)	(2)	(3)	(4)	(5)
Residual from model with CZ + industry, fit to Sample B	--	--	--	0.878	--
(standard error)				(0.008)	
Share of CZ emp. in Industry	--	--	--	--	0.243
(standard error)					(0.019)
R-squared	0.433	0.457	0.884	0.973	0.894
RMSE	0.073	0.072	0.033	0.016	0.032

Notes: Table shows goodness of fit and estimated regression coefficients from regression of estimated CZ-industry effects on CZ effects (column 1), industry effects (column 2), and CZ and industry effects (columns 3-5). Models in columns 4-5 also include the estimated residual from regression model in column 3 (with CZ and industry effects), fit to Sample B. Model in column 5 also includes share of CZ employment in the industry. Models are fit to person-quarter data after assigning estimated CZ-industry effect to each person-quarter observation.

Table 6: Decomposition of Variance of Average CZ Earnings Premium

	All CZ's	Top 50 CZ's Only	
	2-digit industries (1)	2-digit industries (2)	4-digit industries (3)
Standard Dev. of Average CZ premium	0.063	0.061	0.062
Decomposition (variance shares):			
Var(Average Earnings Premium)	1.003	0.978	0.958
Var(Composition Effect)	0.015	0.010	0.007
Var(Interaction Effect)	0.008	0.003	0.006
Cov(Earnings Premium, Composition Effect)	-0.027	-0.010	-0.009
Cov(Earnings Premium, Interaction Effect)	0.002	0.025	0.039
Cov(Composition Effect, Interaction)	0.000	-0.006	0.000

Notes: Table shows decomposition of the variance of estimated average CZ wage premium, based on equation (6) in text. Decomposition in column 1 uses main LEHD sample and 24 2-digit industries to define CZ-by-industry effects. Decompositions in columns 2 and 3 are restricted to observations in 50 largest CZ's only. Decomposition in column 2 uses 24 2-digit industries to define CZ-by-industry effects; decomposition in column 3 uses 312 4-digit industries to define CZ-by-industry effects.

Table 7: Components of Wage Gap Between High- and Low-Education Workers

	Regression Models Relating Wage Gap and Components to Mean Log Wage in CZ		Variance Decomposition of Wage Gap	
	Coefficient (1)	Standard error (2)	Std. Dev. or Correlation (3)	Var. Share (4)
Wage gap (high- versus low-education workers)	0.644	(0.051)	0.109	1.000
Components of Wage Gap (column 3 = std. dev.)				
Difference in mean person effects	0.604	(0.049)	0.096	0.782
Difference in covariate indexes	-0.003	(0.005)	0.005	0.002
Difference in mean CZ wage effect:				
Within-industry wage gap	-0.068	(0.028)	0.030	0.075
Industry sorting	0.092	(0.007)	0.016	0.021
Interaction	0.018	(0.002)	0.004	0.001
Total	0.042		--	--
Covariance Terms (column 3 = correlation of terms)				
Cov(Person effects, cov. index)	--	--	-0.127	-0.010
Cov(Person effects, within-industry gap)	--	--	-0.226	-0.109
Cov(Person effects, industry sorting)	--	--	0.755	0.195
Cov(Person effects, interaction)	--	--	0.576	0.039
All other covariance terms	--	--	--	0.004

Notes: Columns 1 and 2 show coefficient and standard error from univariate regression of CZ-specific value of wage gap term identified in row heading on mean log wage CZ (estimated from ACS). Columns 3 and 4 show components of a decomposition of the variance of the estimated CZ wage gap between high-education and low-education workers, based on equation (8) in text.

Table 8: Summary of Relationships Between CZ-level Outcomes and Log CZ Size

	2010-2018 ACS		LEHD	
	Estimated Coefficient (1)	Standard Error (2)	Estimated Coefficient (3)	Standard Error (4)
Alternative Measures of Earnings (All CZ's)				
1. log hourly wage (Winsorized)	0.0717	(0.0025)	--	--
2. log annual earnings (no trim)	0.0826	(0.0030)	--	--
3. log annual/qtrly earnings (earnings > 3800)	0.0694	(0.0027)	0.0765	(0.0092)
Basic Decomposition of Mean Log Earnings (All CZ's)				
4. Mean skill index / mean person effects	0.0118	(0.0015)	0.0505	(0.0095)
5. CZ wage effect	0.0577	(0.0019)	0.0260	(0.0031)
6. Percent of size effect due to skills (row 4/3)	*17.0*		*66.0*	
Measures of Dispersion in Skill Composition (All CZ's)				
7. Share in decile 1 skill/person effs.	0.0014	(0.0004)	-0.0041	(0.0032)
8. Share in decile 10 skill/person effs.	0.0050	(0.0005)	0.0228	(0.0032)
9. Std. dev. of skill/person effs.	0.0106	(0.0006)	0.0288	(0.0019)
Components of Average CZ Wage Effect (CZ's with All Industries)				
10. CZ-average Wage Effect	--	--	0.0317	(0.0035)
11. CZ-specific premium component	--	--	0.0341	(0.0033)
12. CZ-industry composition component			-0.0010	(0.0003)
13. Interaction component	--	--	-0.0015	(0.0005)
Degree of Assortative Matching within CZ (CZ's with All Industries)				
14. Within-CZ skill-match correlation (correl. of person effect and industry effect)	--	--	0.0612	(0.0012)

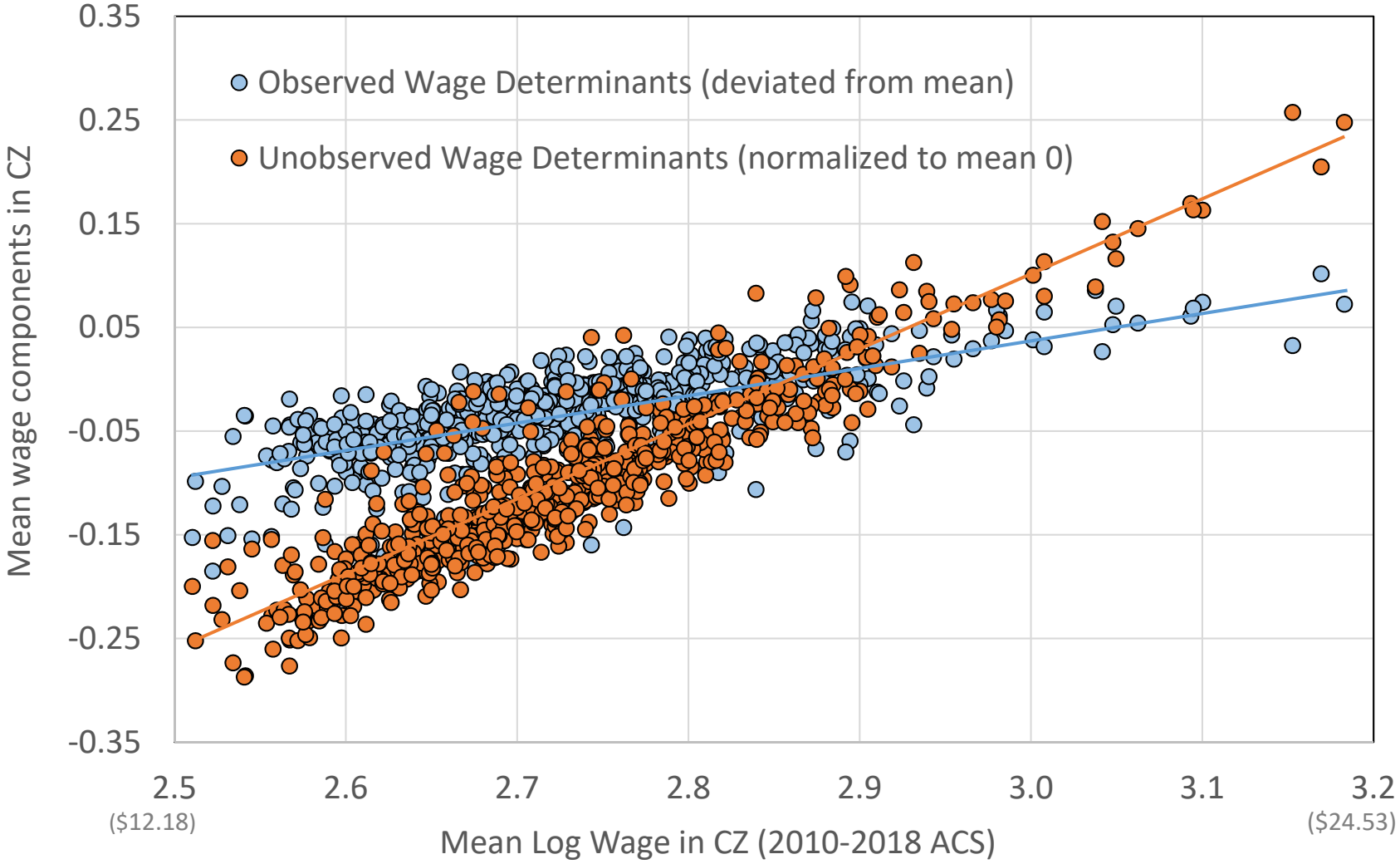
Notes: Each entry is a coefficient from a separate univariate regression of the outcome indicated by the row heading on the log of workforce size in the CZ. "All CZ's" refers to 688 CZ's; CZ's with All Industries refers to a subset of CZ's which have workers in all 24 NAICS industries in all our replication samples. All models are weighed by CZ size. In the ACS, "skill index" is predicted wage or earnings based on a regression model that includes gender, age, education, and country of birth effects. In LEHD "person effects" represent the mean of the estimated person effects for workers in the CZ. See text for explanation of components of average CZ wage effect and within-CZ skill-match correlation.

Table 9: Components of the Return to Education and Log CZ Size

	2010-2018 ACS		LEHD	
	Estimated Coefficient (1)	Standard Error (2)	Estimated Coefficient (3)	Standard Error (4)
a. Outcomes for less-educated workers (education ≤12)				
1. log earnings (earnings > 3800)	0.0152	(0.0018)	0.0386	(0.0056)
2. Mean skill index/person effects	-0.0235	(0.0010)	0.0198	(0.0046)
3. CZ wage effect	0.0388	(0.0017)	0.0193	(0.0024)
4. Share of size effect due to skills (row 2/1)	*-154.6*		*51.3*	
b. Outcomes for more-educated workers (education>12)				
1. log earnings (earnings > 3800)	0.0857	(0.0027)	0.1020	(0.0083)
2. Mean skill index/person effects	0.0168	(0.0014)	0.0744	(0.0087)
3. CZ wage effect	0.0690	(0.0019)	0.0287	(0.0031)
4. Share of size effect due to skills (row 2/1)	*19.6*		*72.9*	
c. Earnings Gap between more and less educated workers				
1. gap in log earnings	--	--	0.0612	(0.0030)
2. Mean skill index/person effects	--	--	0.0540	(0.0050)
3. CZ wage effect components				
Relative wage premium component	--	--	-0.0018	(0.0029)
Compositional component	--	--	0.0075	(0.0007)
Interaction component	--	--	0.0015	(0.0001)

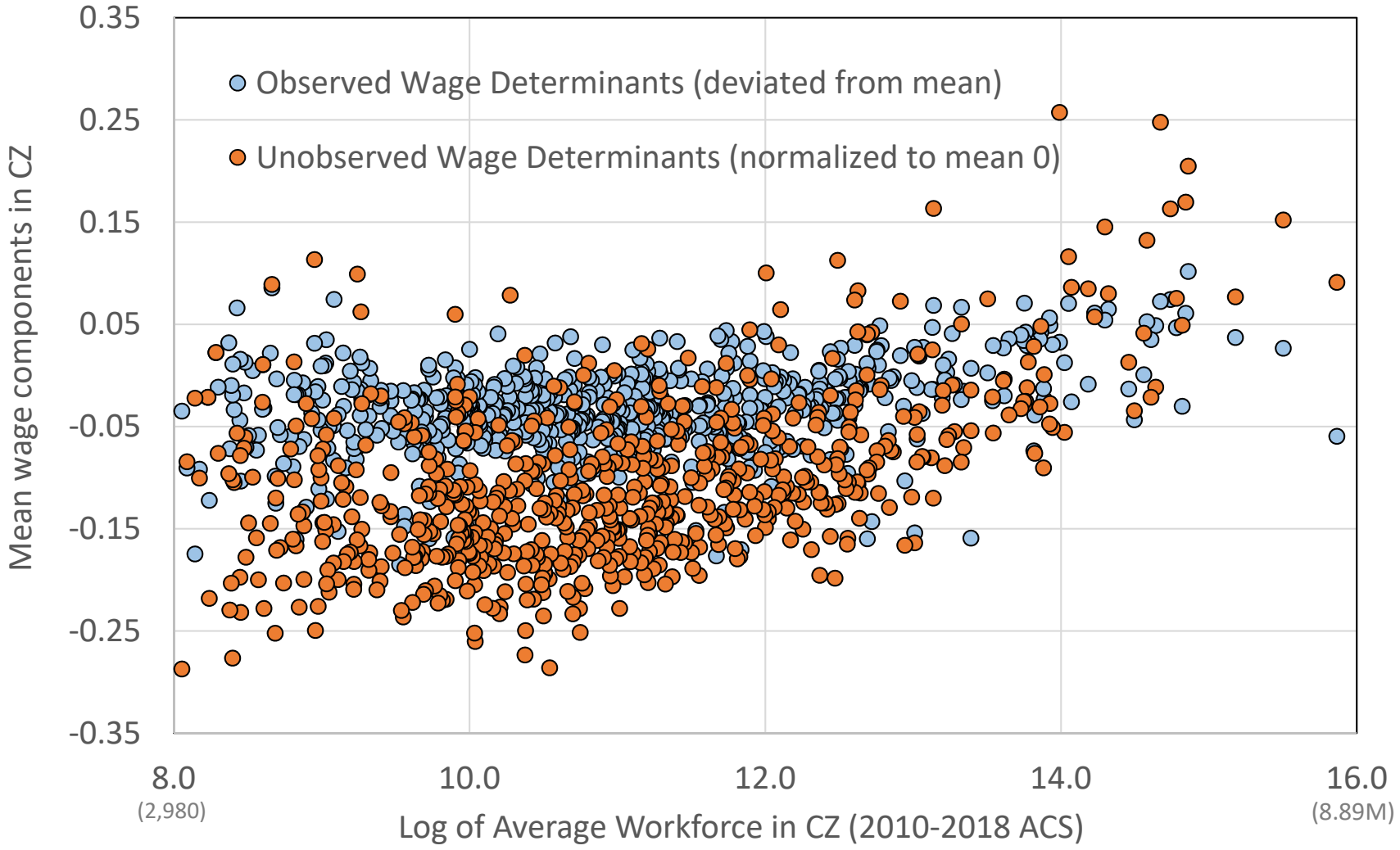
Notes: See notes to Table 8. Each entry is a coefficient from a separate univariate regression of the outcome indicated by the row heading on the log of workforce size in the CZ. The sample includes 688 CZ's; all models are weighed by CZ size. In the ACS, "skill index" is predicted wage or earnings based on a regression model that includes gender, age, education, and country of birth effects. In LEHD "person effects" represent the mean of the estimated person effects for workers in the CZ. For ACS, all estimates are based on log annual earnings, dropping individuals who earned less than $3800 in the previous year. For LEHD, all estimates are based on log quarterly earnings, dropping quarters in which the individual earned less than $3800, and all transitional quarters -- see text.

Figure 1: Observed and Unobserved Components of CZ Mean Log Wages, 2010-2018 ACS



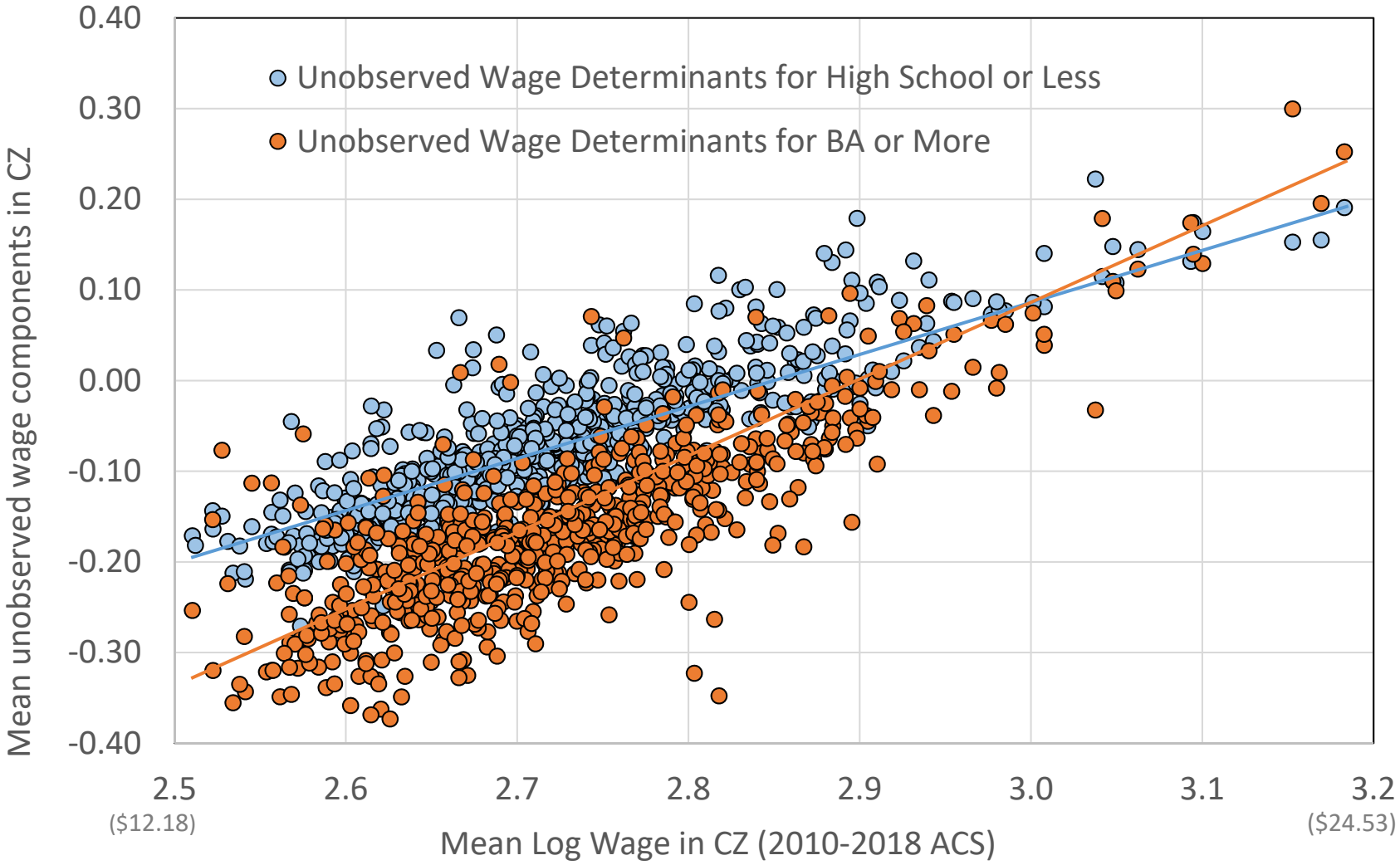
Note: sample contains 688 CZ's based on 1990 CZ definitions (Alaska is excluded). Observed wage determinants represent fitted values from a regression model of log hourly wages that includes education, experience, gender, race/ethnicity, country of origin, and CZ effects, and are deviated from their weighted mean. Unobserved wage determinants are estimated CZ effects from the regression model. Weighted mean of CZ effects is constrained to 0. Fitted lines shown are from weighted OLS regressions.

Figure 2: Observed and Unobserved Components of CZ Mean Log Wages, 2010-2018 ACS



Note: sample contains 688 CZ's based on 1990 CZ definitions (Alaska is excluded). Observed wage determinants represent fitted values from a regression model of log hourly wages that includes education, experience, gender, race/ethnicity, country of origin, and CZ effects, and are deviated from their weighted mean. Unobserved wage determinants are estimated CZ effects from the regression model. Weighted mean of CZ effects is constrained to 0. Fitted lines shown are from weighted OLS regressions.

Figure 3: Unobserved Components of CZ Mean Log Wages: High School or Less vs. BA or More, 2010-2018 ACS



Note: sample contains 688 CZ's based on 1990 CZ definitions (Alaska is excluded). Unobserved wage determinants represent estimated CZ effects from regression models of log hourly wages that include education, experience, gender, race/ethnicity, country of origin, and CZ effects. Models are fit separately for workers with ≤12 years of schooling and workers with ≥16 years of schooling. Weighted means of CZ effects are constrained to 0 in both models. Fitted lines shown are from weighted OLS regressions.

Figure 4: Mean Earnings Before and After a Change of CZ's

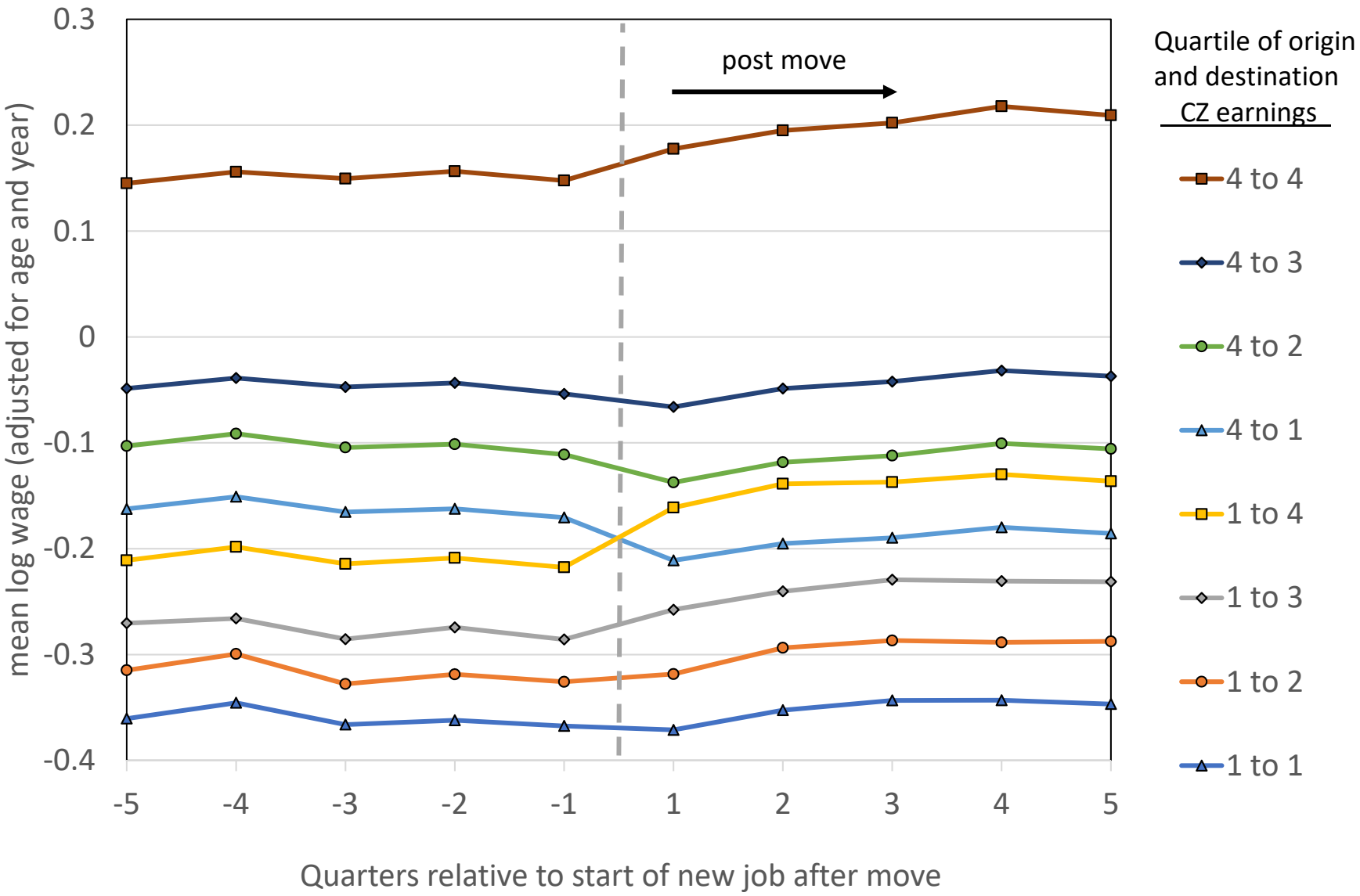
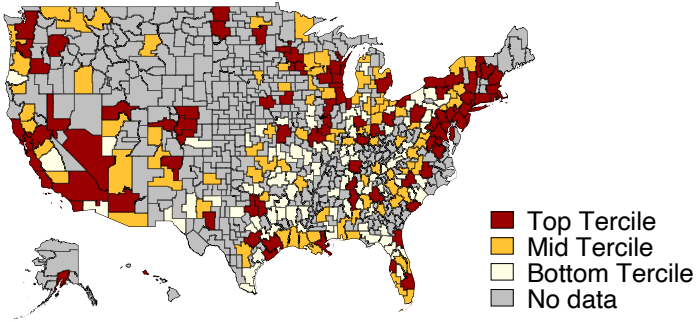


Figure 5: Map of CZ's by Terciles of Wages and CZ Effects

a. Wages (ACS)



b. CZ Effects (LEHD)

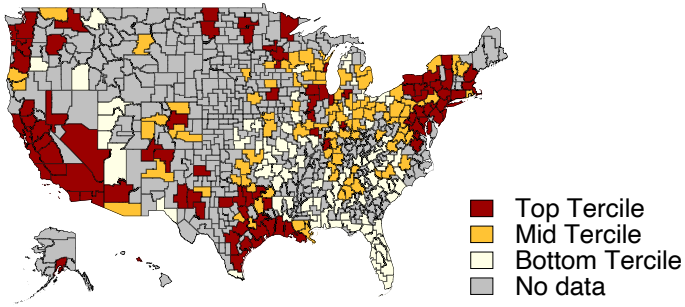


Figure 6: Mean Earnings Before and After a Change of CZ's

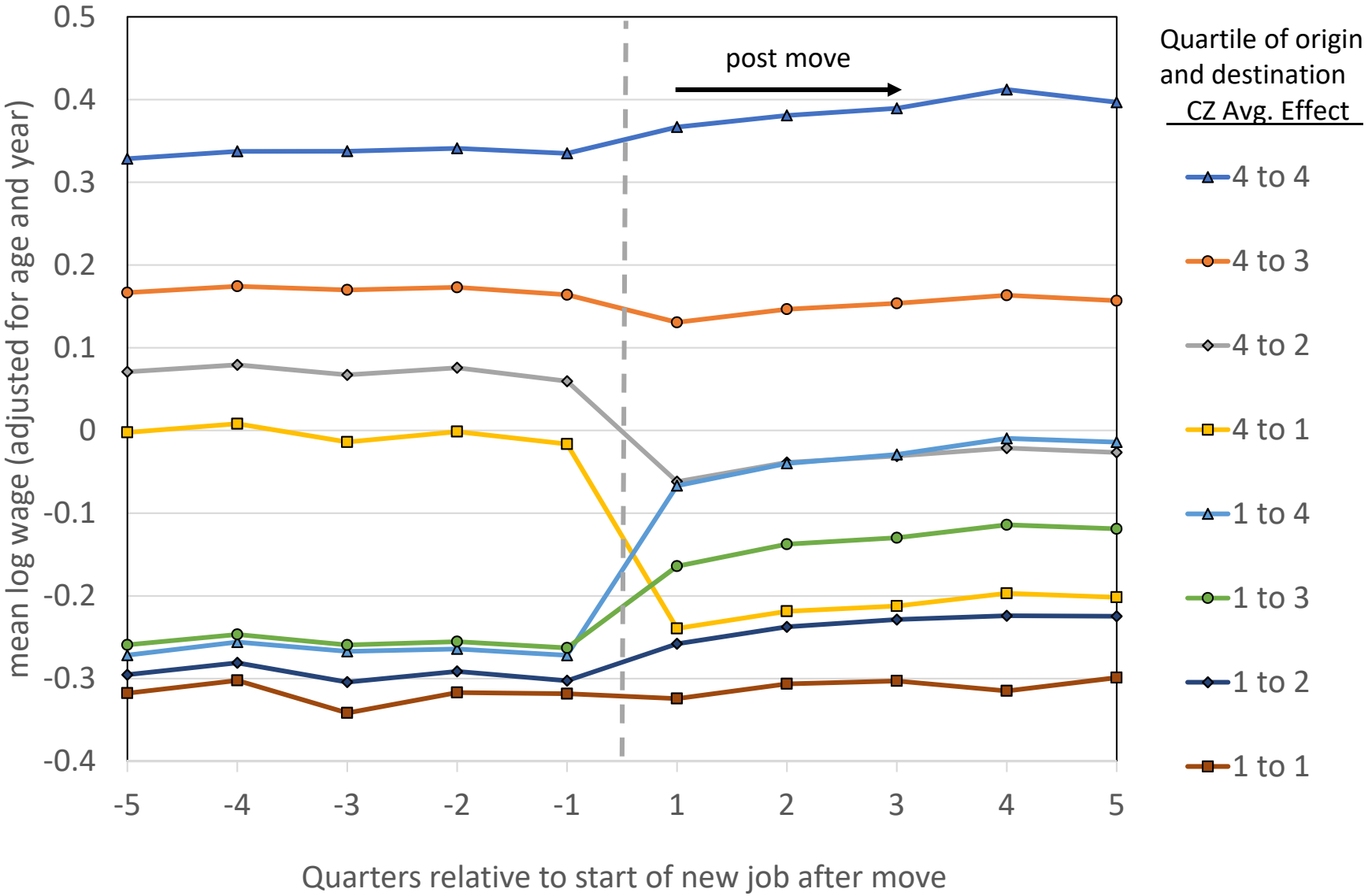


Figure 7: Predicted and Actual Changes in Wages for CZ Movers, by Origin and Destination Vingtile of Average CZ Effect

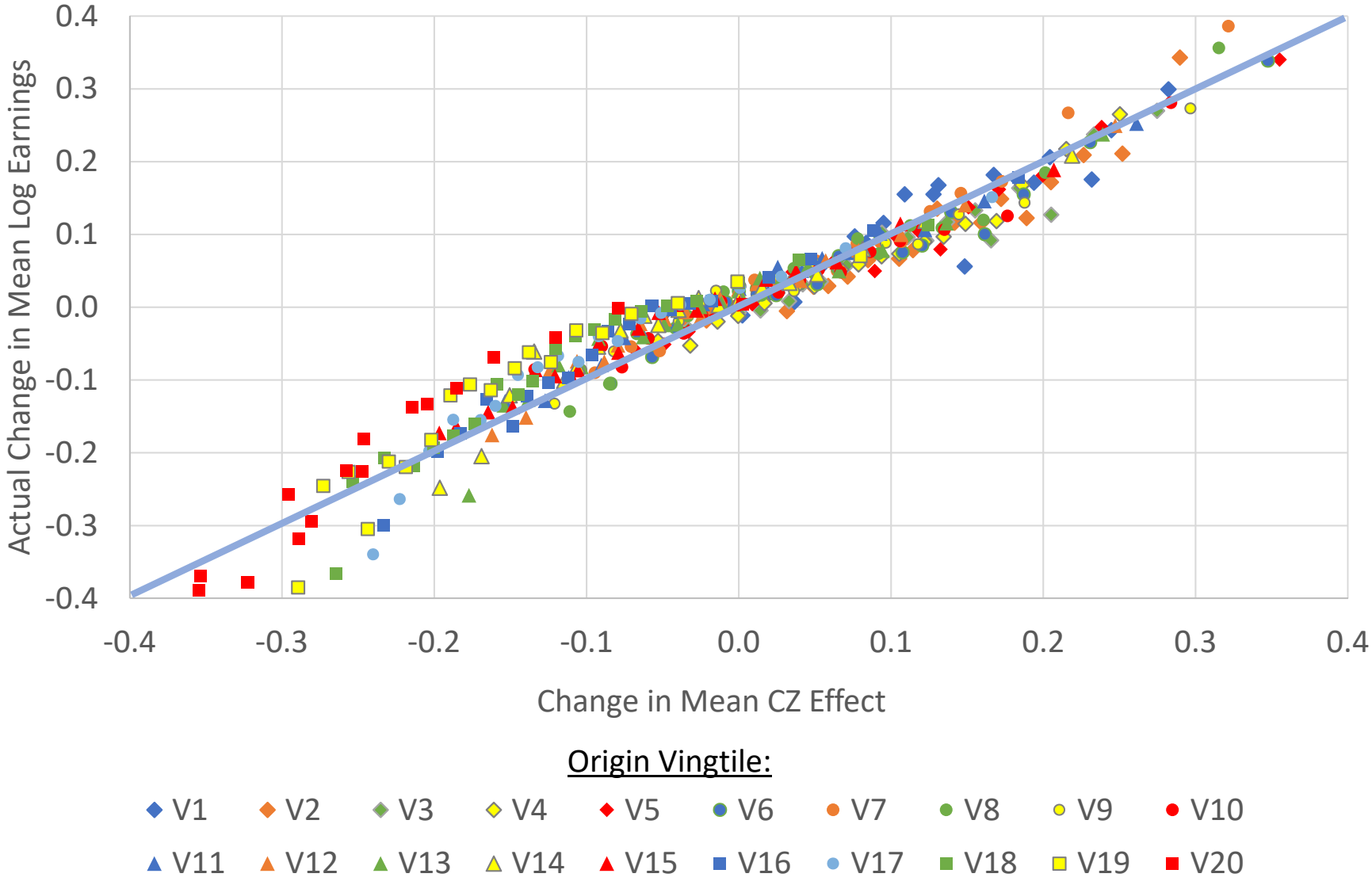
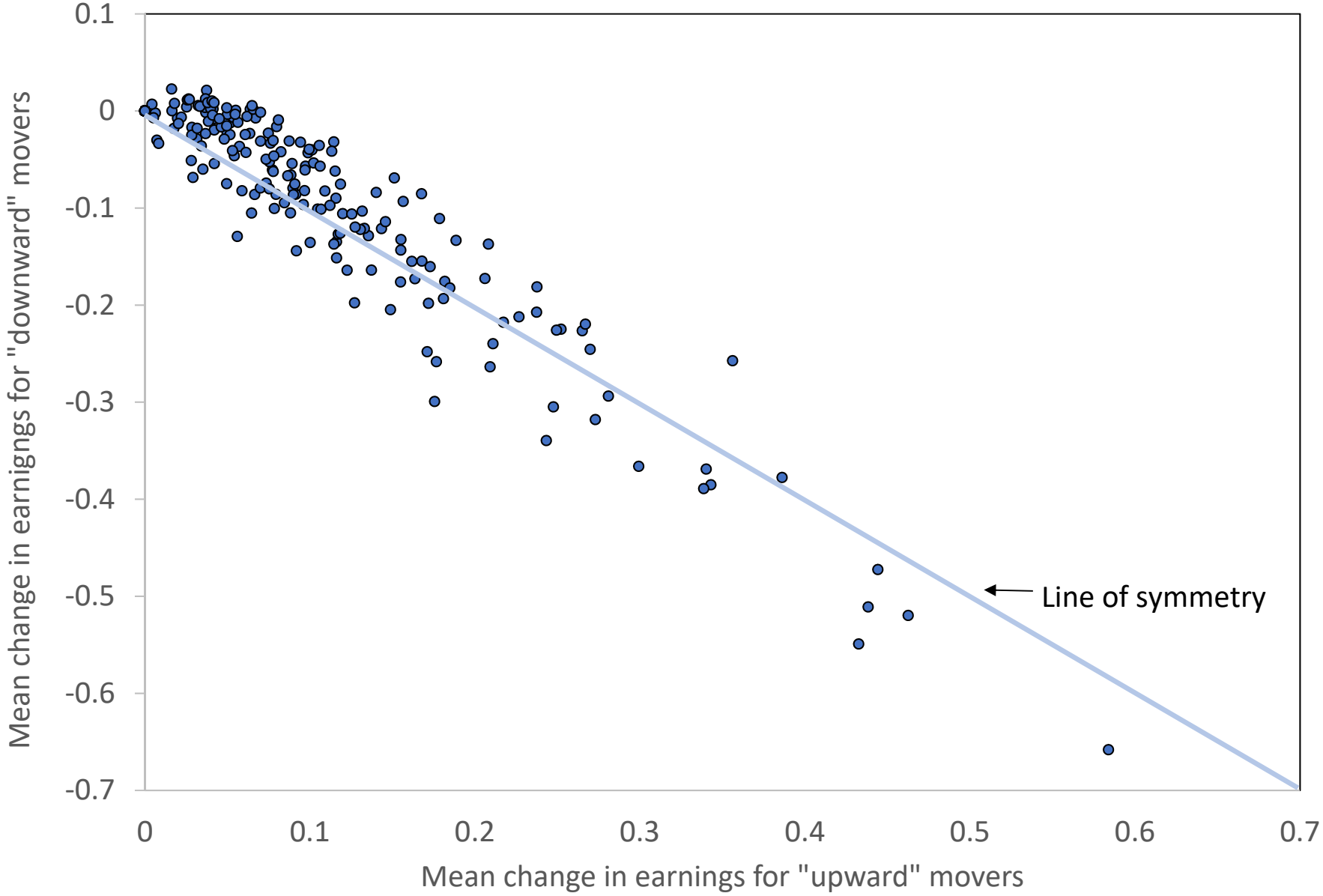
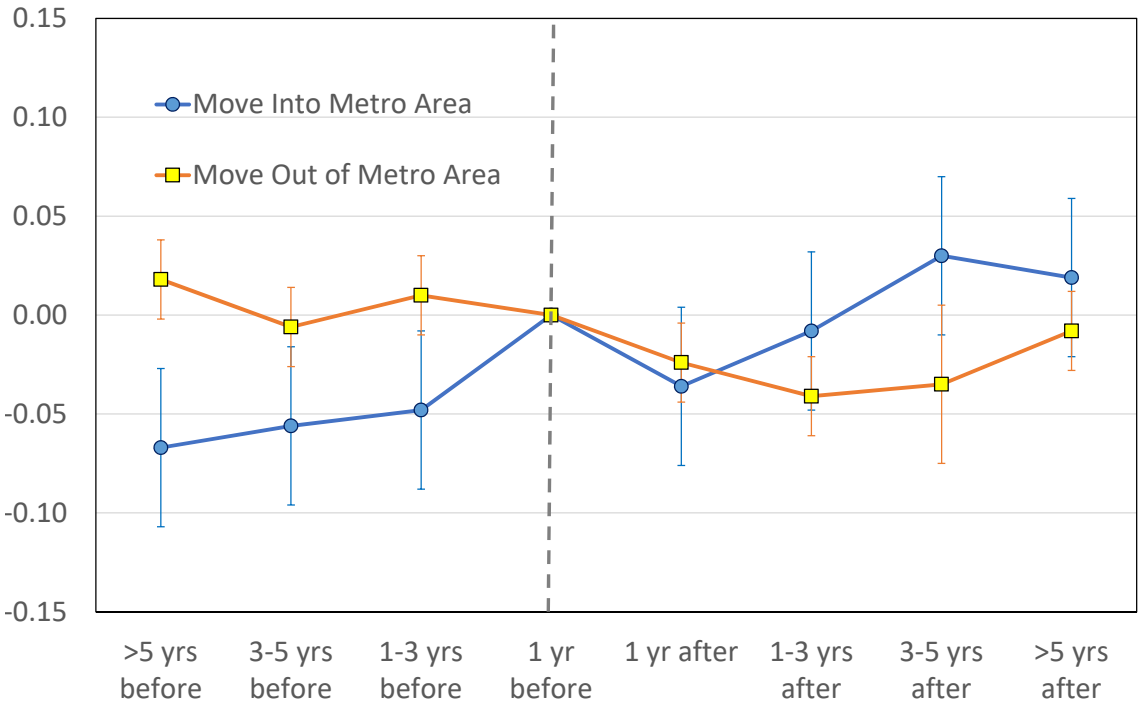


Figure 8: Evaluation of Symmetry for CZ changers

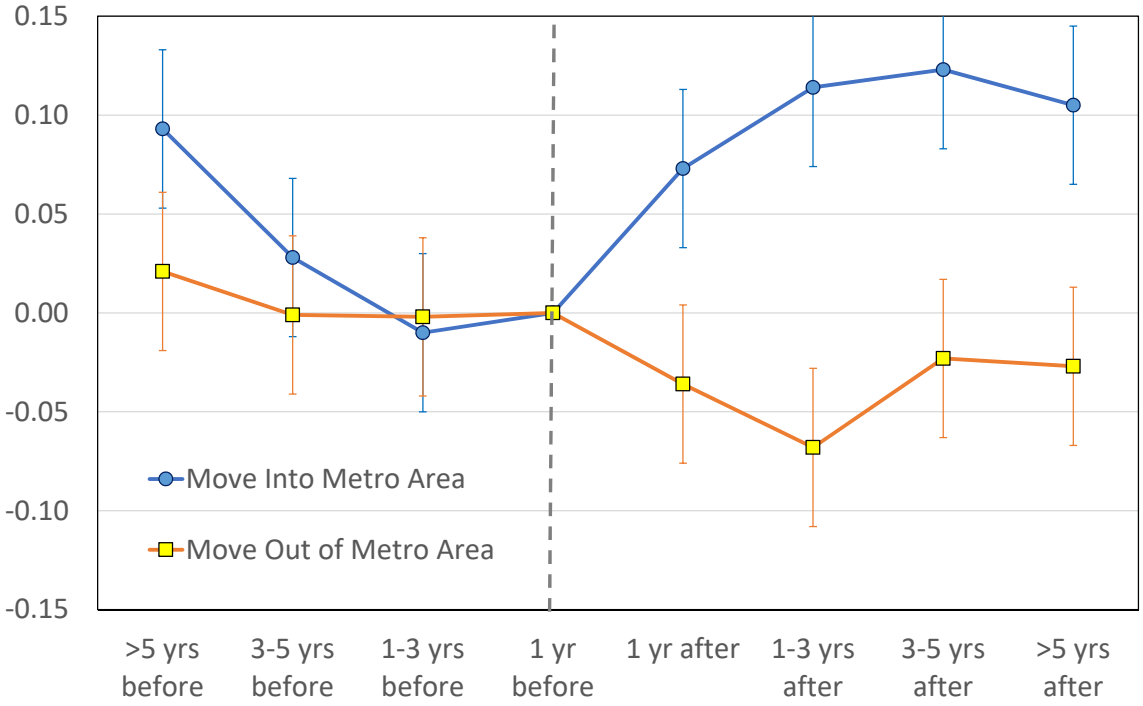


Appendix Figure 1: Wage Changes for Movers in and out of Metro Areas (Glaeser and Mare, 2000)

A. PSID Data

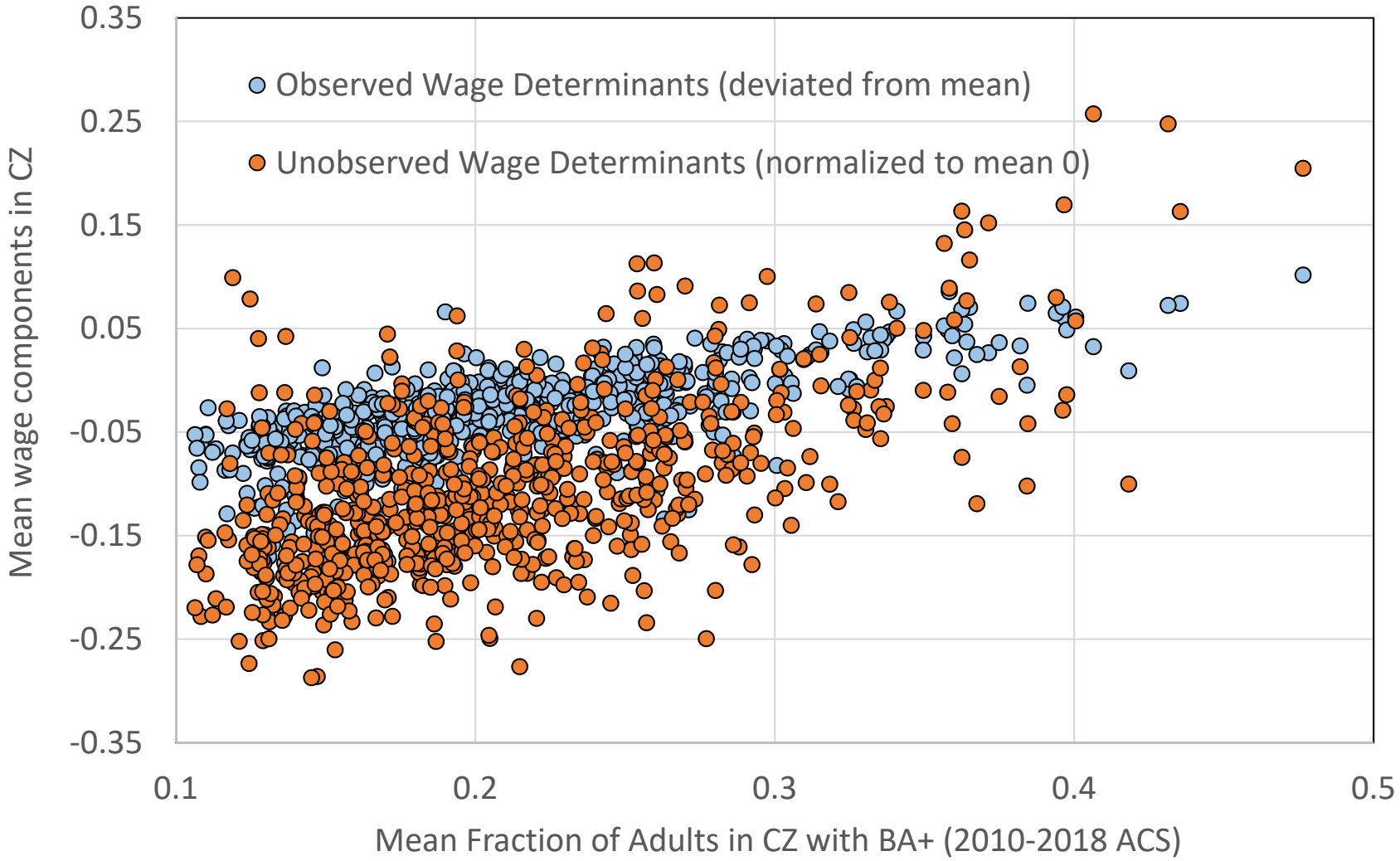


B. NLSY Data



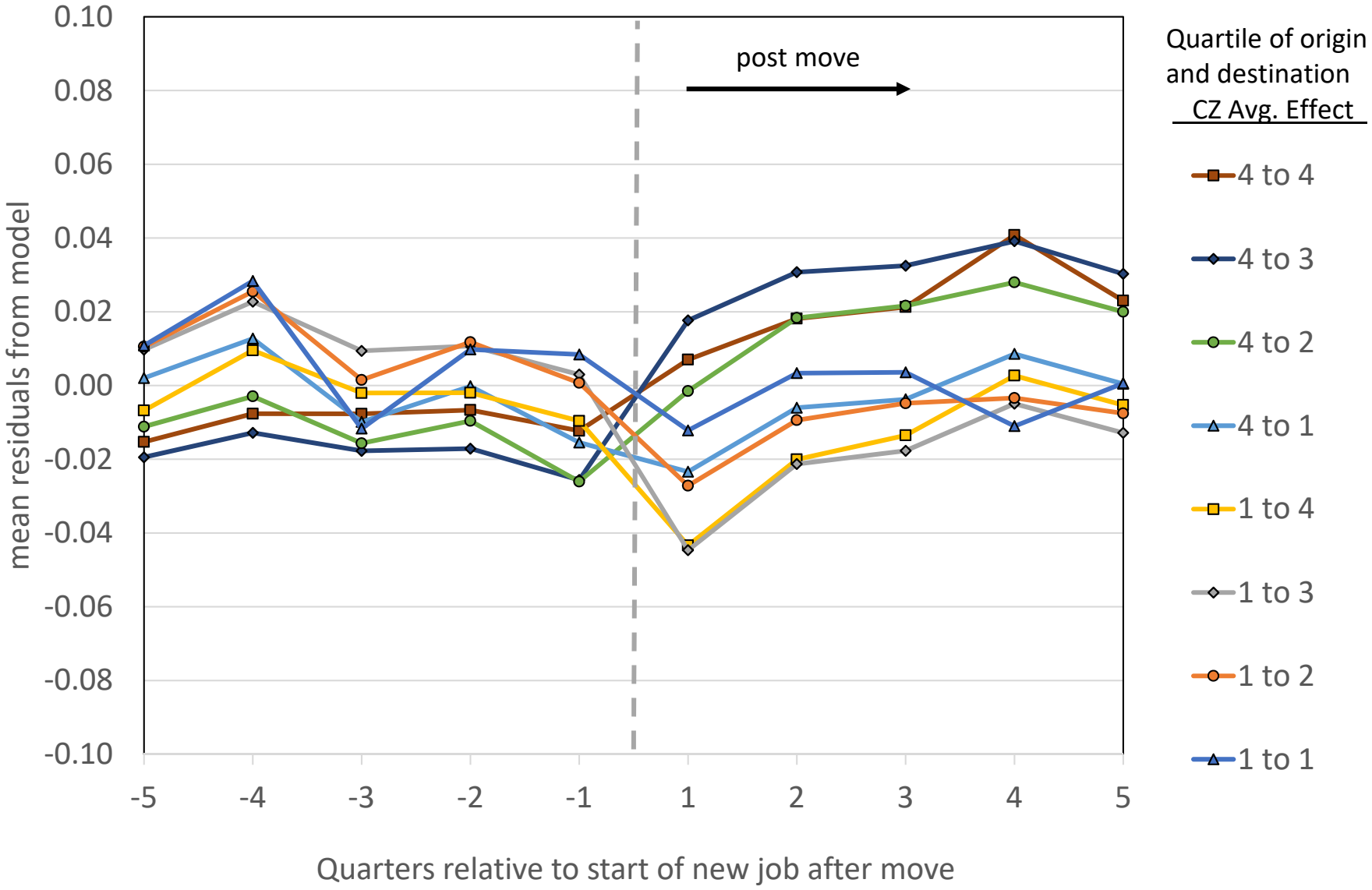
Note: from Glaeser and Mare (2000, Table 5, columns 2 and 4). 95% confidence intervals shown with vertical bars.

Appendix Figure 2: Relation of Observed and Unobserved Components of CZ Mean Log Wages to CZ College Share

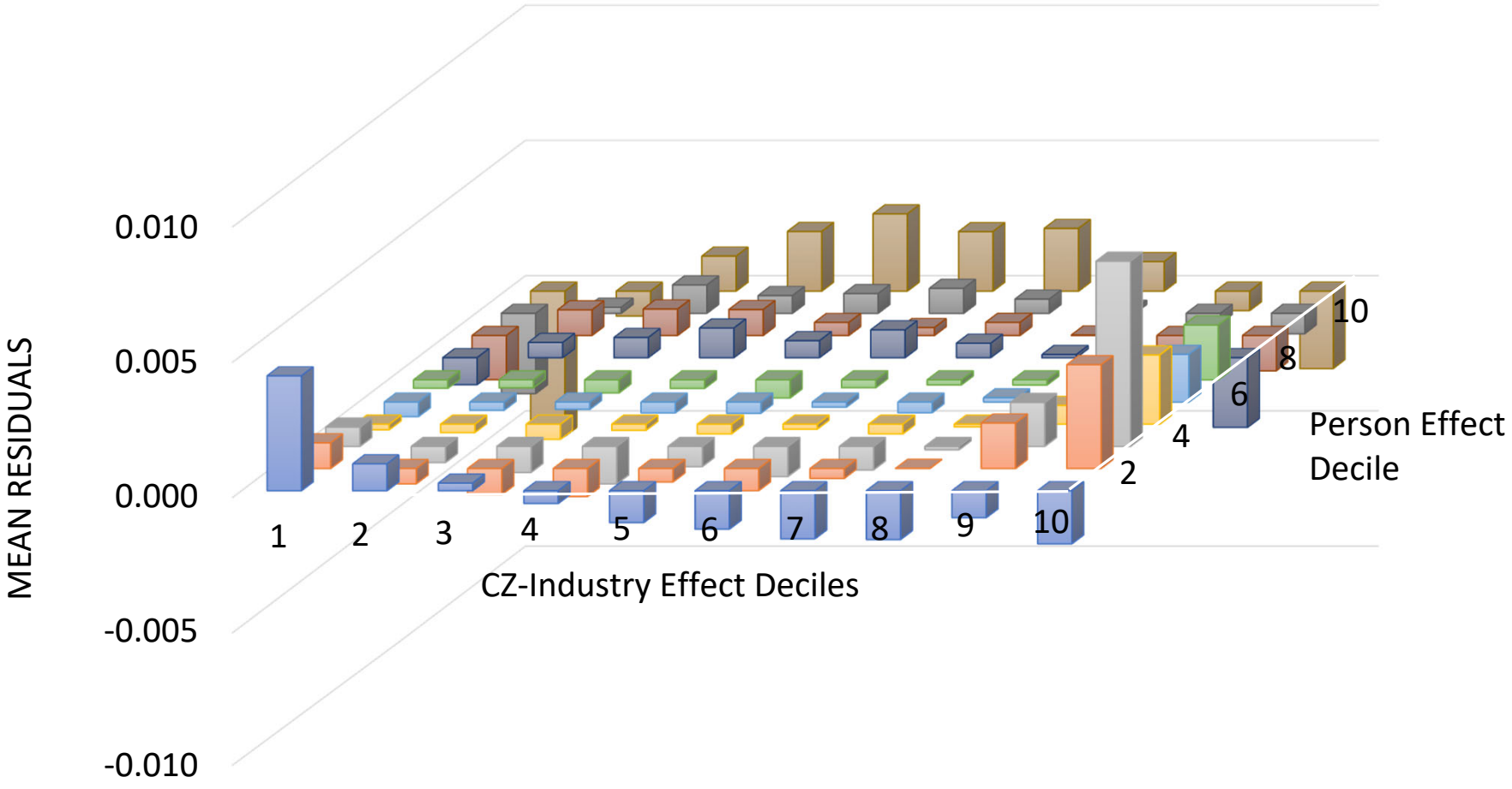


Note: sample contains 688 CZ's based on 1990 CZ definitions (Alaska is excluded). Observed wage determinants represent fitted values from a regression model of log hourly wages that includes education, experience, gender, race/ethnicity, country of origin, and CZ effects, and are deviated from their weighted mean. Unobserved wage determinants are estimated CZ effects from the regression model. Weighted mean of CZ effects is constrained to 0.

Appendix Figure 3: Mean Residuals Before and After a Change of CZ's



Appendix Figure 4: Mean Residuals by Decile of Person Effect and CZ/Industry Effect



Appendix Table 1: Comparison of Variance Decompositions of Two-Way Fixed Effects Models: Plug-In versus Cross-Sample Methods

	Individual Level				CZ-industry Level				CZ Level			
	Plug-in		Cross-sample		Plug-in		Cross-sample		Plug-in		Cross-sample	
	Std Dev or Correl.	Var. Share	Std Dev or Correl.	Var. Share	Std Dev or Correl.	Var. Share	Std Dev or Correl.	Var. Share	Std Dev or Correl.	Var. Share	Std Dev or Correl.	Var. Share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Overall wage/mean wage	0.654	1.000	0.654	1.000	0.275	1.000	0.275	1.000	0.145	1.000	0.145	1.000
Variance components												
Person effects	0.561	0.736	0.561	0.736	0.200	0.528	0.199	0.526	0.100	0.475	0.100	0.472
CZ-industy effects	0.097	0.022	0.097	0.022	0.097	0.126	0.097	0.125	0.064	0.195	0.064	0.196
Covariate index (Xβ)	0.150	0.052	0.150	0.053	0.015	0.003	0.015	0.003	0.005	0.001	0.005	0.001
Residual	0.243	0.138	0.243	0.138	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Covariance components												
Person/CZ-industry	0.209	0.053	0.211	0.054	0.585	0.302	0.593	0.308	0.558	0.339	0.563	0.347
Person/Covariates	-0.010	0.004	-0.010	-0.004	0.389	0.031	0.387	0.032	-0.024	-0.001	-0.029	-0.001
CZ-industy /Covariates	0.026	0.002	0.026	0.002	0.256	0.010	0.256	0.010	-0.293	-0.009	-0.297	-0.009

Notes: See note to Table 3. Table shows variance decompositions based on equation (3) using Plug-in method or cross-sample method. Columns 1-4 pertain to the variance of individual quarterly earnings. Columns 5-8 pertain to the variance of mean earnings by CZ-industry cell. Columns 9-12 pertain to the variance of mean earnings by CZ. Entries in odd-numbered columns for "variance components" are estimated standard deviations of earnings component indicated in row heading; entries in odd-numbered columns for "covariance components" are the estimated correlations of the indicated variance components. Entries in even-numbered columns are variance shares explained by variance or covariance components.

Appendix Table 2: Mean Log Earnings and Components of Earnings for Groups of CZ's

		Ranked By Mean CZ Effect (ψ_c)				Grouped By Tercile of Mean CZ Effect (ψ_c)		
		Top 10		Middle				
	All	Large Urban	Resource	Range	Bottom 10	Lowest	Middle	Top
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mean Log Earnings	9.32	9.70	9.46	9.32	9.18	9.24	9.32	9.41
	(0.11)	(0.11)	(0.07)	(0.10)	(0.08)	(0.08)	(0.07)	(0.12)
Mean Person Effect	9.18	9.40	9.19	9.18	9.13	9.15	9.18	9.21
(normalized)	(0.08)	(0.06)	(0.06)	(0.08)	(0.08)	(0.07)	(0.07)	(0.09)
Mean CZ-industry Effect	0.09	0.24	0.23	0.08	0.00	0.04	0.08	0.14
(normalized)	(0.05)	(0.04)	(0.03)	(0.04)	(0.01)	(0.02)	(0.01)	(0.04)
Mean Covariates	0.05	0.05	0.04	0.05	0.05	0.05	0.05	0.05
	(0.01)	(0.00)	(0.02)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)

Notes: standard error of mean in parentheses. CZ-industry effects are normalized to have mean 0 in bottom 10 CZ's.

Appendix Table 3: Summary of Estimated Two-Way Fixed Effects Models for Low and High Educatic

	Person-quarter level		CZ-industry level		CZ level	
	Std. Dev. or Correlation	Var. Share	Std. Dev. or Correlation	Var. Share	Std. Dev. or Correlation	Var. Share
	(1)	(2)	(3)	(4)	(5)	(6)
A. Low Education (No More than High School) Subsample						
Log earnings or mean log earnings	0.518	1.000	0.229	1.000	0.098	1.000
Variance components (std. deviations in odd-number columns, variance shares in even-numbered columns)						
Person effects	0.445	0.737	0.157	0.470	0.056	0.334
CZ-industry effects	0.101	0.038	0.101	0.196	0.060	0.379
Covariate index (Xβ)	0.110	0.045	0.011	0.002	0.005	0.003
Residual	0.202	0.153	0.000	0.000	0.000	0.000
Covariance components (correlations in odd-number columns, variance shares in even-numbered columns)						
Person/CZ-industry	0.193	0.065	0.548	0.333	0.422	0.300
Person/Covariates	-0.102	-0.037	-0.024	-0.002	-0.148	-0.009
CZ-industy/Covariates	0.000	0.000	0.006	0.000	-0.102	-0.007
B. High Education (Some College or More) Subsample						
Log earnings or mean log earnings	0.685	1.000	0.282	1.000	0.180	1.000
Variance components (std. deviations in odd-number columns, variance shares in even-numbered columns)						
Person effects	0.599	0.764	0.210	0.557	0.131	0.531
CZ-industry effects	0.095	0.019	0.095	0.114	0.066	0.135
Covariate index (Xβ)	0.132	0.037	0.011	0.002	0.005	0.001
Residual	0.256	0.140	0.000	0.000	0.000	0.000
Covariance components (correlations in odd-number columns, variance shares in even-numbered columns)						
Person/CZ-industry	0.225	0.054	0.640	0.322	0.646	0.346
Person/Covariates	-0.042	-0.014	0.070	0.004	-0.171	-0.007
CZ-industy/Covariates	0.005	0.000	0.063	0.002	-0.275	-0.006

Notes: See note to Table 3. Table shows variance decompositions based on equation (3). Columns 1-2 pertain to the variance of individual quarterly earnings. Columns 3-4 pertain to the variance of mean earnings by CZ-industry cell. Columns 5-6 pertain to the variance of mean earnings by CZ. Entries in columns 1-3-5 for "variance components" are standard deviations of earnings component indicated in row heading; for "covariance components" are the estimated correlations of the indicated variance components. Entries in columns 2-4-6 are variance shares explained by

Appendix Table 4: Elasticities of Housing Values and Rents w.r.t. CZ Size and Mean Wages

	All CZ's		**Largest 50 CZ's**	
	Elasticticity w.r.t.:		Elasticticity w.r.t.:	
	log size	mean log wage	log size	mean log wage
	(1)	(2)	(3)	(4)
Housing Prices (log of home value for owners)				
Unadjusted	0.25	3.02	0.38	2.84
	(0.01)	(0.08)	(0.08)	(0.42)
Quality Adjusted	0.22	2.70	0.42	2.87
	(0.01)	(0.09)	(0.08)	(0.49)
Monthly Rent (log of rent for renters)				
Unadjusted	0.17	1.80	0.19	1.25
	(0.01)	(0.05)	(0.04)	(0.24)
Quality Adjusted	0.18	1.93	0.23	1.49
	(0.01)	(0.05)	(0.04)	(0.24)

Note: Table entries are regression coefficients (and standard errors) from weighted OLS regressions of CZ-average housing price measure in row heading on constant and log of number of workers in CZ (columns 1,3) or log of mean log wage in CZ (columns 2,4). Regressions are weighted by number of workers in CZ. Sample in columns 1-2 is set of 678 CZ's in 2018 5-year ACS with non-missing data. Sample in columns 3-4 is 50 largest CZ's, ranked by number of workers.