Re-assessing the Spatial Mismatch Hypothesis

By DAVID CARD, JESSE ROTHSTEIN, AND MOISES YI

ONLINE APPENDIX

A. The spatial mismatch literature

In his paper introducing the spatial mismatch (SM) hypothesis, Kain (1968) presented a series of models that related the share of jobs held by black workers in a given "work zone" in Chicago or Detroit to the Black share of residents in that zone and the distance from the zone to the nearest "ghetto."[1] He then performed a counterfactual analysis, assuming that Black residential shares were equalized across zones (eliminating the ghetto areas) but that the model coefficients remained unchanged. The results suggested that the elimination of residential segregation would lead to a 10% rise in Black employment in Chicago and a 3-7% rise in Detroit.

Two criticisms were raised almost immediately about these findings. First, the conclusions of Kain's counterfactual analysis were reversed if the model was relaxed slightly to allow a nonlinear relation between Black residential shares and Black employment shares (Offner and Saks, 1971). Second, it is not obvious how to derive Kain's specification from an underlying model of the choices of Black and white workers about where to work. Many subsequent studies tried to incorporate broader city-wide measures of the decentralization of jobs (e.g., Mooney, 1969; Masters 1974) or of the average distance between residential locations and potential work locations (e.g., Hutchinson, 1974). While some studies in this vein found support for SM, others did not. In a careful review of the literature up to 1990, Holzer (1991, p. 118) concluded that "… the evidence remains very contradictory." Glaeser, Hanushek and Quigley (2004, p. 76) are less sympathetic, asserting: "It is not generally true that Blacks live further from jobs than Whites do, and it is hard to believe that the physical costs of getting to jobs are really responsible for the pathologies of the ghetto."

[1] The precise definition of a zone is unclear, but maps in Kain (1964), which presents the same empirical evidence as Kain (1968), suggest the zones are related to Census tracts. Similarly, Kain does not discuss how he defines the boundaries of the ghettos in Chicago and Detroit in the mid-1950s.

A useful theoretical perspective on the SM hypothesis is provided by Ellwood (1986), who discusses the conditions under which the employment outcomes of two groups of equally productive workers are affected by their residential locations. He argues that the clearest way to test the SM hypothesis is relate outcomes of workers (or potential workers) in a neighborhood to measures of job proximity for people in that neighborhood. This is the idea that guides our analysis.

There have been important changes in residential patterns since the 1960s that may have reduced the importance of SM. Black households are much less concentrated in formerly redlined neighborhoods in central cities than they were in the 1960s, and the Black population has suburbanized to a significant extent. Moreover, many more workers commute by car rather than public transportation. However, residential segregation remains high, preserving the plausibility of SM as a potential explanation for racial differences in labor market outcomes.

B. A Simple Model

i) Basic Setup

In this section we sketch a simple model of wage outcomes for workers in a spatially differentiated labor market. The model explicitly builds in an AKM-style model of wage setting in which each establishment offers a proportional wage premium that raises or lowers wages of any worker who is employed there relative to other workplaces in the market. Similar to the monopsonistic competition model in Card et al. (2018) the model focuses on worker's preferences over available jobs, ignoring frictions in the matching process.

Specifically, assume that person i gets utility from employer j:

$$u_{ij} = a_i + \delta_j - \beta_i d_{ij} + \epsilon_{ij},$$

where δ_j is the pay premium offered by employer j, d_{ij} is a measure of the commute distance for i to get to workplace j, and ϵ_{ij} is a match effect. If worker *i* takes a job at employer *j* her observed wage is:

$$\ln w_{ij} = \alpha_i + \delta_j + v_{ij},$$

where (as in a standard AKM model) α_i is a common component of wages for i across all jobs, and the residual term v_{ij} is assumed to be independent of d_{ij} and ϵ_{ij}.

Next, assume that a worker who is searching for a job has an "offer set" O_i representing the potential set of job opportunities that are available. She takes the job with the highest utility in the set:

$$j^*(i) = argmax_{j \in O_i}[\delta_j - \beta_i d_{ij} + \epsilon_{ij}],$$

and we observe the combination of the wage premium and commute distance $(\delta_{j^*(i)}, d_{ij^*(i)})$ for that worker.

ii) Comparing Job Opportunities of Different Groups

Suppose there are two groups of workers G_1 and G_2. If the joint distributions of (δ_j, d_{ij}) in the offer sets are the same for the two groups, and they have the same distribution of $\beta_i's$, then they will have the same probability distributions over $(\delta_{j^*(i)}, d_{ij^*(i)})$. In particular, the conditional expectation of the wage premium, given commute distance

$$E[\delta_{j^*(i)} | d_{ij^*(i)} = d]$$

will be the same for the two groups. This provides the basis for a simple "outcome test": if two groups have the same access to jobs, and the same preferences for wages versus commuting distance, then we would expect the observed relationship between wage premiums and commute distances to be the same for the two groups.

To facilitate comparisons between workers with different offer sets, suppose that commute distances are discrete, $d_{ij} \in \{d_1, d_2, \ldots d_N\}$, and that wage premiums are also discrete, $\delta_j \in \{\delta_1, \delta_2, .. \delta_M\}$. In this case the offer set for a given worker is summarized by which particular combinations of (δ_u, d_v) are available (i.e., the support of the joint distribution of wage premiums and commute distances). For example a high-wage premium job at close proximity may not be available in a given worker's choice set.

Suppose that the offer sets for individuals in group G_1 have the property that jobs with wage premiums $\delta_j \in \{\delta_1, \delta_2, .. \delta_M\}$ are available at every commute distance, while the offer sets for individuals in group G_2 have the property that jobs with wage premiums $\delta_j > \bar{\delta}$ are only available with commute distances $d_{ij} > \bar{d}$. In this case we would say that the job opportunities of group G_2 are negatively affected by their residential locations, relative to group G_1. In particular we would expect that the observed wage premiums for workers in G_2 with relatively short commuting distances would be lower than the premiums for workers in G_1 in the same range of commute distances. We would also expect that the slope of the conditional expectation of the wage premium, given the commute distance, will be higher for the disadvantaged group.

C. LEHD sample and comparisons to American Community Survey

(i) LEHD Sample

We use data from Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program. These data are derived from quarterly earnings reports provided by employers to state unemployment insurance (UI) agencies. The core data set includes total wages paid by a given employer to each worker in a quarter. This is supplemented with information on employers and workers derived from other sources (e.g., the decennial census and ACS files) – see Abowd et al. (2009). The LEHD covers about 95% of private sector employment, as well as state and local government employees, but excludes federal employees, members of the armed services, and self-employed workers. From 2010 forward it includes data from all 50 states.

Our sample construction follows Card et al. (2023). We begin with person-employer-quarter (PEQ) observations from 2010Q1 to 2018Q2 where the worker is between 22 and 62 years of age. To help screen out part-time jobs and/or partial-quarter job spells we exclude PEQs with earnings below $3,800 (roughly the earnings from a full-time job at the federal minimum wage), quarters where an individual had multiple jobs, and all *transitional* quarters (the first or last quarter of any person-employer spell). We also drop PEQs with an unknown industry and/or establishment location. Finally, we drop individuals with fewer than 8 quarters of earnings that satisfy the previous restrictions over our 8½ year sample window, and individuals who are neither white non-Hispanic or Black non-Hispanic. We assign individuals to 1990 Commuting Zones (CZs) (Tolber and Sizer, 1996) based on the county of their establishment.

The upper rows of **Table A-1** reports summary statistics for non-Hispanic white and Black workers in three groups of CZs – the "older industrial" and "newer Sunbelt" CZs discussed in the main paper, and a residual group consisting of all other CZs in the top-200. (We discuss the lower rows of the table below.) Not surprisingly, the first two groups – which are drawn exclusively from the largest 30 CZs – have somewhat higher earnings for both white and Black workers than does the latter group. They also have larger white-Black gaps in log earnings – 0.37 and 0.38, respectively, vs. 0.25 in the other CZs.

(ii) Comparisons to American Community Survey (ACS)

In this appendix, we also report some results for a sample constructed from the 2010-2018 ACS. We select people age 22-62 inclusive from the ACS with at least 1 year of Mincer experience (i.e., age-education-6>0). For our analyses of earnings outcomes we further limit attention to "full year earners" with annual wage and salary earnings of $15,200 or higher (a threshold 4x higher than the quarterly threshold for full time work we impose on the LEHD). We assign 1990 CZs to Public Use Micro Areas (PUMA's) identified in the ACS using PUMA-county population files for the 2000 and 2010 Census created by David Dorn.[2] For PUMA's that contain observations from multiple CZs we probabilistically assign one CZ based on the relative share of the PUMA population in that CZ. Finally, we limit attention to individuals in the 30 largest CZs (based on counts of person quarter observations in our LEHD samples) and group the CZs into 4 groups: (1) Older industrial cities; (2) Newer Sunbelt cities; (3) Northeast Corridor; (4) Other CZs (which are mainly in the West). The resulting sample of workers in large CZs contains 6.49 million observations, representing a weighted population of roughly 721 million 22-62 year olds (80.2 million per sample year). In this sample 52% are white non-Hispanic and 13.7% are Black non-Hispanic; roughly 61.8% are classified as full year earners.

Table A-2 reports summary statistics for the working age populations in all larger CZs and in the four groups of cities, as well as statistics for the subset of full-year earners. (Note that in contrast to the statistics in Table A-1, these results include people of all ethnicities and racial groups). We note first that the population share of white and Black non-Hispanics varies across the four CZ groups, being relatively low in the older industrial cities and higher in the

[2] See https://www.ddorn.net/data.htm. We downloaded two files from this site: [E5] 2000 Census and 2005-2011 ACS Public Use Micro Areas to 1990 Commuting Zones; and [E6] 2010 Census and 2012-ongoing ACS Public Use Micro Areas to 1990 Commuting Zones.

Sunbelt cities. Mean years of education and the share of people with a BA or higher also vary somewhat across CZ groups and are particularly low in the Sunbelt (largely driven by the high share of Hispanics in these cities). Average employment rates (based on having earnings in the previous year) are fairly similar across city groups, ranging from 79 to 82%; the fraction of full-time earners varies a little more and is particularly low in the Sunbelt cities.

The middle rows of the table show characteristics of full year earners (i.e., with at least $15,200 of earnings last year). This group is 12-15% Black non-Hispanic except in the Western CZs, and is about 45% female. Mean annual earnings range from 60,000 in the Sunbelt cities to 75,000 in the Northeast corridor; mean hourly wages show a similar range. On average about 86% of full-time earners in the top 30 CZs commute to work in their own car: this rate is higher in the Sunbelt (around 95%) and lower in the Northeast corridor (67%). Mean commute times average about 30 minutes (one way), but are a little higher in the Northeast corridor (partly reflecting the fact that commuters by bus and rail have relatively long average commute times).

Finally, the bottom three rows of the table show mean log hourly earnings of white non-Hispanics, Black non-Hispanics, and the Black-white earnings gap. The mean gap is about 32 log points and is slightly lower in the older industrial cities than the Sunbelt cities or the Northeast corridor. Importantly, the magnitudes of the Black-white gaps in annual earnings in our ACS sample are similar to the gaps in quarterly earnings in our LEHD sample. About one-eighth of the gap in earnings for full-year earners appears to be due to a lower hours among Black workers – the Black-white gap in hourly wages for full-year earners in the largest 30 CZs is 28 log points. This gap, in turn, is not too different than the 26 log point gap in log hourly wages for 2010-2018 reported by Wilson and Darity (2022), based on hourly or weekly wages reported in the monthly Current Population Surveys.

Table A-3 reports some statistics on each of the 30 CZs in our ACS sample. Most of the older industrial cities have around a 20% share of Black workers (the exceptions are Minneapolis and Pittsburgh) while nearly all the newer Sunbelt cities have a relatively high share of Hispanics (the exception is Atlanta). Average one-way commute time are pretty similar across CZs, but higher in New York, and to a lesser extent Washington DC.

(iii) Imputation of Establishment Locations in LEHD

The UI data in the LEHD contain an identifier for the employing firm and the state, but not for the specific establishment if the firm has multiple establishments in the same state. The Census Bureau uses data on workers' residential addresses and the locations of establishments owned by the firm to impute establishments for workers employed at multi-establishment firms (Vilhuber 2018). We use the first of the multiple imputations available to assign PEQs to establishments. For some analyses in this appendix, we classify establishments by the characteristics of the firm to which they belong. We measure firm size as the largest number of PEQs associated with all of that firm's establishments in any quarter in our period. We use this to define three strata of firm size: Ten or fewer workers; 11-276 workers; and greater than 276 workers. We further divide firms in the latter two size categories into those with just a single establishment and those with multiple establishments, yielding a total of five firm categories. Note that we use State Employer Identification Numbers (SEINs) to define firms, so strictly speaking a "firm" for our purposes is a firm-by-state combination.

(iv) Coding of geographic locations in LEHD

The Census Bureau assigns geographic locations at a highly granular level to workers' residences and establishment locations, at an annual frequency. We use the latitude and longitude of home and workplace compute the as-the-crow-flies commute distance for each worker, in miles. To analyze the number of jobs within a radius r of each worker, we coarsen the locations of firms and workers to a set of grid points spaced 0.5 miles apart in both the North-South and East-West locations. Commute distances computed based on this grid are extremely highly correlated with those that use the original uncoarsened locations, so we do not believe much precision is lost with this coarsening, but it dramatically reduces computational burden. Because CZs differ dramatically in their scales, to make multiple-CZ averages meaningful we standardize distances across CZs by rescaling so that the 75^{th} percentile commute distance in each CZ equals 16 miles. So if the 75^{th} percentile in a particular CZ is 12 miles, we multiply all distances by 4/3, whereas if the 75^{th} percentile is 24 miles, we multiply all distances by ¾.

D. AKM Model and the role of establishment pay premiums in the racial wage gap

(i) AKM Model

Using our LEHD sample for each CZ, we fit an AKM model with worker and establishment fixed effects and time-varying observables:

(A-1) $$y_{it} = \alpha_i + \delta_{f(i,t)} + X_{it}\beta + \epsilon_{it}.$$

The variables included in X are a full set of calendar quarter indicators and a cubic in worker age. We estimate (A-1) separately for each CZ pooling Black and white workers but limiting to the largest connected set in the CZ (which typically includes well over 95% of PEQs in the CZ). We normalize the average pay premium ($\delta_{f(i,t)}$) of all firms in the restaurant industry in each CZ to zero. Thus, $\hat{\delta}_f$ can be interpreted as the average pay premium at establishment *f* relative to the average pay premium at restaurants in the same CZ.

Post-estimation we average the left-hand and right-hand sides of (1) by CZ and race, then take the difference between whites and Blacks in each CZ, yielding:

(A-2) $$\bar{y}_{cw} - \bar{y}_{cb} = (\bar{\alpha}_{cw} - \bar{\alpha}_{cb}) + \left(\bar{\delta}_{cw} - \bar{\delta}_{cb}\right) + (\bar{X}_{cw} - \bar{X}_{cb})\hat{\beta},$$

where $\bar{y}_{cw}$ and $\bar{y}_{cb}$ represent the means of log earnings for white and Black workers in CZ *c,* respectively, $\bar{\alpha}_{cw}$ and $\bar{\alpha}_{cb}$ represent the means of the estimated person effects for white and Black workers in that CZ, $\bar{\delta}_{cw}$ and $\bar{\delta}_{cb}$ represent the means of the estimated establishment effects for the two groups, and $\bar{X}_{cw}$ and $\bar{X}_{cb}$ represent the mean vectors of covariates. This is equation (2) in the text.

Let s_{fcw} and s_{fcb} represent the shares of all PEQ's of white and Black workers in CZ *c* that worked at establishment *f*. Then

(A-3) $$\bar{\delta}_{cw} - \bar{\delta}_{cb} = \sum_{f \in C} \left(s_{fcw} - s_{fbw} \right) \hat{\delta}_f.$$

Thus, the second term in equation (2) can be interpreted as measure of the differential sorting of whites relative to Blacks to workplaces with a higher estimated pay premium. If Black workers are less likely than whites to be employed at such workplaces, this term will be negative.

The lower rows of Table A-1 report the average values of the three terms on the right hand side of equation (A-2) for the three groups of larger CZs. Table 1 of the paper reports these components expressed as a percentage of the average Black-white gap in mean log earnings. Our estimated models imply that differences in the average pay premiums received by Black workers relative to white workers account for close to zero of the racial wage gap, and that the differential sorting of Black and white workers to high-paying establishments is very small.

In Card et al. (2024) we present a simple decomposition of establishment pay premiums into the mean by industry, and the deviation of the establishment premium from the average for its industry, which we call a within-industry "hierarchy effect". The bottom rows of Table A-1 use this approach to decompose the Black-white difference in mean establishment pay premiums into the difference in mean industry wage effects and the difference in mean hierarchy effects. Interestingly, for the two major CZ groups that are the focus of this paper these have opposite signs: Black workers work in slightly lower-paying industries than whites (particularly in the older industrial cities), but within a given industry they are employed at slightly higher-paying establishments.[3]

(ii) Interpretation

The fact that estimated average pay premiums for white and Black workers are nearly the same in our LEHD sample is surprising. Gerard et al. (2021) find that the under-representation of Black workers relative to whites at higher-paying workplaces in Brazil explains about 20% of the Black-white pay gap in that country. Moreover, our estimated AKM models, like the models in most other recent studies, imply that there is tendency for workers with higher values of α_i to work at establishments with higher pay premiums. In particular, our estimates imply that, other things equal, a 10% increase in α_i is associated with about a 1% increase in $\delta_{f(i,t)}$ within a CZ. Given the 30-35% gap in the mean of α_i between Black and white workers in our setting, one might have expected a roughly 3 log point gap in average pay premiums between Blacks and whites just because of assortative matching, rather than the 0 that we observe.

We do find that Black workers are slightly less likely to work in higher-paying <u>industries</u>, but this is offset by the tendency to be employed at higher-premium workplaces within a given

[3] We also find a tendency for whites to work in higher-premium industries in the ACS. Using estimated premiums for 295 4-digit industries (discussed below) we find that the average industry premium is 3 log points higher for whites than Blacks.

industry. This pattern is potentially consistent with a longstanding fact about the U.S. labor market, which is that Black workers are more likely to be covered by unions than whites (e.g., Ashenfelter, 1972). Data from the unionstats.com website shows that the ratio of the Black to white union coverage rate for male workers was around 130% in the late 1970s, and averaged about 120% in years 2010-2018. Similarly, the relative coverage rate of Black females relative to white females was around 160% in the late 1970s and averaged about 120% in years 2010-2018.

One possibility that we are exploring in ongoing research is that Black workers are actually closer to higher-premium jobs than white workers in many CZs – rather than further away, as is suggested by the Spatial Mismatch hypothesis – and that this relative proximity offsets the tendency we would expect from assortative matching for Blacks to work at lower-paying jobs. It is an open question whether this can explain the interesting pattern of between- and within- industry wage differences we see in the LEHD.

E. Job access and commuting patterns by race

i) Job access

Figure 1 of the main paper shows the share of all jobs, or of all good jobs, within radius r of the average Black and white worker in each group of CZs. We interpret this as showing that Black workers are not systematically farther from jobs or from good jobs than are white workers. But this may be misleading about access to jobs if not all jobs are available to all workers. To try to assess this, we separate workers by their level of education, high school or less vs. college or more. We similarly separate jobs by whether they are held by high school or college workers, assuming that jobs are only available to workers of the same education level as their incumbent workers. For each education group, we compute the share of all jobs for that education group that are within radius r of the average white and Black worker of that education group. Results are shown in **Figure A-1**. The general story is similar to that seen in Figure 1 – in both the high school and college labor markets, Black workers tend to live closer to jobs, and closer to good jobs in particular, than do white workers.

ii) Commute distances and job quality

Figure A-2 shows the distribution of commute distances for Black and white workers in our two groups of CZs. This is constructed from estimates of the kernel density of log commutes, which we then convert to CDFs. **Table A-4** reports selected quantiles for various groups of CZs. For comparison, we also show the quantiles of 1-way commute times (not distances) from the ACS. The LEHD data on commute distances show that in the older industrial cities, Black workers have generally shorter commute distances than whites (e.g., a median distance of 7.8 miles for Blacks versus 9.5 miles for whites), whereas in the Sunbelt cities Black workers have longer commute distances at quantiles up to the median (e.g., a median commute distance of 9.9 miles for Blacks versus 9.2 miles for whites). Looking at commute times in the ACS, we see identical quantiles for Blacks and whites in the older industrial cities, but slightly higher commute times at the median for Black workers in the Sunbelt. Since Black workers are more likely than whites to use buses and subways, and these modes have longer average travel times, we interpret the ACS travel time data as being consistent with the LEHD travel distances.[4]

Figure 2 shows the relationship between commute distance and the establishment wage premium for Black and white workers in each group of CZs. One concern is that errors in the imputation of establishments within multi-establishment firms may create bias in this relationship. To assess this, we estimate the relationship separately for workers in five groups of firms – the smallest firms, with no more than 10 workers in any quarter; larger firms that are still below median in size, separately by whether they have one or multiple establishments; and above-median firms, again separately by whether they have one or multiple establishments. Any imputation error would affect only the workers at multi-establishment firms. **Figure A-3** shows that the relationship is generally similar for all groups of firms.

However, the decline in establishment premiums for workers with the longest commutes is seen only among those at multi-establishment firms. This is consistent with the

[4] In the 30 largest CZs 88% of whites commute by car versus 81% of Blacks. Mean commute times are about 20 minutes longer for commuters who use buses or other transit modes (except walking) relative to car.

hypothesis that it might be spurious, due to imputation errors. The establishment imputation model does not take account of establishment pay premiums, and thus will not reflect that workers may be more likely to commute past one establishment to another when the farther establishment offers better pay.

Another interesting relationship that our framework lets us examine is the relationship between worker skill and commute distance. This is ex ante ambiguous – commute time has higher opportunity cost for high ability workers, but these workers may consume some of their higher income via more housing space, requiring a move to the suburbs. **Figure A-4** shows the distribution of commuting time by worker type (deciles of α_i), race, and CZ grouping. In both groups of CZs, high-skill workers commute modestly farther than do low-skill workers, though the dispersion within decile is quite large. In the older industrial CZs, the white commute distance distribution stochastically dominates the Black distribution in each ability decile, except possibly among the lowest-skill workers. In the newer Sunbelt CZs, Black workers tend to commute farther than similarly-skilled white workers, but this reverses for the lowest-skilled workers at the highest commute lengths.

Table A-5 presents the estimated elasticities of earnings, and the components of earnings attributable to person effects and workplace pay premiums, with respect to commute distance by CZ group and race. The elasticity of earnings is slightly higher for whites than Blacks in the older industrial cities, but lower for whites than Blacks in the Sunbelt cities and in the remainder group of CZs. Interestingly, however, when we look separately at the parts of earnings attributable to worker skills and establishment wage premiums, we see that the elasticity of the personal skills component is uniformly <u>higher</u> for Black workers, whereas the elasticity of workplace premiums is uniformly <u>lower</u> for Black workers. This suggests that the positive relationship between potential earnings capacity and commute distance is stronger for Blacks than whites, but that access to better-paying jobs, conditional on worker skills, is if anything better for Black workers (consistent with the main findings in our paper).

Table A-6 presents a parallel set of estimates based on our ACS samples. Since the ACS only reports the average (one-way) time taken for commuting, we include a set of dummies for the mode of transit. The elasticities of earnings with respect to commute time range from 0.08

to 0.10 for whites and 0.06 to 0.07 for Blacks. In three of the four city groups we estimate that the elasticity is *lower* for Blacks than whites, though in the Sunbelt cities the elasticity is very slightly higher for Black workers (0.071 versus 0.069).

We cannot decompose earnings in the ACS into person and workplace effects. As an alternative, we estimated a relatively rich cross-sectional wage model (separately by race) that included 295 4-digit industry effects. This allows us to decompose an individual earnings observation into a part attributable to the industry of employment, a part due to other observed covariates, and an unexplained part. We then regressed the industry component on commute time and obtained the set of elasticities shown in the second row of each panel in Table A-6. For both Blacks and whites we estimate that longer commute times are associated with employment in higher-paying industries: the elasticities are in the range of 0.024-0.028 for whites and 0.018 to 0.026 for Blacks – not too different from the elasticities of workplace pay premiums with respect to travel distance we obtained in the LEHD. Again, the elasticity of industry pay premiums tends to be slightly lower for Blacks, suggesting that if anything Black workers have slightly *better access* to higher-paying industries, except in the Sunbelt cities where whites and Blacks have similar access.

Additional References

Abowd, John, Francis Kramarz, and David N. Margolis (1999). "High wage workers and high wage firms." *Econometrica* 67(2): 251-333.

Ashenfelter, Orley (1972). "Racial discrimination and trade unionism." *Journal of Political Economy* 80(3, part 1): 488-508.

Card, David, Ana Rute Cardoso, Jörg Heining and Patrick Kline (2018). "Firms and labor market inequality: Evidence and some theory." *Journal of Labor Economics* 36(S1): 13-70.

Card, David, Jesse Rothstein, and Moises Yi (2023). "Location, Location, Location." NBER Working Paper No. 31587, August 2023.

Card, David, Jesse Rothstein, and Moises Yi (2024). "Industry wage differentials: A firm-based approach." *Journal of Labor Economics*, forthcoming. Available as NBER Working Paper No. 31588, August 2023.

Ellwood, David T. (1986). "The spatial mismatch hypothesis: Are there teenage jobs missing in the ghetto?" In Richard Freeman and Harry Holzer, editors, *The black youth employment crisis*. Chicago: University of Chicago Press, pp. 147-190.

Gerard, François, Lorenzo Lagos, Edson Severnini, and David Card (2021). "Assortative matching or exclusionary hiring? The impact of employment and pay policies on racial wage differences in Brazil." *American Economic Review* 111(10): 3418-57.

Glaeser, Edward L., Eric A. Hanushek, and John M. Quigley (2004). "Opportunities, race and urban location: the influence of John Kain." *Journal of Urban Economics* 56: 70-79.

Holzer, Harry J. (1991). "The spatial mismatch hypothesis: what has the evidence shown?" *Urban Studies* 28(1): 105-122.

Hutchinson, Paul (1974). "The effects of accessibility and segregation on the employment of the urban poor." In George von Furstenburg, Bennett Harrison, and Ann R. Horowitz, editors. *Patterns of Racial Discrimination* Volume 1 . Lexington, MA: Lexington Books.

Kain, John F. (1964). "The effect of the ghetto on the distribution and level of nonwhite employment in urban areas." Proceedings of the Annual Meetings of the American Statistical Association.

Kain, John F. (1968). "Housing segregation, negro employment, and metropolitan decentralization." *Quarterly Journal of Economics* 82 (2): 175-197.

Masters, Stanley H. (1974). "A note on John Kain's `Housing segregation, Negro employment, and metropolitan decentralization.'" *Quarterly Journal of Economics* 88 (3): 505-512.

Mooney, Joseph D. (1969). "Housing segregation, Negro employment, and metropolitan decentralization: an alternative perspective." *Quarterly Journal of Economics* 83(2): 299-311.

Offner, Paul and Daniel H. Saks (1971). "A note on John Kain's `Housing segregation, Negro employment, and metropolitan decentralization.'" *Quarterly Journal of Economics* 85(1): 147-160.

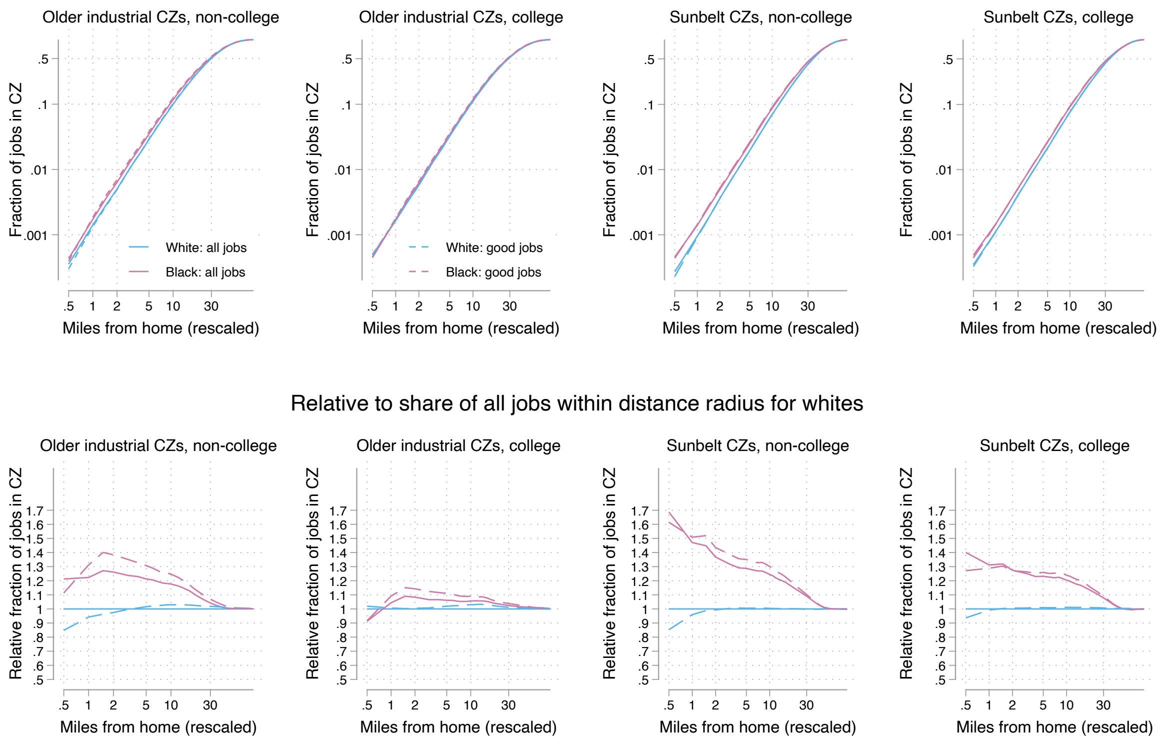
Tolbert, Charles M., and Molly Sizer (1996). "U.S. commuting zones and labor market areas: A 1990 update." U.S. Department of Agriculture, Economic Research Service, staff paper number AGES-9614.

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

Wilson, Valerie and William Darity Jr. (2022). "Understanding black-white disparities in labor market outcomes requires models that account for persistent discrimination and unequal bargaining power." Unpublished working paper. Washington DC: EPI

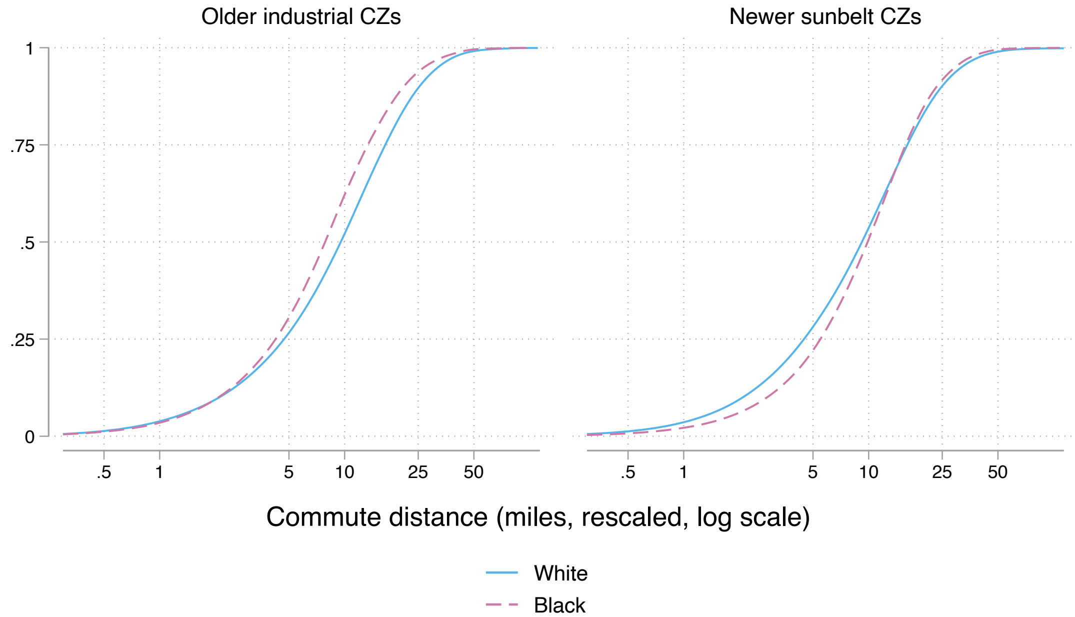
Wilson, William Julius (1987). *The truly disadvantaged: The inner city, the underclass, and public policy*. Chicago: University of Chicago Press.

Figure A-1. Job access in college and non-college labor markets



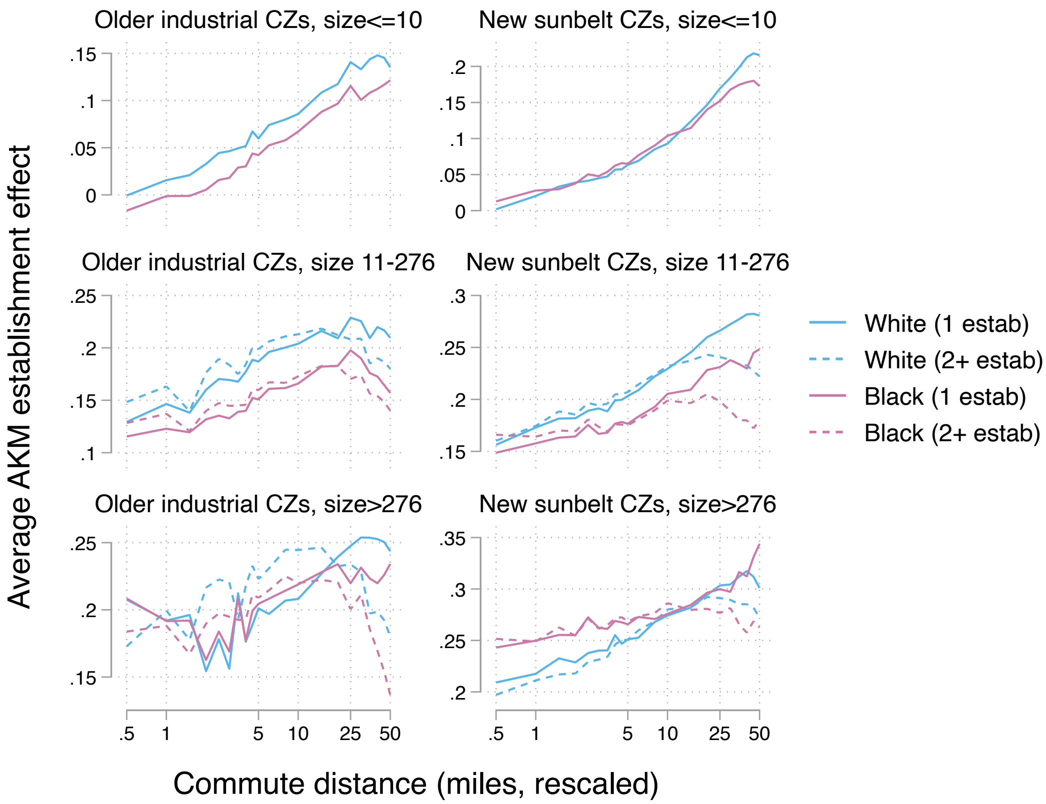
Notes: Distances for each CZ are rescaled to set the 75th percentile commute distance to 16 miles. "Good jobs" are those at establishments with AKM establishment effects in the top tercile. Bottom panels show fraction of jobs in a CZ relative to fraction of jobs available within distance radius for whites with the same level of education.

Figure A-2. CDFs of commute distribution, by CZ group and race



Notes: We estimate kernel densities of log standardized commute time using pooled samples across CZs in each group, separately by race. Commute distances are standardized to set the 75th percentile commute distance in each CZ to 16 miles. Kernel density estimates use an Epanechnikov kernel and Stata's default bandwidth for the white sample, and are evaluated at each multiple of 0.1 miles up to 10 miles, and then at each mile up to 100 miles. We linearly interpolate the estimated densities and construct CDFs from them.

Figure A-3. AKM establishment effect and commute distance, by firm type, CZ group, and race



Notes: Commute distances are standardized to set the 75th percentile commute distance in each CZ to 16 miles. Rows categorize firms by the maximum number of workers observed at the firm in any quarter; within each panel, series distinguish firms with a single establishment vs. more than one.

Figure A-4. Commute distances by AKM worker effect decile, race, and CZ group

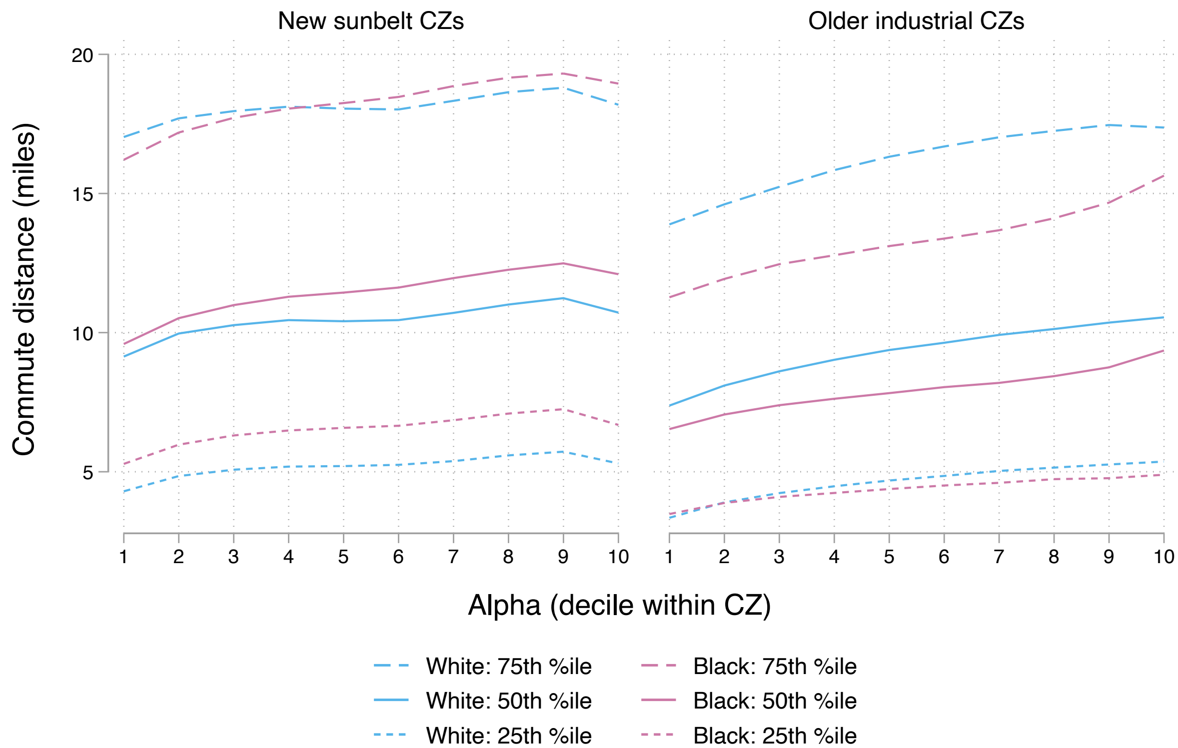


Table A-1. LEHD summary statistics and AKM decomposition

	Older industrial CZs	Newer sunbelt CZs	All other CZs among largest 200
All workers			
Geometric mean earnings	13,521	14,458	11,673
SD of log earnings	0.64	0.69	0.58
No. of person-quarters (millions)	311.4	203.1	1077.0
White workers			
Geometric mean earnings	14,271	15,835	12,064
No. of person-quarters (millions)	262.1	156.6	931.7
Black workers			
Geometric mean earnings	10,067	10,993	9,396
No. of person-quarters (millions)	49.2	46.5	145.3
White-Black gap			
Log earnings	0.35	0.36	0.25
Components of AKM decomposition			
Person effect	0.37	0.38	0.28
Establishment effect	0.00	0.00	-0.01
Xb	-0.01	-0.02	-0.01
Decomposition of estab. effects			
Industry effect	0.02	0.01	0.00
Within-industry effect	-0.02	-0.01	-0.01

Notes: Source is 2010-2018 LEHD. Sample includes only white and Black non-Hispanic people with quarterly earnings above $3,800, and excludes the first and last quarter of each employment spell. Person effects and establishment effects are obtained from AKM model (equation 1) with controls for age (cubic) and calendar quarter (indicators).

Table A-2. Summary Statistics for Four Groups of Larger CZ's

	Top 30 CZ's	Older Industrial	Newer Sunbelt	Northeast Corridor	Remainder (mostly West)
Share of Obs (%)	100.0	28.6	32.4	19.0	20.0
Demographics of Working Age Population					
White non-Hispanic (%)	52.0	64.1	38.7	51.9	56.4
Black non-Hispanic (%)	13.7	16.6	13.2	16.7	7.5
Hispanic (%)	22.6	11.3	36.3	19.0	20.0
Asian non-Hispanic (%)	11.8	8.1	11.8	12.5	16.3
Mean Years of Education	13.6	13.8	13.1	14.0	13.7
BA or higher (%)	36.3	37.1	31.2	43.2	36.8
Employed (%)	80.5	81.0	79.0	81.8	81.0
Full-time Earner (%)	61.8	63.3	58.4	64.7	62.2
Mean Earnings (with 0's)	41,976	42,197	36,913	49,322	42,878
Characteristics of Full Time Wage and Salary Earners					
Black non-Hispanic (%)	12.3	13.6	12.7	15.7	6.6
Female	45.1	46.0	43.7	46.8	44.4
Mean Years of Education	14.2	14.4	13.8	14.6	14.2
BA or higher (%)	44.4	45.1	39.2	51.7	44.2
Mean Earnings	66,064	64,812	60,986	74,618	67,170
Mean Hourly Wage	31.33	30.62	29.14	35.09	31.99
Use Car to Commute (%)	86.0	88.4	94.3	67.2	88.8
Mean Commute Time (mins)	30.9	29.9	30.0	35.3	29.4
Earnings of Full Time Wage and Salary Workers					
Mean Log Earnings - White NH	10.96	10.90	10.97	11.09	10.93
Mean Log Earnings - Black NH	10.64	10.61	10.61	10.74	10.61
Black-white Gap	-0.32	-0.30	-0.36	-0.35	-0.32

Source: 2010-2018 ACS public use files. Adult population includes people age 22-62 with age> education+6. Full time earners have annual earnings above $15,200. Older industrial CZs are Philadelphia, Detroit, Pittsburgh, Cleveland, Newark NJ, Buffalo, Baltimore, Chicago, Minneapolis, and St. Louis. Newer Sunbelt CZs are Los Angeles, Houston, Atlanta, Miami, San Diego, Phoenix, and Dallas. Northeast Corridor CZs are New York, Washington DC, Boston, and Hartford. Remaining CZs are San Francisco, Seattle, Denver, Sacramento, San Jose, Portland, Tampa, Orlando and Fort Worth.

Table A-3. Characteristics of Commuting Zones (CZ's) in Four Major Groups of Larger CZ's

	Working Age Population x1000	Working Age Adults (22-62 with Positive Experience) White NH	Black NH	Hispanic	Asian NH	Immigrant	BA or more	Full Year Earner	Full Year Earners Only Annual Wage & Sal. Earnings	One-way Comm. Time
Older Industrial Cities:										
Chicago	4,764	53.4	16.4	21.5	8.8	25.6	38.9	63.7	66,155	33.3
Newark	3,429	48.6	13.8	23.4	14.2	36.6	42.3	65.0	75,162	34.2
Philadelphia	3,237	63.2	19.5	9.4	7.9	14.0	36.3	62.4	65,846	30.7
Detroit	2,827	69.2	20.8	3.8	6.2	11.1	30.9	57.2	60,445	27.8
Minneapolis	1,871	77.7	7.7	5.3	9.3	14.2	41.8	70.1	65,041	25.8
Baltimore	1,517	57.8	29.1	5.2	7.9	13.2	39.0	66.7	67,591	31.9
Cleveland	1,377	74.2	18.0	3.6	4.2	6.6	31.5	61.1	56,743	25.2
St Louis	1,336	73.4	19.4	2.6	4.6	6.5	34.9	63.6	59,286	26.2
Pittsburgh	1,334	87.1	7.7	1.5	3.7	4.2	35.4	63.1	57,925	28.0
Buffalo	1,257	79.3	10.9	5.2	4.7	7.4	32.7	62.3	54,583	22.2
Newer Sunbelt Cities:										
Los Angeles	10,272	32.6	6.5	44.6	16.4	40.3	29.1	55.7	61,037	31.0
Houston	3,446	38.0	17.3	35.3	9.4	32.4	30.2	59.8	64,352	30.3
Atlanta	2,822	46.5	34.8	10.5	8.2	20.1	38.3	61.7	62,790	32.3
Dallas	2,562	46.0	16.6	27.7	9.7	28.4	34.8	63.7	63,323	28.8
Miami	2,577	24.2	20.3	51.6	3.9	53.1	29.7	57.1	53,723	30.2
Phoenix	2,421	57.5	5.3	29.0	8.2	20.2	28.4	59.4	57,027	27.0
SanDiego	1,826	47.7	5.0	31.5	15.8	30.6	35.1	59.8	62,988	26.0
Northeast Corridor Cities:										
NYC	6,993	41.6	18.3	25.5	14.6	40.5	39.5	60.8	73,448	39.2
Washington DC	3,296	45.9	25.4	15.1	13.5	30.3	50.4	70.9	78,587	35.4
Boston	2,967	72.2	7.3	10.4	10.1	23.0	46.8	67.0	73,401	32.0
Hartford	1,946	68.3	10.1	14.9	6.7	18.5	39.0	64.9	73,128	27.1
Remaining Cities in Top 30:										
San Francisco	2,981	41.6	7.8	21.6	29.1	36.8	45.5	63.5	81,992	33.1
Seattle	2,616	68.9	5.1	8.7	17.4	20.1	37.8	64.8	67,805	30.2
Denver	1,765	68.4	4.9	20.1	6.7	15.8	42.7	67.0	65,303	27.6
Sacramento	1,711	52.1	6.7	23.6	17.6	25.3	27.3	55.3	60,366	29.4
Tampa	1,571	64.3	11.7	18.3	5.7	17.0	28.8	59.4	54,652	27.8
San Jose	1,488	35.5	2.4	31.1	31.0	46.3	43.4	62.5	83,949	27.6
Ft. Worth	1,311	55.7	13.4	24.2	6.7	19.8	28.0	62.3	57,988	28.6
Orlando	1,326	50.4	15.2	27.8	6.6	21.9	29.8	58.9	51,650	28.5
Portland	1,286	75.5	2.8	10.4	11.3	16.8	37.1	62.3	61,163	26.7

Source: 2010-2018 ACS Public Use Files. Working age population includes people 22-62 with positive experience. Full year earnings have at least $15,200 in annual wage and salary earnings. Size of working age population is based on average weighted count of ACS sample in 2010-2018. Commuting zones are based on 1990 CZ definitions.

Table A-4. Quantiles of commute distance or commute time by CZ group and race

	Older Industrial CZs		Newer Sunbelt CZs		Northeast Corridor CZs	Remaining CZs
	Miles (LEHD)	Minutes (ACS)	Miles (LEHD)	Minutes (ACS)	Minutes (ACS)	Minutes (ACS)
Percentiles for White Non-Hispanics						
10	2.1	10	2.1	10	10	10
25	4.7	15	4.5	15	15	15
50	9.5	25	9.2	25	30	25
75	16.6	40	16.2	40	45	40
90	25.3	60	24.9	60	60	55
Percentiles for Black Non-Hispanics						
10	2.1	10	2.8	10	15	10
25	4.2	15	5.5	15	20	15
50	7.8	25	9.9	30	30	25
75	13.4	40	15.9	40	55	40
90	20.9	60	23.3	60	70	60

Notes: Miles represent distances from home to work, from LEHD. Distances are standardized across CZs to set the 75th percentile commute distance in each CZ to 16 miles. Minutes represent commute times, from ACS.

Table A-5: Elasticity of Earnings and Earnings Components w.r.t. Commute Distance

	Older Industrial	Newer Sunbelt	Remainder (Top 200)
White non-Hispanic			
Log Quarterly Earnings	0.059	0.024	0.031
	(0.000)	(0.000)	(0.000)
Person Effects	0.027	0.001	0.007
	(0.000)	(0.000)	(0.000)
Firm Premium	0.030	0.024	0.023
	(0.000)	(0.000)	(0.000)
Black non-Hispanic			
Log Quarterly Earnings	0.056	0.048	0.056
	(0.000)	(0.000)	(0.000)
Person Effects	0.035	0.026	0.032
	(0.000)	(0.000)	(0.000)
Firm Premium	0.016	0.019	0.020
	(0.000)	(0.000)	(0.000)

Source: 2010-2018 LEHD. Sample includes only people age 22-62 with quarterly earnings above $3,800. Person effects and firm premium represents estimated person and establishment earning effect obtained from AKM model (equation 1) with controls for age and calendar quarter. Coefficient estimates in table are obtained from specifications that regress log of quarterly earnings and AKM components on log of commute distance with CZ controls.

Table A-6: Elasticity of Annual Earnings w.r.t. Commute Time

	Top 30 CZs	Older Industrial	Newer Sunbelt	Northeast Corridor	Remainder (mainly West)
White non-Hispanic					
Log Annual Earnings	0.090	0.103	0.070	0.099	0.079
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Industry Premium	0.028	0.030	0.026	0.028	0.027
(295 industries)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Black non-Hispanic					
Log Annual Earnings	0.069	0.063	0.077	0.066	0.069
	(0.002)	(0.003)	(0.003)	(0.003)	(0.005)
Industry Premium	0.023	0.021	0.028	0.021	0.023
(295 industries)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)

Source: 2010-2018 ACS public use files. Sample includes only people age 22-62 with positive experience (age-education>6) and annual earnings above $15,200. Industry premium represents estimated industry wage effect received by worker, obtained from model fit by gender to all 30 of the largest CZ's, with controls for education, experience, race, immigrant status and CZ effects. Coefficient estimates in table (with robust standard errors) are obtained from specifications that regress log of annual earnings on log of commute time with controls for gender, mode, and CZ.