Appendices to "Location, location, location"
David Card, Jesse Rothstein, and Moises Yi
August 2023

Appendix A. Samples

1. ACS Worker Level Files

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, constructing an estimated hourly wage from data on total earnings in the previous year, hours per week, and information on weeks worked. To limit the effect of outliers we Winsorize the hourly wage at $5 and $500.

We assign local labor markets based on the 1990 CZ boundaries developed by Tolbert and Sizer (1996), which define 741 CZs, each comprised of one or more whole counties. The lowest level of geography in the ACS is the public use micro area (PUMA), which can contain multiple counties (or parts of counties) in sparser 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 CZs. We pool all CZ's in Alaska, yielding a total of 691 commuting zones. We note that the size distribution of CZs is highly skewed: The 50 largest CZs have nearly 60% of workers, and the 200 largest have over 85%.

Our sample has 11.7 million workers, providing relatively large samples for even modest-sized CZ's. We have around 10,000 observations for the CZ's ranked at roughly 200th in size (e.g., Binghamton, NY; Morgantown, WV; and Byron, TX), and 3,000-3,500 for CZ's ranked at roughly 400th in size.

The mean of log nominal hourly wages in our ACS sample is 2.863 (about $17.50), and the (weighed) 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 (weighted) regression of mean log hourly wages on log size is 0.068 with a robust standard error of 0.01 (Appendix Table 1). Log size explains about 50% of the cross-CZ variance in mean log wages. It is also highly correlated with the share of college-educated workers, the fraction of immigrants, and the share of white non-Hispanic workers.

We construct three measures of skill-adjusted CZ effects using a (weighted) regression of log wages on CZ effects and controls. In the first (basic) model, the model includes 18 individual controls (education, a quartic of experience interacted with gender, and dummies for race/ethnicity, interacted with gender, along with year effects. This model has an R-squared of 0.3183, and a root mean squared error of 0.6029.

Our second model generalizes this by allowing dummies for individual years of completed education (interacted with gender, immigrant status, and whether an immigrant's years in the US is less than their potential experience + 5 (implying that the probably completed some years of schooling in the U.S.), as well as dummies for 3 major immigrant source regions (Latin America, Asia, and Europe/Canada/Australia/New Zealand) and controls for years in the U.S. for immigrants (interacted with source region). This model has an R-squared of 0.3490, and a root mean squared error of 0.5892.

Our third model generalizes the second model by adding controls for field of degree for people with a BA or higher education (in 16 main categories), interacted with gender, and controls for detailed industry, based on Census industry codes used in the ACS (a total of 267 codes). We adjust the coding of Census industry across years of the ACS to the 2018 coding system. This model has an R-squared of 0.400, and a root mean squared error of 0.5654.

We also fit two extra versions of this third model, using data on people with exactly 12 and exactly 16 years of education. The high school sample has 3.755 million observations; the R-squared of the model is 0.253 and the root mean squared error is 0.5296. The college sample has 2.516 million observations; the R-squared of the model is 0.302 and the root mean squared error is 0.5875.

Our wage models are limited to people who were employed in the reference year. As a final exercise, we fit versions of the second model to the full ACS sample, including non-workers, where the dependent variables are an indicator for annual employment and the continuous number of hours worked in the year (including non-workers as zeros). We use these below to explore geographic variation in employment and hours.

2. ACS Household Level File (For House Price/Rent Analysis)

Our ACS household file is formed from the household records of the 2014-2018 5-year ACS public use sample. We extract all non-group-quarters records, providing a sample of 6.804 million household records. We assign CZs using the same procedures described above, using PUMA codes and information from the 2010 Census to probabilistically allocate households to counties and CZs. The sample has 692 separately identified CZs. We adjust reported property values and rents to real values using the deflator variable ("ADJINC") provided in the ACS. We set reported property values in the interval of $1,000 to $4,500 to $4,500, and drop observations with reported value below $1,000. Similarly we drop rental values of less than $100 per month. This leaves 4.374 million observations on property values (for households that own their home), and 1.836 million observations on rents (for households that rent). The mean of log housing value is 12.14 (implying a geometric mean of $187,213) with a standard deviation of 1.03. The mean of log monthly rents is 6.70 (implying a geometric mean of $815) with a standard deviation of 0.67.

To obtain adjusted property values and rents, we fit simple models for the logs of property values and log of monthly rents that include CZ effects and housing unit characteristics. For property values, the controls include: dummies for whether the unit is a mobile home, a single attached home, or a unit in buildings in 6 different size ranges, controls for the number of bedrooms (5 dummies), the log of the total number of rooms, dummies for the year the unit was built (in 22 ranges), and a dummy for whether the homeowner has a mortgage. This model has an R-squared of 0.50 and a root mean squared error of 0.70. For rents the controls include the same characteristics, plus indicators for whether electricity, gas, water, other fuels, or meals are included in the rent (a total of 5 indicators). This model has an R-squared of 0.34 and a root mean squared error of 0.52.

3. LEHD

We discuss here a few details of the LEHD sample that were not discussed fully in the text.

One is that the unemployment insurance records from which the LEHD is constructed identify the firm at which a worker is employed and the state, but not the specific establishment. Only one state reports establishment locations. The Census Bureau uses this state's data to train a model that it then uses to impute establishment assignments probabilistically to data from other states, using the distance between the worker's residential location and each of the establishments of the employing firm (Vilhuber 2018). Ten imputations are reported for each PEQ. When the firm has only a single establishment within reasonable commuting distance of the worker, all ten of these are identical.

Our analysis uses establishments both for the AKM model and to assign CZs and industries. We classify PEQs based on whether there is variation in the implied CZ or industry across the ten imputations. We present an analysis below of the no-uncertainty sample.

Section VII segments the sample by worker education (high school or less, vs. some college or more). Although education is imputed for most workers in the LEHD, these imputations carry some error. As an alternative, we take advantage of linkages that the Census Bureau has created between LEHD records and the 2001-17 ACS files. Any individual captured in one of these survey samples was asked about his or her completed education. Our education analyses restrict attention to observations with linked education information from one of these surveys, using only survey responses when respondents were at least 30 years old. This yields about 15% of the full LEHD sample with education information.

Appendix B. Additional results

1. Additional ACS results

Section II of the paper discusses four stylized facts that we derive from our analysis of the ACS sample introduced in Appendix A: (1) mean wages vary widely across CZs; (2) only a modest share of this variation is explained by differences in observed characteristics of the workers in different CZs; (3) mean wages are higher in larger CZs; and (4) pay premiums for working in larger cities or in cities with higher average wages are higher for better-educated workers. We present here evidence in support of these claims.

Appendix Figure 2 compares mean log wages in each CZ, on the x-axis, to estimated CZ wage effects from a model that controls for worker observables, on the y-axis. This is our third, most detailed observational model discussed above. The wide spread of points on the x-axis, with a standard deviation of 0.14 and an interquartile range of 0.2, establishes the first stylized fact. The slope of the fitted line on the figure, 0.69, demonstrates that less than one-third of this variation is explained by observed worker characteristics. The standard deviation of the adjusted CZ effects is 0.106, about 75% as large as the variation in unadjusted mean wages. In Appendix Figure 2 we plot the adjusted CZ effects against the mean wages in each CZ. The slope is 0.692, suggesting that only about 30% of the variance of CZ wages can be explained by even a rich set of observed worker and job characteristics.

Appendix Figure 3 compares unadjusted CZ mean wages and adjusted CZ effects to log CZ size. Both unadjusted means and adjusted CZ effects are higher in large CZs, with elasticities of 0.068 and 0.056, respectively. Appendix Table 1 presents regressions of unadjusted and adjusted wages on CZ size, as well as similar estimates for annual earnings. Elasticities of earnings with respect to size are very similar to those of wages, suggesting that labor supply does not vary dramatically across CZs (at least among those who work).

Appendix Figure 4 compares our adjusted CZ effects fit separately for workers with 12 and with 16 years of education to the estimates from the model that pools all education groups. (All models are normalized to mean zero in the average CZ.) The slope of the 12-year estimates with respect to the pooled estimates is 0.906, while that for the 16-year estimates is 1.209. The implication is that the return to education is larger in higher-premium CZs.

Finally, we used information on annual employment and hours in the ACS to study geographic variation in labor supply. Appendix Table 2 presents results. Rows 1 and 3 show models relating CZ mean labor supply to CZ size (column 1), the unadjusted mean log wage in the CZ (column 2), or the adjusted log wage from the second ACS specification described above (column 3). Both labor supply measures are positively related to CZ size and CZ unadjusted and adjusted wages. In rows 2 and 4, we replace the unadjusted means of the dependent variables with adjusted CZ effects, using the same specification. The adjusted labor supply measures remain positively related to unadjusted mean wages, though the relationship is much weaker. They are not positively related either to CZ size or to adjusted CZ mean wages. Columns 4 through 6 repeat the exercise separately by gender, with similar results for men and women.

We conclude that there is no evidence that employment varies systematically at either the intensive or extensive margin with CZ pay premiums.

2. Evaluating the AKM specification

In this section, we present a number of analyses aimed at validating the restrictions of the AKM specification (2) – exogenous mobility with respect to the error term, and additive separability of the person and establishment effects. These analyses replicate and extend analyses proposed by Card et al. (2013), and are intended in part to allow for comparisons between U.S. data and results seen in other countries.

We begin with an analysis of additive separability. 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 establishments (classified by the estimated values of δ_f). Appendix Figure 5 shows mean residuals by decile of α_i crossed with decile of δ_f. Systematic patterns might point to violations of the assumption that log wages are additively separable in person and firm effects. For example, if log wages were determined by a function $g(\alpha_i, \delta_f)$ with a positive cross-partial derivative, we would expect residuals to be positive at the far right corner of the Figure and negative at the near left corner. We do not see this. Generally, mean residuals are quite small. The main pattern we see is a positive average residual at the near left corner, when the lowest-skill workers are seen at the lowest-premium establishments, and corresponding negative residuals in the near right and far left corners, very high (low) skill workers at very low (high) premium establishments. (Note that residuals must average to zero within each row and column.) This could be consistent with a $g()$ function with a negative cross-partial derivative, or, perhaps, with minimum wage effects that put a floor on the very lowest wages. In any event, the deviation is quite small even at the extremes, and is almost entirely concentrated in the lowest decile of person and firm effects. Thus, violations of additive separability appear to be small.

The main text presented event study plots of mean earnings in the quarters leading up to and following moves across commuting zones (e.g., Figure 3). Evidence that workers who moved had systematically different trends or shocks prior to the move would tend to contradict the exogenous mobility assumption. Appendix Figure 6 presents a different view of this event study, this time plotting the evolution of the equation (2) *residuals*, rather than total earnings, around the move. We see some evidence here that residuals dip by as much as 2% in the first quarter after a move to a new CZ, with the larger dips among the workers moving up the most in the CZ premium distribution. However, this tip is quickly recovered.

Appendix Figure 7 presents a similar event study, this time centered around moves across *establishments*. As before, we show series for moves out of the top- and bottom-quartile of establishments classified by δ_f. Patterns are generally similar to those for CZ moves, though more extreme reflecting the greater variance of establishment effects than CZ means. Earnings are generally quite flat prior to moves, with perhaps a bit of indication that 4-to-1 downward movers had already been experiencing very slow earnings declines prior to the move. Levels are consistent with expectations: Workers in quartile-4 establishments have much higher earnings than those in quartile-1 establishments, while within the quartile-4 group those who will eventually move to lower-premium establishments already have lower earnings prior to the move (which our model interprets as indicating lower α_is) than those who will not. Earnings changes following moves are also consistent with expectations, with larger changes for those with the biggest shifts in establishment premia. As in earlier analyses, we see some indication that quarter-1 earnings are a bit depressed, with most groups rising a bit after that but little indication of differential patterns in this rise.

The right panel of Appendix Figure 7 zooms in by plotting just the residual component of earnings. With the much-enlarged scale, we can see better that the model assumptions do not seem to hold exactly. Workers who make big upward moves to higher-δ_f firms see non-trivial declines in their residuals in the first quarter following the move, suggesting that they do not receive the full expected wage increase immediately. These residuals disappear fairly quickly afterward; by the 4th quarter after the move, residuals for all of the groups have settled into the same 0.05 range that they were in two quarters before the move. The ranking within this range has changed slightly, but the range is small enough not to leave much scope for match effects.

(Recall that the standard deviation of estimated establishment effects is 0.22, so a residual shift of, say, 0.03 is not large in comparison.)

Appendix Figure 8 repeats the exercise from Figure 4, this time for between-establishment rather than between-CZ moves. We divide establishment into vingtiles based on their estimated earnings premiums, and plot the average change in log earnings (removing the contribution of $X_{it}\hat{\beta}$ to obtain age-adjusted earnings) from period -1 to period 1 for movers in each of the 400 possible origin and destination cells against the predicted change in δ_f given the origin and destination vingtiles. This is a more direct test of the AKM specification than was Figure 4, as equation (2) specifies that age-adjusted earnings are $y_{it} - X_{it}\hat{\beta} = \delta_{f(i,t)} + \varepsilon_{it}$, so only systematic changes in the residual could produce deviations from the 45-degree line.

The scatterplot shows that changes in earnings are generally quite close to that line, with a slope of about 0.87. What deviations there are indicate that movers see somewhat smaller changes in wages (in either direction) than would be predicted from the change in firm effects. Panel B shows the same plot, this time with residuals. It is easier to see the deviations here: People who move to firms where the premium increases (decreases) by about 0.4 see their residuals decrease (increase) by about 0.1 in the first quarter after the move, with smaller changes for more extreme moves.

Appendix Figure 6 indicated that residual changes in the first quarter after a move are particularly extreme. When we repeat the analysis in Appendix Figure 8, panel B, using the change in residuals between the last pre-move quarter and the *fourth* post-move quarter, the slope diminishes, to just -0.08. Thus, while there is some evidence that match effects or residuals offset some of the differences in pure firm effects, the magnitude is quite small.[1]

A close examination of Appendix Figure 8 suggests that there may be some asymmetry in wage changes for movers, with a somewhat larger gap between predictions and actual wage changes for those who move to firms with higher premiums than for those who move to firms with lower premiums. We investigate this further in Appendix Figure 9. Here, we use the same 20-by-20 grid of vingtile-to-vingtile moves. For each pair (i,j) where j is a higher vingtile than i,

[1] We have conducted this analysis separately for workers in each tercile of the distribution of estimated α_i, to gauge violations of additive separability. We find quite similar slopes across terciles.

we compute the average observed earnings change from period -1 to +1 for i-to-j movers (i.e., for those who move to higher-premium firms) and for the corresponding group of j-to-i movers. Appendix Figure 9 arrays the averages for downward moves on the y-axis against the corresponding averages for upper moves on the x-axis.

The AKM specification (2) implies that these should be equal and opposite, at least approximately. (Because there is variation in δ_f within vingtiles, it is possible that the predicted changes for observed movers in the two directions are not exactly opposite.) This is not quite what we see. The average downward mover loses a bit more than the average upward mover gains. This is an indication that the AKM model does not fit the data exactly. However, the deviations remain small. Moreover, we have already seen that the first post-move quarter is somewhat anomalous. When we repeat the exercise using data from the 4^{th} post-quarter move, the deviation is much smaller (yielding a slope of -1.11, as compared with the -1.30 in Appendix Figure 9 for the first post-quarter move).

We also show a similar symmetry test for between-CZ moves, in Appendix Figure 10. We have much less variation to work with here, as the variation in firm-level premiums is so much larger than that in CZ-level premiums. However, there is no indication here of the lack of symmetry we saw at the firm level. One interpretation is that within-CZ, between-firm moves, particularly to higher-premium firms, are more likely to depend on changes in match effects or non-wage amenities that partially offset the change in firm premiums than are between-CZ moves.

Overall, our investigation indicates that equation (2), with its implicit assumptions of additive separability and exogenous mobility, fits the data reasonably well. We do find some evidence of small violations of the exogenous mobility assumption, particularly in the first few quarters after a move, that could be incorporated in future work. However, so far as we are able to assess, these violations seem to be most severe for within-CZ moves, and wage changes surrounding between-CZ moves seem to come much closer to the AKM model ideal.

3. Public estimates of CZ wage premiums

There would be value in distributing our estimates of CZ wage premiums for use in other analyses. Unfortunately, we have not been able to disclose CZ-level statistics from the LEHD, beyond the classification into quintiles shown in Figure 2. As an alternative, we have produced approximations to the LEHD-based premiums based on public information from the ACS and disclosed regression coefficients (presented in Tables 6 and 8). These are predicted values from regressions of the LEHD premiums on three ACS variables: The estimated CZ wage premium from our first cross-sectional model, the log of the CZ size, and the share of the workforce with some college or more. The fitted values from this regression have standard deviation 0.072, so are correlated 0.90 with the LEHD estimates (and thus account for 83% of the variance of those estimates). They are unbiased predictors, so substitution of these ACS-based measures for the undisclosed LEHD measures as independent variables in regressions should not bias estimates (Hyslop and Imbens 2001).

The ACS-based CZ wage premium estimates are reported for the 50 largest CZs in Appendix Table 3, and for all CZs in a data file that will be distributed with this appendix.

4. Extended analyses

We present here a few additional results.

First, Appendix Table 4 repeats the variance decomposition of our AKM estimates from Table 2, both for the full sample shown earlier (columns 1-2) and for two subsamples. (In each case, we use an AKM model fit to the entire sample, and restrict to subsamples only in summarizing that model's results.) In columns 3-4, we show an analysis of non-movers, people observed only in a single CZ in our sample. Because the AKM model is identified from movers, with results then extrapolated to stayers, evidence that stayers differ substantially from movers would reduce confidence in the analysis. We see that the decomposition for stayers is similar to that in the full sample. In columns 5-6, we limit to person-quarter observations for which the worker is at a single-establishment firm or for which the establishments are sufficiently geographically dispersed (or sufficiently similar) that all ten imputations of the establishment assignment yield the same industry and CZ assignment. Again, this makes little difference to the results.

Next, we presented analyses in the main text of differences across education groups. We present additional results for this analysis in Appendix Table 5. As in Appendix Table 4, we fit the main AKM model with the full sample, then analyze separately the low-education and high-education subsets of this sample. (This is necessary because we observe education only for a small share of observations; limiting the AKM estimation only to those with observed education at a particular level would eliminate many of the connections needed to identify the AKM model.) We show the variance decomposition separately for the two groups. Firm effects account for a larger share of the variance of earnings among low-education workers, while among high-education workers person effects are a bit more important in variance terms.

Implicit in this analysis' reliance on a pooled AKM model is an assumption that firm premiums δ_f are the same for high- and low-education workers. To assess this, we computed worker-firm match effects as the average residual for each worker-firm combination. Insofar as firm premiums differ for the two types of workers, that difference would be captured in the match effects (e.g., with positive match effects for H workers and negative match effects for L workers). We then regressed these match effects on firm-by-education indicators. The adjusted R-squared of this regression was negative, offering no indication of differences in firm premiums by education.

Appendix Table 6 addresses uses the five-part decomposition arising from equations 12 and 13 to explore the overall variance of the CZ-level return to education. As in Table 11, we see a large role for person effects, with differences in the skill gap between college and high school workers accounting for about one-half of the variance in returns. Most of the rest is explained by two important covariance terms: 14% is attributable to covariance between person effects and the pure CZ wage premium (the 1st term in equation (13)), and 28% is attributable to covariance between person effects and industry composition. The latter effect confirms the conclusion from Table 11 that differences in the degree of sorting of higher-education workers to high-wage industries is an important source of variation in the wage gap between college and high school workers across different CZ's.

Finally, we discuss our analysis of housing price differences across CZs, reported in Appendix Table 7. We construct four CZ-level housing measures - mean log home values for homeowners and mean log rents for renters, each unadjusted and then adjusted for housing characteristics[2]) on log CZ size. Characteristics are 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 rents (for rents only) and an indicator for whether the owner has a mortgage (homeowner model only).

We then regress our CZ housing price measure on log CZ size. Estimates are quite stable across the different measures, and point to an elasticity of housing costs with respect to log size that is around 0.2 or larger, substantially so when we limit to the 50 largest CZs.

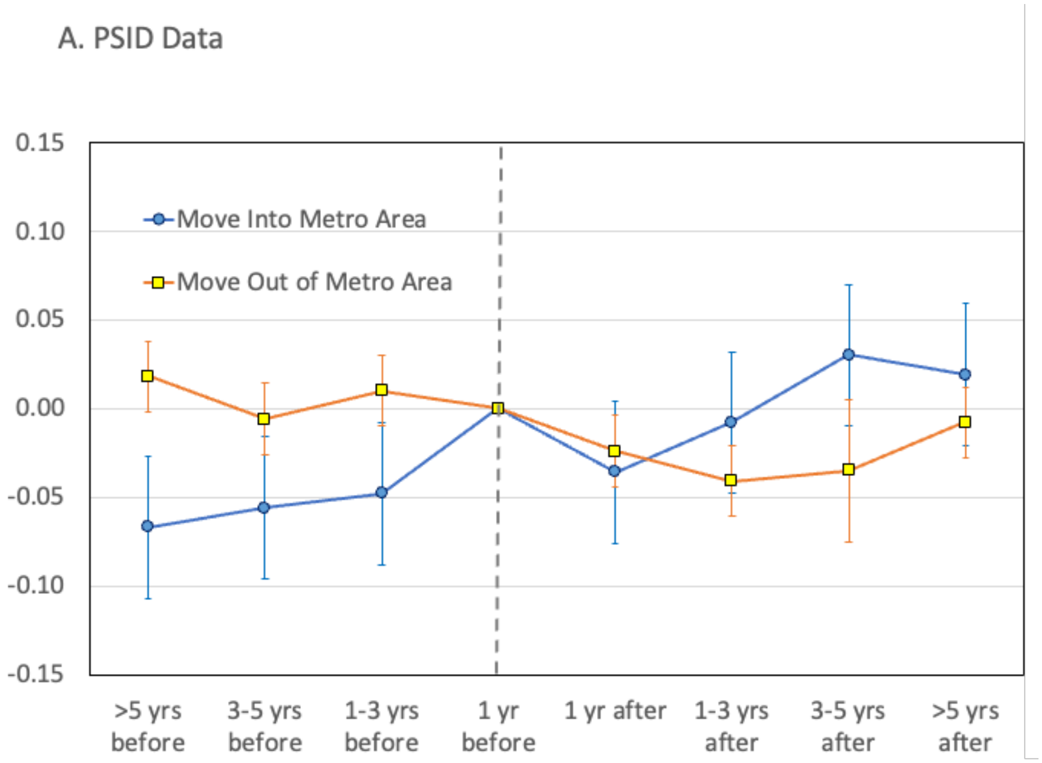
References

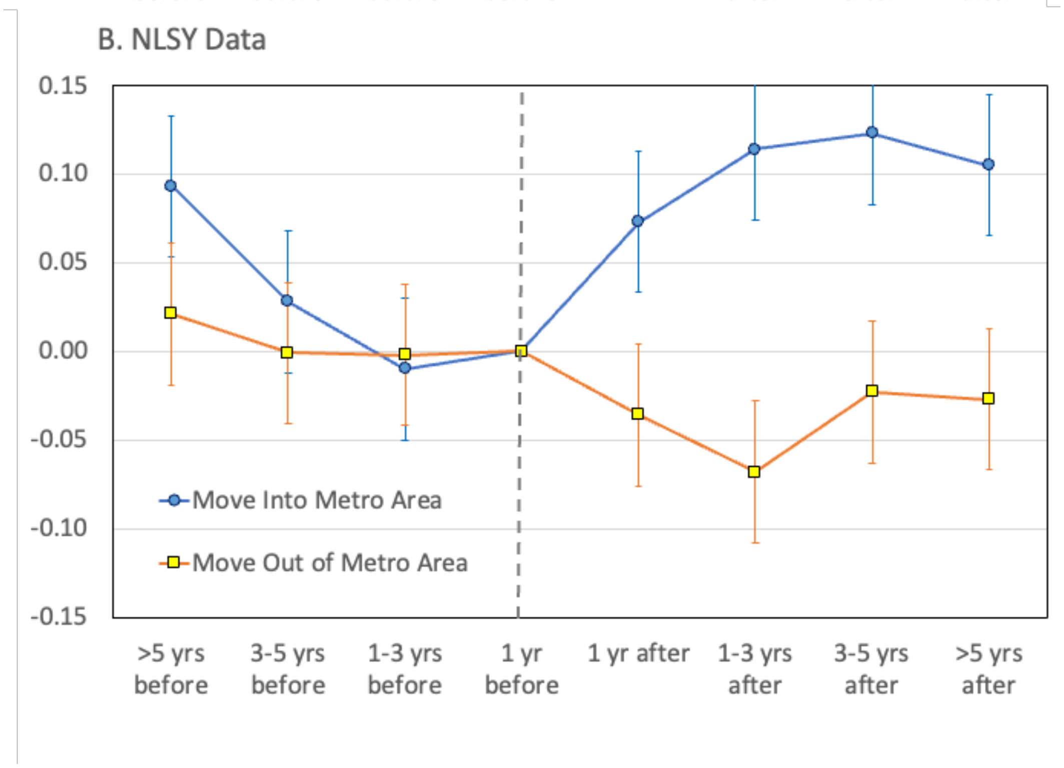
Tolbert, Charles M., and Molly Sizer. "U.S. commuting zones and labor market areas: A 1990 update." U.S. Department of Agriculture, Economic Research Service, staff paper number AGES-9614 (1996).

Vilhuber, Lars. "LEHD Infrastructure S2014 files in the FSRDC." U.S. Census Bureau, Center for Economic Studies Discussion Paper CES 18-27 (May 2018).

[2] 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 R-squared of 0.34, while that for home values has an R-squared of 0.50.

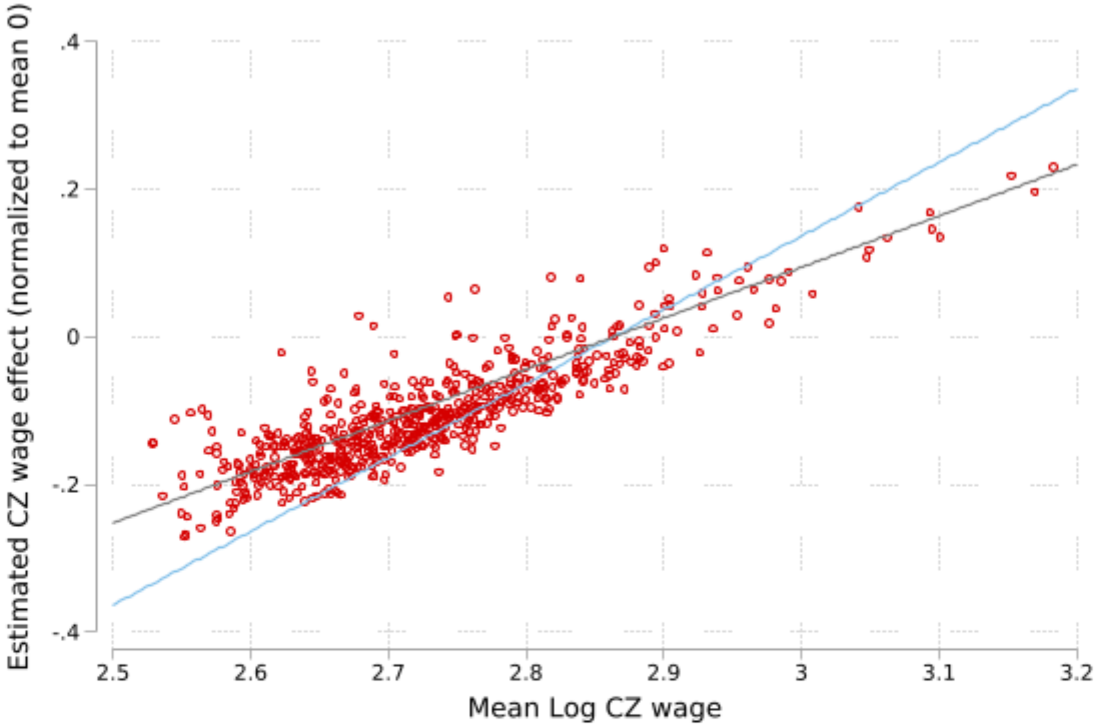
Appendix Figure 1. Wage changes for movers in and out of metropolitan areas (Glaeser and Mare, 2000)





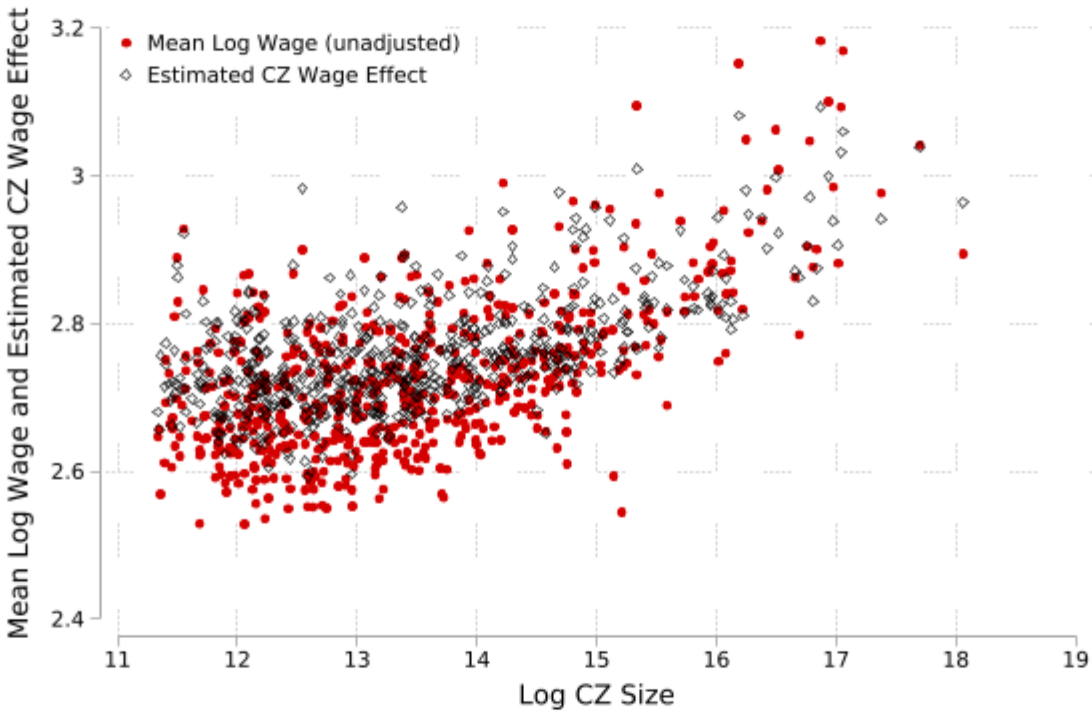
Notes: From Glaeser and Mare (2000, Table 5, columns 2 and 4). 95% confidence intervals shown with vertical bars.

Appendix Figure 2. Estimated CZ Wage Effects from Cross-Sectional Model versus Mean Log Wage in CZ, ACS data



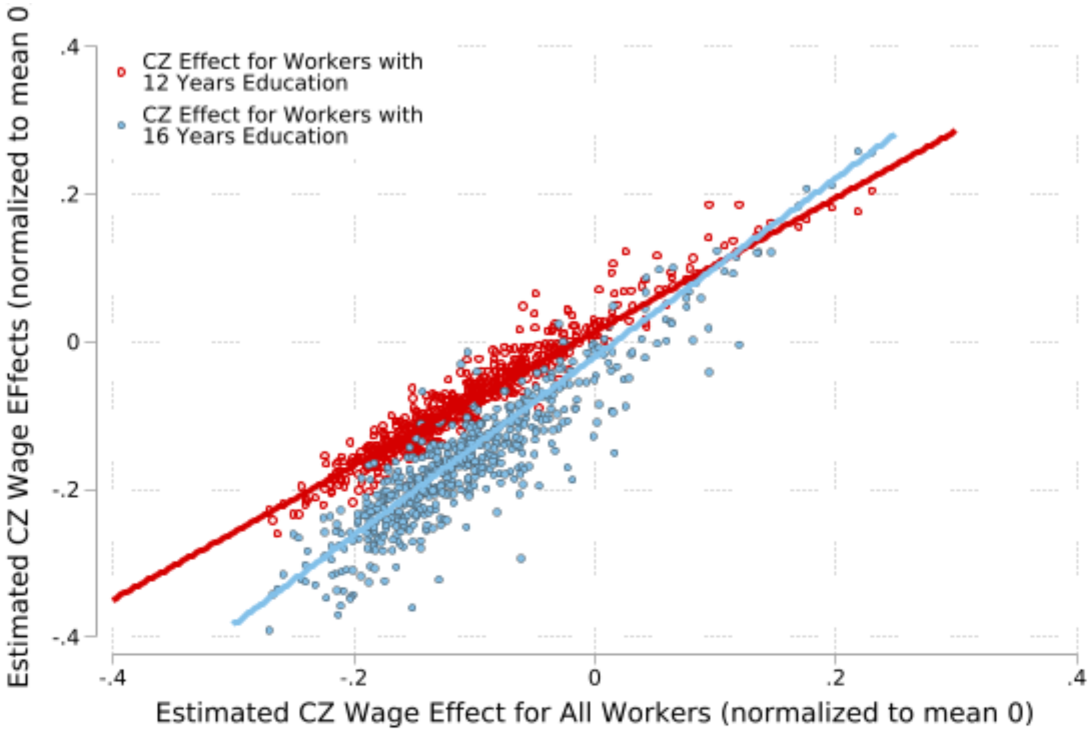
Notes: CZ effects on vertical axis are fixed effects from our third cross-sectional model fit to the ACS, controlling for years of completed education (fixed effects), interacted with gender, immigrant status, and whether the immigrant has been in the U.S. for less than their potential experience + 5; experience (quartic, interacted with gender); race and ethnicity dummies (interacted with gender), year fixed effects; three immigrant source regions, interacted with years in the U.S.; field of degree (16 categories, interacted with gender, only for those with a BA or more); and detailed industry. Fixed effects are normalized to weighted mean zero across CZs. Scatterplot shows the 600 largest CZs. Overlayed lines show the 45-degree line (with intercept adjusted to equal zero for the average CZ) and a weighted regression line fit to all CZs (slope=0.69; robust standard error=0.02).

Appendix Figure 3. Relationship of Mean Log Wage and Cross-Sectional CZ Wage Effect to Log of CZ Size, ACS data



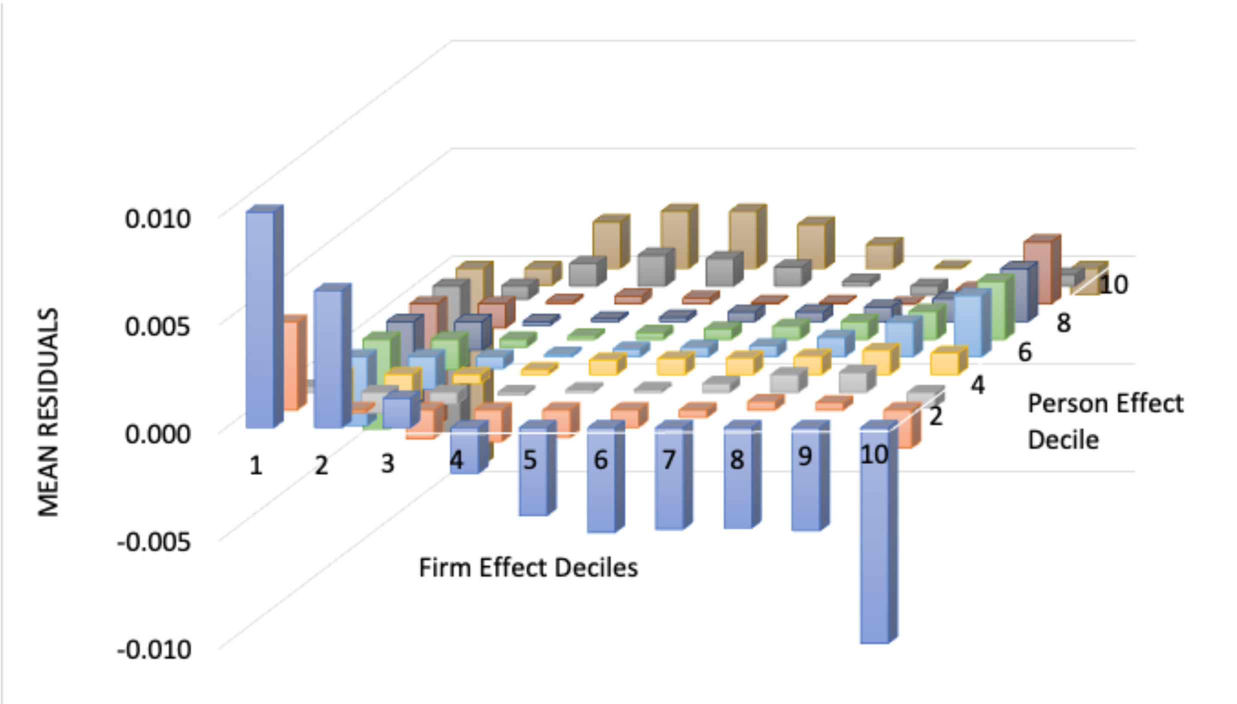
Notes: See notes to Appendix Figure 2 for description of estimated CZ wage effects. Points are shown for the 600 largest CZs. Wage effects are adjusted to have the same weighted mean as mean CZ wages.

Appendix Figure 4. Relation of Cross-Sectional CZ Wage Effects for High School and College Workers with Cross-Sectional CZ Wage Effect for All Workers, ACS data



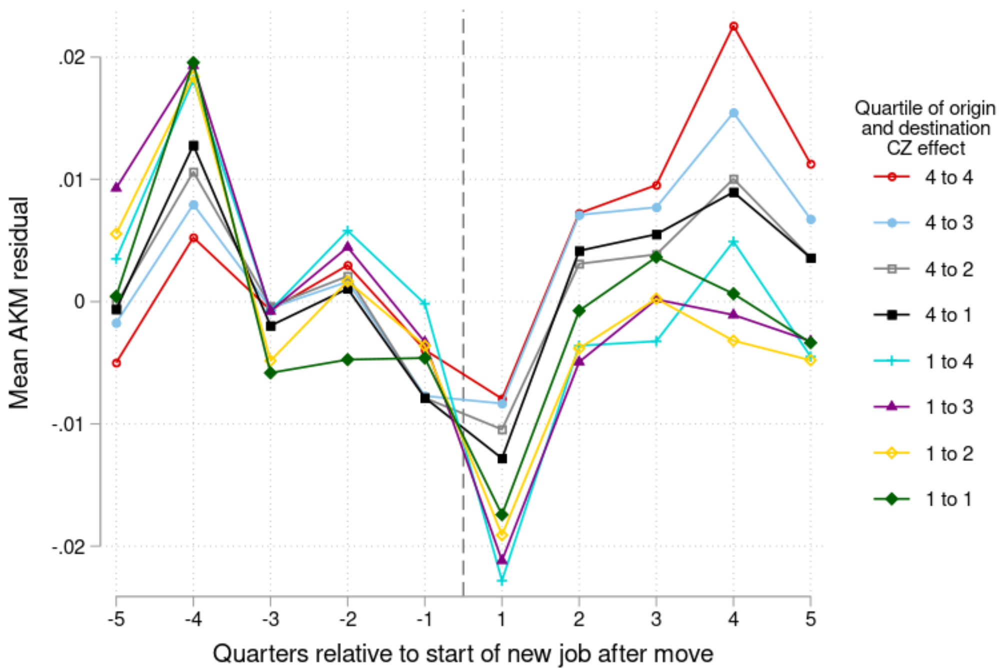
Notes: CZ effects on X-axis are from the model described in the notes to Appendix Figure 2. The model is then re-fit separately for workers with exactly 12 and exactly 16 years of education. CZ fixed effects from these models are plotted on the Y-axis for the 600 largest CZs. Overlayed lines are lines of best fit, fit to the full sample and weighted by the number of workers at each education level.

Appendix Figure 5. Mean AKM residuals by decile of person effect and CZ/industry effect



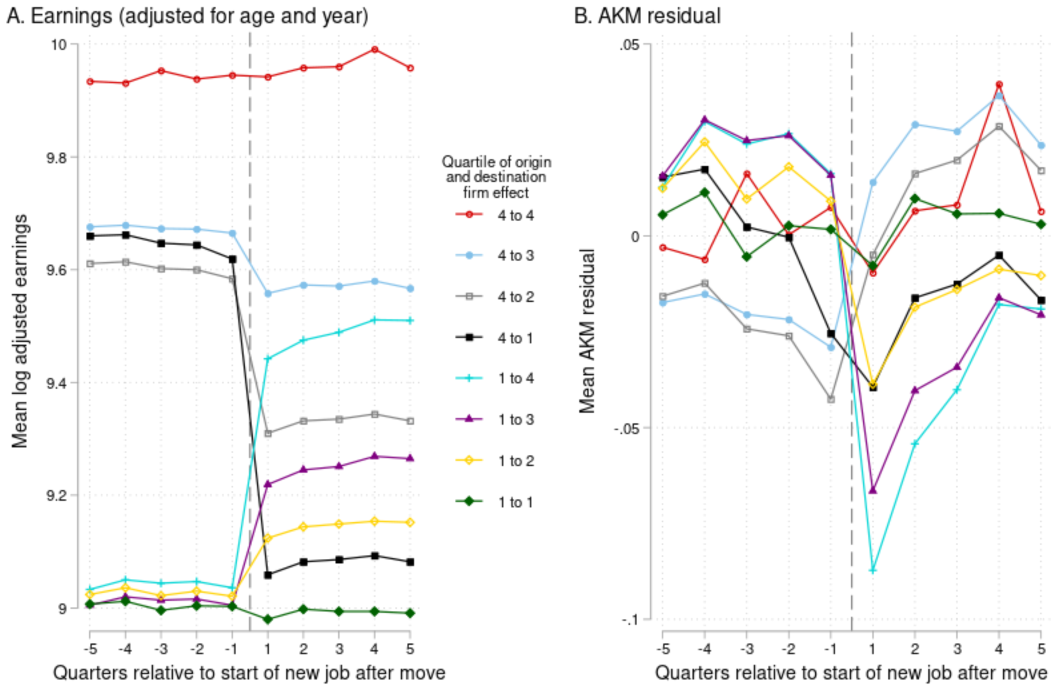
Note: Figure shows mean residuals from AKM specification by decile of the estimated person and firm effects.

Appendix Figure 6. Mean earnings residuals from AKM model before and after a change of CZs, by change in CZ earnings premium



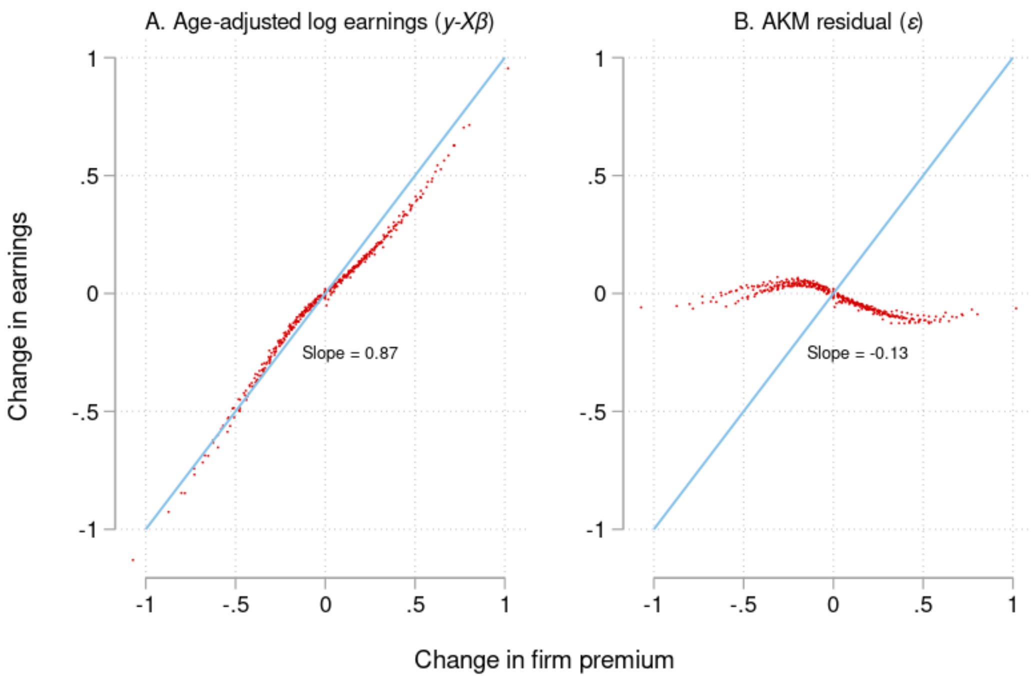
Notes: Figure shows event-time means of the AKM residual for workers who move between CZs, separately for different quartiles of the origin and destination AKM-based CZ earnings premiums. Quarter -1 represents the last full quarter (i.e., the second-to-last observed quarter) in the origin CZ, and quarter +1 represents the first full quarter (second observed quarter) in the destination CZ. We allow up to four quarters of non-employment between the two observed spells. Sample is limited to workers who move between CZs only once in our sample and are observed with stable jobs at the same firm for at least five consecutive quarters (not including transition quarters) before and after the switch.

Appendix Figure 7. Mean earnings and mean residuals from AKM model before and after move across firms, by change in firm earnings premium



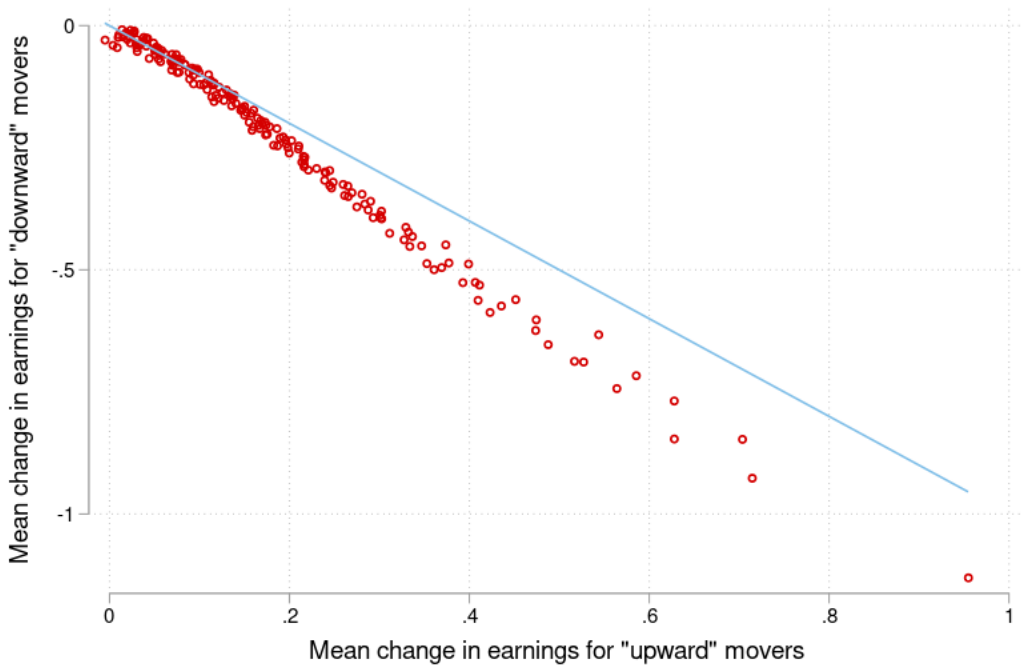
Notes: Figure shows event-time means for workers who move between establishments, separately for different quartiles of the origin and destination AKM-based establishment earnings premiums. Quarter -1 represents the last full quarter (i.e., the second-to-last observed quarter) in the origin establishment, and quarter +1 represents the first full quarter (second observed quarter) in the destination establishment. We allow up to four quarters of non-employment between the two observed spells. Sample is limited to workers who move between establishments only once in our sample and are observed with stable jobs at the same firm for at least five consecutive quarters (not including transition quarters) before and after the switch.

Appendix Figure 8. Change in earnings and AKM residuals of firm movers from last pre-move quarter to first post-move quarter, by change in firm premium



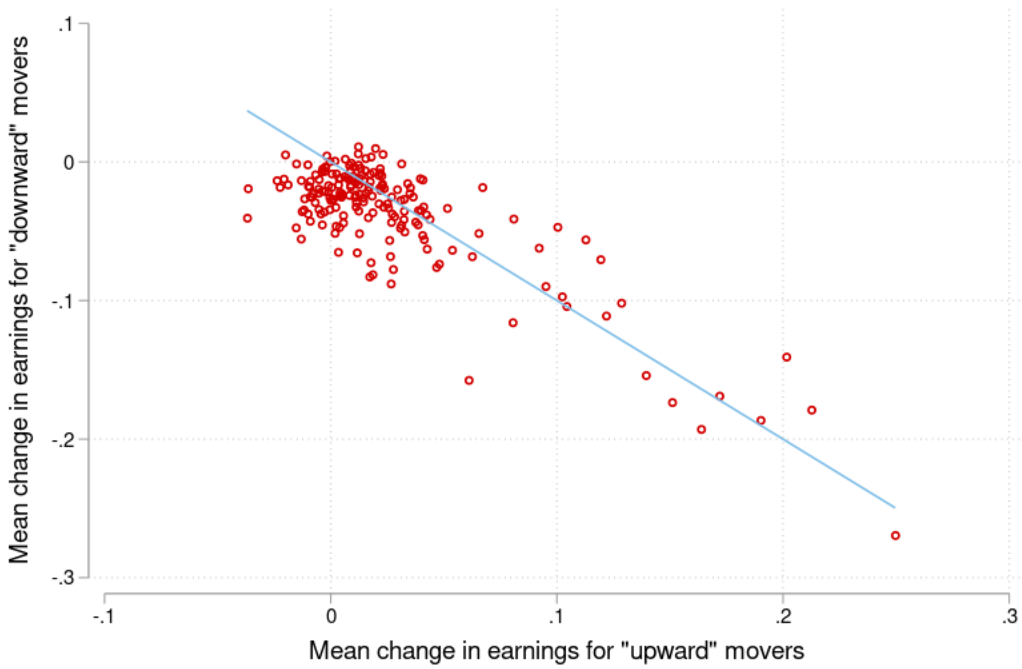
Notes: Firms are classified into 20 vingtiles based on their AKM-based establishment earnings premiums. We then classify firm movers based on the change in mean premium from their origin to their destination vingtile. Figure shows the average change in the indicated earnings components from the last full quarter in the origin firm to the first full quarter in the destination firm for each of the resulting 400 cells. Slopes correspond to the unweighted best linear fit line fit to these 400 points. Sample is limited to workers who move between establishments only once in our sample and are observed with stable jobs at the same firm for at least five consecutive quarters (not including transition quarters) before and after the switch.

Appendix Figure 9. Comparison of earnings changes for “upward” and “downward” firm movers



Note: Firms are classified into 20 vingtiles based on their AKM-based establishment earnings premiums. We identify all 190 origin-destination pairs of vingtiles where the destination has higher premiums than the origin and compute the average change in earnings from the last full pre-move quarter to the first full post-move quarter (excluding transition quarters) for this origin-destination combination. These values are plotted on the x-axis. On the y-axis, we plot the average change for movers going in the opposite direction, toward a lower-premium firm. The overlayed line represents what would be expected if average earnings changes were equal and opposite. Sample is limited to workers who move between establishments only once in our sample and are observed with stable jobs at the same firm for at least five consecutive quarters (not including transition quarters) before and after the switch.

Appendix Figure 10. Comparison of earnings changes for "upward" and "downward" CZ movers



Note: CZs are classified into 20 vingtiles based on their AKM-based CZ earnings premiums. We identify all 190 origin-destination pairs of vingtiles where the destination is has higher premiums than the origin and compute the average change in earnings from the last full pre-move quarter to the first full post-move quarter (excluding transition quarters) for this origin-destination combination. These values are plotted on the x-axis. Note that a few are negative because the mean change in earnings has the opposite sign from the difference in estimated CZ premiums from our AKM model. On the y-axis, we plot the average change for movers going in the opposite direction, toward a lower-premium CZ. The overlayed line represents what would be expected if average earnings changes were equal and opposite. Sample is limited to workers who move between CZs only once in our sample and are observed with stable jobs at the same firm for at least five consecutive quarters (not including transition quarters) before and after the switch.

Appendix Table 1: CZ size effects on log hourly wages and log annual earnings, ACS data

	Unadjusted	CZ Effect from Model 2	CZ Effect from Model 3
	(1)	(2)	(3)
All Workers			
Log hourly wage	0.068	0.059	0.056
	(0.010)	(0.005)	(0.004)
	[0.497]	[0.607]	[0.605]
Log annual earnings	0.078	0.059	0.060
	(0.013)	(0.007)	(0.006)
	[0.489]	[0.541]	[0.590]
High School-educated Workers (12 years education)			
Log hourly wage	0.031	0.050	0.049
	(0.004)	(0.004)	(0.004)
	[0.308]	[0.548]	[0.547]
Log annual earnings	0.028	0.046	0.050
	(0.005)	(0.005)	(0.005)
	[0.214]	[0.426]	[0.511]
Some College or More Workers(>12 years education)			
Log hourly wage	0.082	0.070	0.065
	(0.009)	(0.005)	(0.005)
	[0.591]	[0.658]	[0.655]
Log annual earnings	0.092	0.072	0.069
	(0.013)	(0.007)	(0.006)
	[0.551]	[0.607]	[0.638]

Note: Table entries are estimated regression coefficient, robust standard error (in parentheses) and R-squared [in square brackets] from OLS regression of estimated CZ wage or earnings differential on log of CZ size. Models in column 1 are fit to unadjusted mean log hourly wages or mean log annual earnings. Models in columns 2 and 3 are fit to adjusted CZ earnings differentials derived from models described in text. All models are fit by weighted least squares using as a weight the total number of worker observations for the CZ in the 2010-2018 ACS sample.

Appendix Table 2: Models for CZ employment rates and average hours, ACS data

	Both Genders			Males only			Females only		
	Regressed on log(size)	Regressed on mean log wage in CZ: Unadjusted	Regressed on mean log wage in CZ: Adjusted	Regressed on log(size)	Regressed on mean log wage in CZ: Unadjusted	Regressed on mean log wage in CZ: Adjusted	Regressed on log(size)	Regressed on mean log wage in CZ: Unadjusted	Regressed on mean log wage in CZ: Adjusted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable = CZ-mean of:									
Employed last year	0.005	0.15	0.12	0.010	0.17	0.17	0.001	0.13	0.08
	(0.003)	(0.02)	(0.03)	(0.003)	(0.02)	(0.03)	(0.004)	(0.02)	(0.04)
Employed last year (adjusted)	0.000	0.07	0.04	0.001	0.06	0.03	-0.001	0.08	0.04
	(0.002)	(0.02)	(0.03)	(0.002)	(0.02)	(0.03)	(0.002)	(0.02)	(0.03)
Hours last year	11.42	337	283	15.62	359	346	9.04	325	230
	(8.27)	(49)	(83)	(8.21)	(53)	(86)	(9.00)	(50)	(87)
Hours last year (adjusted)	-3.82	80	34	-7.42	34	-8	0.02	128	79
	(5.37)	(40)	(63)	(6.12)	(48)	(75)	(4.90)	(34)	(54)

Notes: Table entries are regression coefficients from univariate regressions of dependent variable shown in row heading on independent variable indicated in column heading. Hours last year includes zeros for non-workers. All models are estimated using data for 691 CZ's and characteristics estimated using 2010-2018 ACS files. "Adjusted" means that the dependent variable or the independent variable (or both) is estimated from a model with CZ dummies and a set of individual characteristics. All regression models are weighted by the (weighted) count of working-age people in the CZ in the ACS files. Robust standard errors in parentheses.

Appendix Table 3. Predictions of CZ wage premiums based on public data (top 50 CZs)

Rank	CZ number	Largest city in CZ	Approx. premium	Rank	CZ number	Largest city in CZ	Approx. premium
1	38300	Los Angeles, CA	0.050	26	16300	Pittsburgh, PA	-0.053
2	19400	New York, NY	0.098	27	33000	Fort Worth, TX	0.010
3	24300	Chicago, IL	0.046	28	7400	Orlando, FL	-0.064
4	11304	Washington, DC	0.155	29	38801	Portland, OR	0.014
5	19600	Newark, NJ	0.113	30	18000	Buffalo, NY	-0.040
6	32000	Houston, TX	0.046	31	31301	San Antonio, TX	-0.025
7	19700	Philadelphia, PA	0.046	32	37901	Las Vegas, NV	0.017
8	20500	Boston, MA	0.094	33	31201	Austin, TX	0.022
9	37800	San Francisco, CA	0.169	34	12701	Cincinnati, OH	-0.020
10	9100	Atlanta, GA	0.006	35	29502	Kansas City, MO	-0.021
11	11600	Detroit, MI	-0.032	36	1701	Raleigh, NC	-0.011
12	39400	Seattle, WA	0.074	37	15900	Columbus, OH	-0.023
13	33100	Dallas, TX	0.039	38	900	Charlotte, NC	-0.011
14	7000	Miami, FL	-0.035	39	14200	Indianapolis (remaind, IN	-0.028
15	35001	Phoenix, AZ	0.003	40	36100	Salt Lake City, UT	-0.023
16	21501	Minneapolis, MN	0.036	41	24100	Milwaukee, WI	-0.011
17	20901	Bridgeport, CT	0.090	42	5600	Nashville-Davidson (remai	-0.037
18	28900	Denver, CO	0.036	43	7100	West Palm Beach, FL	-0.021
19	38000	San Diego, CA	0.055	44	20401	Providence, RI	0.038
20	37400	Sacramento, CA	0.051	45	7600	Jacksonville (remaind, FL	-0.031
21	11302	Baltimore, MD	0.095	46	37200	Fresno, CA	0.008
22	6700	Tampa, FL	-0.047	47	12200	Grand Rapids, MI	-0.065
23	37500	San Jose, CA	0.183	48	20600	Manchester, NH	0.016
24	15200	Cleveland, OH	-0.045	49	33803	Oklahoma City, OK	-0.057
25	24701	St. Louis, MO	-0.024	50	2000	Virginia Beach, VA	-0.020

Notes: CZ premiums are best linear predictors of non-disclosed LEHD-based premiums from public ACS data, using as predictors the estimated premium from our first cross-sectional model, the log size of the CZ, and the share of workers in the CZ with some college or more. Premiums are normalized to weighted mean zero across CZs.

Appendix Table 4: Comparison of CZ-Level Variance Decompositions for Main Sample, No-Uncertainty Sample, and Non-Mover Sample

	Main sample		Non-mover sample		No-uncertainty sample	
	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	0.147	100%	0.140	100%	0.145	100%
Variance components (std. deviations in odd-number columns, variance shares in even-numbered columns)						
Person effects	0.081	30.3%	0.076	29.2%	0.081	31.0%
Firm effects	0.079	29.3%	0.077	30.4%	0.077	28.3%
Covariate index (Xβ)	0.006	0.2%	0.006	0.2%	0.006	0.2%
Covariance components (correlations in odd-number columns, variance shares in even-numbered columns)						
Person & firm	0.642	38.2%	0.617	36.8%	0.643	38.1%
Person & covariate index	0.228	1.0%	0.337	1.5%	0.259	1.2%
Firm & covariate index	0.245	1.0%	0.406	1.9%	0.239	1.1%
Sample size (billions of person-quarter observations)	2.52		1.84		1.90	

Notes: Table shows variance decompositions of CZ-level mean earnings based on equation (6). Entries in odd-numbered columns "variance components" are standard deviations of the earnings components indicated in row headings; for "covariance components" they are the estimated correlations of the indicated components. Entries in even-numbered columns are variance shares explained by the variance or covariance components.

Appendix Table 5: Summary of AKM model estimates by education group

	Low education sample				High education sample			
	Person-quarter level		CZ level		Person-quarter level		CZ level	
	Std. Dev. or Correlation	Var. Share	Std. Dev. or Correlation	Var. Share	Std. Dev. or Correlation	Var. Share	Std. Dev. or Correlation	Var. Share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log earnings or mean log earnings	0.514	100%	0.097	100%	0.678	100%	0.178	100%
Variance components (std. deviations in odd-number columns, variance shares in even-numbered columns)								
Person effects	0.409	63.5%	0.055	32.3%	0.559	67.8%	0.101	32.2%
Firm effects	0.212	17.0%	0.059	36.8%	0.223	10.8%	0.087	24.1%
Covariate index (Xβ)	0.127	6.1%	0.007	0.6%	0.130	3.7%	0.005	0.1%
Residual	0.192	14.0%	0.001	0.0%	0.244	13.0%	0.000	0.0%
Covariance components (correlations in odd-number columns, variance shares in even-numbered columns)								
Person & firm	0.112	7.3%	0.433	29.9%	0.167	9.0%	0.753	41.9%
Person & covariate index	-0.177	-7.0%	-0.247	-2.1%	-0.139	-4.4%	0.318	1.1%
Firm & covariate index	-0.017	-0.4%	0.226	2.1%	0.015	0.2%	0.255	0.7%

Notes: Table shows variance decompositions based on equation (4). Columns 1-2 and 5-6 pertain to the variance of individual quarterly earnings. Columns 3-4 and 7-8 pertain to the variance of mean earnings by CZ. Entries in odd-numbered columns for "variance components" are standard deviations of the earnings components indicated in row headings; for "covariance components" they are the estimated correlations of the indicated components. Entries in even-numbered columns are variance shares explained by the variance or covariance components.

Appendix Table 6: Variance Decomposition of CZ Wage Gap between High- and Low-education Workers

	Variance Decomposition of Wage Gap	
	Std. Dev. or Correlation (1)	Var. Share (2)
Wage gap (high- versus low-education workers)	0.107	100%
Components of Wage Gap (column 1 = std. dev.)		
Difference in mean person effects	0.074	49%
Difference in covariate indexes	0.006	0%
Difference in mean CZ wage effect:	0.038	13%
Of which:		
Relative wage premium	0.014	2%
Composition	0.028	7%
Interaction	0.002	0%
Covariance Terms (column 1 = correlation of terms)		
Cov(Person effects, cov. index)	-0.251	-2%
Cov(Person effects, CZ effect)	0.831	42%
Of which:		
Cov(Person effects, relative wage premium)	0.785	14%
Cov(Person effects, composition)	0.785	28%
Cov(Person effects, interaction)	-0.311	-1%
Sum of all other covariance terms	--	3%

Notes: Table shows variance decomposition of the difference in mean log earnings of college and non-college workers in a CZ, using equations (12) and (13). Analysis is weighted by CZ size.

Appendix Table 7: Elasticities of Housing Values and Rents with respect to CZ Size

	All CZ's (1)	Largest 50 CZ's (2)
Housing Prices (log of home value for owners)		
Unadjusted	0.25	0.37
	(0.02)	(0.05)
Quality Adjusted	0.23	0.41
	(0.03)	(0.06)
Monthly Rent (log of rent for renters)		
Unadjusted	0.17	0.19
	(0.01)	(0.02)
Quality Adjusted	0.18	0.22
	(0.01)	(0.02)

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. Regressions are weighted by number of workers in CZ. Sample in column 1 is set of 678 CZ's in 2018 5-year ACS with non-missing data. Sample in column 2 is 50 largest CZ's, ranked by number of workers. Quality adjusted values and rents derive from regressions on indicator for type of housing unit, number of bedrooms, log of total number of rooms, year of construction, and indicator for mortgage (for home values) or set of indicators for inclusion of utility costs (for rents).