Online Appendix

The Rising Value of Time and the Origin of Urban Gentrification

Yichen Su

The appendix contains a few sections. Section A presents some extra details regarding my spatial equilibrium model. Section B contains the data appendix, in which I discuss how I obtain census tract residential locations, zip-code level job locations, and the procedure through which I obtain the travel time matrix. Section C describes some extra descriptive statistics that augment the analysis in the main text. Section D contains the estimation appendix, in which I discuss the estimation of long-hour premium and various validation tests for long-hour premium, lasso analysis of occupation characteristics, alternative measurement of the value of time, and some technical details of the estimation procedure. Section E contains a set of alternative exercises that shed light on the full effect of the changing value of time on gentrification.

A Model

A.1 Worker's location choice problem

This section details the solution procedure that derives workers' indirect utility function, given location characteristics (rents and amenities) and workers' occupation, from the basic assumption of Cobb-Douglas utility function.

The utility-maximization problem is

$$\max_{C,H} U\left(C, H, A_{jmt}, c_{jnmt}\right) = C^{\theta} H^{1-\theta} A_{jmt}^{\tilde{\gamma}_k} \exp\left(-\tilde{\omega}_t c_{jnmt}\right) \exp\left(\sigma_k \varepsilon_{i,jmt}\right)$$

subject to budget constraint

$$C + R_{jmt} H = \exp(y_{0kt} + v_{kt}\left(T - c_{jnmt}\right))$$

By Cobb-Douglas functional form, the demand for tradable consumption and housing services is (let I denote weekly earnings):

$$C^* = \theta I$$

$$H^* = \frac{1-\theta}{R_{jmt}} I$$

The log-transformed partial indirect utility is then:

$$V_{i,jnmt} = \theta \log\left(\theta I\right) + (1-\theta) \log\left(\frac{1-\theta}{R_{jmt}} I\right) - \tilde{\omega}_t c_{jnmt} + \tilde{\gamma}_k a_{jmt} + \tilde{\gamma}_k \zeta_{jmt} + \sigma \varepsilon_{i,jmt}.$$

The equation can be simplified with some algebra manipulation and by substitute I with the earnings equation.

$$
\begin{aligned}
V_{i,jnmt} = \theta \log(\theta) + (1-\theta)\log(1-\theta) + (y_{0kt} + v_{kt}(T - c_{jnmt})) \\
-(1-\theta)\log(R_{jmt}) - \tilde{\omega}_t c_{jnmt} + \tilde{\gamma}_k a_{jmt} + \tilde{\gamma}_k \zeta_{jmt} + \sigma \varepsilon_{i,jmt}
\end{aligned}
$$

I then re-normalize the indirect utility function by dividing the entire utility function by σ_k, so that I can interpret the all coefficients as migration elasticities.

$$
\begin{aligned}
V_{i,jnmt} = \frac{1}{\sigma_k}(\theta \log(\theta) + (1-\theta)\log(1-\theta)) \\
+\frac{1}{\sigma_k}(y_{0kt} + v_{kt}(T - c_{jnmt})) - \frac{(1-\theta)}{\sigma_k}\log(R_{jmt}) \\
-\frac{\tilde{\omega}_t}{\sigma_k} c_{jnmt} + \frac{\tilde{\gamma}_k}{\sigma_k} a_{jmt} + \frac{\tilde{\gamma}_k}{\sigma_k}\zeta_{jmt} + \varepsilon_{i,jmt}
\end{aligned}
$$

I simplify the above equation by combining terms. By doing so, I arrive at the following equation which is the one presented in the main body of the paper.

$$
V_{i,jnmt} = \delta_{kt} - \mu_k v_{kt} c_{jnmt} - \omega_{kt} c_{jnmt} - \beta_k r_{jmt} + \gamma_k a_{jmt} + \gamma_k \zeta_{jmt} + \varepsilon_{i,jmt}
$$

Each coefficient is written in terms of the underlying parameters:

$$
\delta_{kt} = \frac{1}{\sigma_k}(\theta \log(\theta) + (1-\theta)\log(1-\theta)) + \frac{1}{\sigma_k}\underbrace{(y_{0kt} + v_{kt}T)}_{\text{gross earnings}}
$$

$$
\begin{aligned}
\mu_k &= \frac{1}{\sigma_k} \\
\beta_k &= \frac{1-\theta}{\sigma_k} \\
\gamma_k &= \frac{\tilde{\gamma}_k}{\sigma_k} \\
\omega_{kt} &= \frac{\tilde{\omega}_t}{\sigma_k}
\end{aligned}
$$

A.2 Worker's problem with leisure choice

In this section, I derive a utility specification with leisure as a choice. I show that such specification is empirically equivalent to the original specification used in the paper. I solve the problem in two steps. In the first step, I solve the utility-maximization problem holding leisure hours fixed, making the problem a standard utility-maximization of the Cobb-Douglas utility function. Once I obtain the partial indirect utility function given each level of leisure, I then solve for optimal leisure hours and the indirect utility function. Finally, I normalize the indirect utility function with the standard deviation of the idiosyncratic component of worker's preference.

Step 1: Solve for partial indirect utility given leisure consumption.

Given the workers' utility function of C, H and L, I fix L first and solve for C and H first. Let $\theta_L = 1 - \theta_C - \theta_H$.

$$\max_{C,H} U\left(C, H, L, A_{jmt}, c_{jnmt}\right) = C^{\theta_C} H^{\theta_H} L^{\theta_L} A_{jmt}^{\tilde{\gamma}_k} \exp\left(-\tilde{\omega}_t c_{jnmt}\right) \exp\left(\sigma_k \varepsilon_{i,jmt}\right)$$

subject to budget constraint

$$C + R_{jmt} H = \exp(y_{0kt} + v_{kt}\left(T - L - c_{jnmt}\right))$$

By Cobb-Douglas functional form, the demand for tradable consumption and housing services is (let I denote weekly earnings):

$$C^* = \frac{\theta_C}{\theta_C + \theta_H} I$$

$$H^* = \frac{\theta_H}{R_{jmt}\left(\theta_C + \theta_H\right)} I.$$

The log-transformed partial indirect utility given leisure consumption L is then:

$$V_{i,jnmt}\left(L\right) = \theta_C \log\left(\frac{\theta_C}{\theta_C + \theta_H} I\right) + \theta_H \log\left(\frac{\theta_H}{R_{jmt}\left(\theta_C + \theta_H\right)} I\right) + \theta_L \log\left(L\right) - \tilde{\omega}_t c_{jnmt} + \tilde{\gamma}_k a_{jmt} + \tilde{\gamma}_k \zeta_{jmt} + \sigma_k \varepsilon_{i,jmt}.$$

The equation can be simplified with some algebra manipulation and by substitute I with the earnings equation:

$$\begin{aligned} V_{i,jnmt}\left(L\right) = {} & \theta_C \log\left(\frac{\theta_C}{\theta_C + \theta_H}\right) + \theta_H \log\left(\frac{\theta_H}{\theta_C + \theta_H}\right) + \left(\theta_C + \theta_H\right)\left(y_{0kt} + v_{kt}\left(T - L - c_{jnmt}\right)\right) \\ & - \theta_H \log\left(R_{jmt}\right) + \theta_L \log\left(L\right) - \tilde{\omega}_t c_{jnmt} + \tilde{\gamma}_k a_{jmt} + \tilde{\gamma}_k \zeta_{jmt} + \sigma_k \varepsilon_{i,jmt}. \end{aligned}$$

Step 2: Solve for optimal leisure choice and the indirect utility function.

The maximization of the second step is the following.

$$\max_{L} V_{i,jnmt}\left(L\right)$$

It can be seen that leisure consumption increases utility through the term $\theta_L \log\left(L\right)$. But higher leisure hours means lower working hours, which reduces log earnings. Optimal leisure is obtained by solving the tradeoff problem.

A simple first-order condition leads to:

$$L^* = \frac{1 - \theta_C - \theta_H}{\left(\theta_C + \theta_H\right) v_{kt}}$$

Substituting optimal leisure back into the partial indirect utility, I obtain the indirect utility function.

$$
\begin{gathered}
V_{i,jnmt} = \theta_C \log\left(\frac{\theta_C}{\theta_C+\theta_H}\right) + \theta_H \log\left(\frac{\theta_H}{\theta_C+\theta_H}\right) \\
+\left(\theta_C+\theta_H\right)\left(y_{0kt} + v_{kt}\left(T - \frac{1-\theta_C-\theta_H}{\left(\theta_C+\theta_H\right)v_{kt}} - c_{jnmt}\right)\right) \\
-\theta_H \log\left(R_{jmt}\right) + \theta_L \log\left(\frac{1-\theta_C-\theta_H}{\left(\theta_C+\theta_H\right)v_{kt}}\right) - \tilde{\omega}_t c_{jnmt} + \tilde{\gamma}_k a_{jmt} + \tilde{\gamma}_k \zeta_{jmt} + \sigma_k \varepsilon_{i,jmt}
\end{gathered}
$$

I then re-normalize the indirect utility function by dividing the entire utility function by σ_k, so that I can interpret all the coefficients as migration elasticities.

$$
\begin{gathered}
V_{i,jnmt} = \frac{1}{\sigma_k}\left(\theta_C \log\left(\frac{\theta_C}{\theta_C+\theta_H}\right) + \theta_H \log\left(\frac{\theta_H}{\theta_C+\theta_H}\right)\right) \\
+\left(\frac{\theta_C+\theta_H}{\sigma_k}\right)\left(y_{0kt} + v_{kt}\left(T - \frac{1-\theta_C-\theta_H}{\left(\theta_C+\theta_H\right)v_{kt}} - c_{jnmt}\right)\right) - \frac{\theta_H}{\sigma_k}\log\left(R_{jmt}\right) \\
+\frac{1-\theta_C-\theta_H}{\sigma_k}\log\left(\frac{1-\theta_C-\theta_H}{\left(\theta_C+\theta_H\right)v_{kt}}\right) - \frac{\tilde{\omega}_t}{\sigma_k}c_{jnmt} + \frac{\tilde{\gamma}_k}{\sigma_k}a_{jmt} + \frac{\tilde{\gamma}_k}{\sigma_k}\zeta_{jmt} + \varepsilon_{i,jmt}
\end{gathered}
$$

I simplify the above equation by combining terms. By doing so, I arrive at the following equation, which is the one presented in the main body of the paper:

$$
V_{i,jnmt} = \delta_{kt} - \mu_k v_{kt} c_{jnmt} - \omega_{kt} c_{jnmt} - \beta_k r_{jmt} + \gamma_k a_{jmt} + \gamma_k \zeta_{jmt} + \varepsilon_{i,jmt}.
$$

Each coefficient is written in terms of the underlying parameters:

$$
\begin{aligned}
\delta_{kt} \quad = \quad & \frac{1}{\sigma_k}\left(\theta_C \log\left(\frac{\theta_C}{\theta_C+\theta_H}\right) + \theta_H \log\left(\frac{\theta_H}{\theta_C+\theta_H}\right)\right) \\
& +\left(\frac{\theta_C+\theta_H}{\sigma_k}\right)\left(y_{0kt} + v_{kt}\left(T - \frac{1-\theta_C-\theta_H}{\left(\theta_C+\theta_H\right)v_{kt}}\right)\right) + \frac{1-\theta_C-\theta_H}{\sigma_k}\log\left(\frac{1-\theta_C-\theta_H}{\left(\theta_C+\theta_H\right)v_{kt}}\right)
\end{aligned}
$$

$$
\mu_k = \frac{\theta_C+\theta_H}{\sigma_k}
$$

$$
\beta_k = \frac{\theta_H}{\sigma_k}
$$

$$
\gamma_k = \frac{\tilde{\gamma}_k}{\sigma_k}
$$

$$
\omega_{kt} = \frac{\tilde{\omega}_k}{\sigma_k}
$$

The specification with leisure choice is empirically equivalent to the derivation without leisure choice. The optimal leisure choice is accounted for by the occupation/time fixed effects.

A.3 Derivation of location demand equation

I start from the normalized indirect utility:

$$V_{i,jnmt} = \delta_{kt} - \mu_k v_{kt} c_{jnmt} - \omega_{kt} c_{jnmt} - \beta_k r_{jmt} + \gamma_k a_{jmt} + \gamma_k \zeta_{jmt} + \varepsilon_{i,jmt}.$$

Worker i then chooses residential neighborhood j within MSA m to maximize indirect utility. Since $\varepsilon_{i,jmt}$ is distributed as Type I Extreme Value, the probability that worker i would choose neighborhood j is given by a multinomial logit function (McFadden (1973)). Given city m where a worker lives and works, the worker's occupation k, and the neighborhood n which the worker works in, the probability of that worker choosing to live in neighborhood j is given by

$$s_{j|nmkt} = \frac{\exp\left(\tilde{V}_{jnmkt}\right)}{\sum_{j' \in J_m} \exp\left(\tilde{V}_{j'nmkt}\right)}$$

where $\tilde{V}_{jnmkt} = \delta_{kt} - \mu_k v_{kt} c_{jnmt} - \omega_{kt} c_{jnmt} - \beta_k r_{jmt} + \gamma_k a_{jmt} + \gamma_k \zeta_{jmt}$ is the mean utility of occupation k living in j and working in n.

If I observe the residential location choice conditional on work location in the data, I can back out $\tilde{V}_{jnmkt}$ directly from the data and model the mean utility directly. Unfortunately, I only observe unconditional location demand s_{jmkt}. To proceed, I assume, in equilibrium, for workers in each occupation k, the unconditional expected utility of working in any neighborhood n within the MSA is identical and remains identical over time. Under this simplifying assumption, I essentially take a partial equilibrium framework in which a firm's location choice would not be affected by the change in residential sorting over the period of the analysis. I denote the expected utility value of working in each neighborhood in MSA m as Λ_{mkt}. The worker's conditional residential choice probability is then given by the following equation:

$$s_{j|nmkt} = \exp\left(\tilde{V}_{jnmkt} - \Lambda_{mkt}\right)$$

$\Lambda_{nmkt} = \log\left(\sum_{j' \in J_m} \exp\left(\tilde{V}_{j'nmkt}\right)\right)$, which is the expected utility for workers in occupation k working in neighborhood n. Under the assumption that the expected utility of working in each location is identical, I set $\Lambda_{nmkt} = \Lambda_{mkt}$. Due to the limitation of the unconditional location choice data, this simplifying assumption is necessary.

Given the residential choice probability conditional on working in n, the unconditional residential choice probability is computed by weighting these conditional probabilities with the unconditional probability of working in neighborhood n in MSA m, which I denote as π_{nmkt}. Thus, the residential choice probability is:

$$s_{jmkt} = \sum_{n' \in J_m} \pi_{n'mkt} \cdot s_{j|n'mkt}$$

I assume the spatial distribution of jobs for each occupation- $\pi_{n'mkt}$ as exogenous to the model within the time frame of this analysis, and the cross-sectional variation in job location is driven by factors such as path-dependent patterns of industry clustering and firm agglomeration (Ellison and Glaeser (1997), Rosenthal and Strange (2004), Ellison, Glaeser, and Kerr (2010)). One example to illustrate this point is the concentration of financial-industry jobs in Lower Manhattan. This area has a high presence of financial jobs because financial firms are historically clustered around the southern tip of Manhattan, not because the southern tip of Manhattan is an ex-ante desirable place for financial workers to live.

After log transformation, I write the log location choice probability as a linear function of various location preference components.

$$\log\left(s_{jmkt}\right) = \underbrace{\tilde{\delta}_{mkt}}_{\text{fixed effects}} + \underbrace{\log\left(\sum_{n'\in J_m} \pi_{n'mkt} \exp\left(-\left(\omega_{kt} + \mu_k v_{kt}\right) \cdot c_{jn'mt}\right)\right)}_{\text{valuation of proximity to employment}}$$
$$- \underbrace{\beta_k r_{jmt}}_{\text{valuation of rent}} + \underbrace{\gamma_k a_{jmt}}_{\text{valuation of amenities}} + \underbrace{\gamma_k \zeta_{jmt}}_{\text{valuation of unobserved amenities}}$$

where $\tilde{\delta}_{mkt} = \delta_{kt} - \Lambda_{mkt}$

B Data

B.1 Residential location imputation procedure

The key dependent variable in this research is the location choice by workers of different occupations and how their location choice changes over time. The choice set for workers is the set of neighborhoods in the cities that the workers live in. The best geographic unit that captures the essence of a neighborhood would be the census tract. The boundary of a census tract is relatively stable over time, and census tracts are designed to be fairly homogeneous in terms of population and economic characteristics. Therefore, the census tract is the natural choice for the definition of neighborhood. Nevertheless, the lowest geographic identifier in the Census microdata released to the public in IPUMS is PUMA, which is a much more aggregate level than the census tract. The data that I use for occupation-specific location data at census tract level are resident count by occupation group from each census tract, provided by the NHGIS. I impute census tract level occupation-specific count of residents using census tract level summary statistics and PUMA level microdata. I document the imputation procedure below.

Since NHGIS only provide counts of residents at census tract level for aggregate occupation level K, I would only observe n_K^j for each census tract j. My goal is to impute the count of residents by detailed occupation level k, namely n_k^j. I do so by first imputing $\hat{\theta}_{k|K}^j$, which is the conditional probability of being in occupation k given one is in occupation-group K. I compute $\hat{\theta}_{k|K}^j$ using IPUMS microdata at PUMA level, assuming that $\hat{\theta}_{k|K}^j$ is the same for every census tract j within the same PUMA area. Then, finally compute the census tract level count of residents in occupation k, by multiplying the count of residents in occupation group K with the imputed probability of a worker being occupation k given he/she is in occupation group K.

$$\hat{n}_k^j = \hat{\theta}_{k|K}^j \cdot n_K^j$$

Once I get $\hat{n}_k^j$, I generate the location choice probability for each occupation and in each city in each year s_{jmkt}, which is the probability of living in census tract j, conditional on living in MSA m, working in occupation k and at year t. As I have noted in the footnote in the main manuscript, before I compute the location choice probability, I add one to each imputed $\hat{n}_k^j$ to avoid having to take log over probability zero. The share of each neighborhood among each type of workers will be used in the location choice model.

B.2 Employment location imputation procedure

The employment location information is derived from the ZCBP from the U.S. Census Bureau, which provides establishment counts by the employment size of business establishments. The dataset comes at the level of detailed SIC and NAICS code for each zip code from 1994 on, annually. Unfortunately, the dataset does not go back farther than 1994. Therefore, I use the employment location data in 1994 to proxy those in 1990. The spatial distribution of employment changes fairly slowly over time, so I expect the four-year difference in data is unlikely to bias the data significantly.

For each zip code z, I first impute the employment count n_h^z for each industry h using establishment count and establishment sizes. Establishment size data are in the form of tabulated count: count of establishments with 1-4 employees, 5-9 employees, etc. I sum up these establishment counts weighted by the mid-value of the employee counts to impute the total employment count for each industry in each zip code.[1] Then I use $\hat{\theta}_{k|h}$, which is the conditional probability of working in occupation k, given he/she works in industry h, to impute the number of employment in occupation k at zip code z. $\hat{\theta}_{k|h}$ is computed using contemporaneous national microdata from IPUMS.

$$\hat{n}_k^z = \sum_h n_h^z \cdot \hat{\theta}_{k|h}$$

The set of $\hat{n}_k^z$ measured for each zip code and each occupation will form the basis of the spatial distribution of employment. I use these data and the travel time matrix to compute the expected commute time for each census tract and for workers of each occupation.

B.3 Data acquisition procedure for travel time matrix

I acquire the travel time and travel distance from the Google Distance Matrix API (Application Programming Interface). The number of entries in the travel matrix from every census tract to every zip code within every MSA is more than 7 million (7,363,850), which is too large to extract from the API directly. One reason that such travel matrix suffers from the curse of dimensionality is

[1] The IPUMS uses CIC (Census Industry Codes) to classify industries. I obtain the CIC-SIC and CIC-NAICS crosswalks from U.S. Census Bureau (U.S. Census Bureau (2017b)).

that large metro areas such as New York contain a very large number of entries connecting numerous locations that are very far apart. For example, from eastern Long Island to Manhattan, there are tens of thousands of entries connecting all zip codes to all census tracts in Manhattan and eastern Long Island, even though most of these entries have almost identical travel times and distances. Hence, it is, in fact, not necessary to compute distance and time for all entries between census tracts and zip codes. I can group various zip code destinations and compute travel distance and time from all census tracts to one destination per zip code group if the trip distance is very long, and thereby reducing the dimensionality of the data dimension.

An intuitive real-life example that demonstrates this logic would be the use of GPS navigation for a long trip. When taking a long trip by car (such as from Palo Alto to San Francisco), setting the GPS destination in whichever specific location near downtown San Francisco would not make much of a difference because one has to get on the freeway and the exact location of the destination makes *relatively* little impact on the ETA. However, if one takes a trip that is around 3 to 4 miles that starts and ends within San Francisco, ETA would be sensitive to the exact location of the destination.

Motivated from this observation, I only directly extract travel distance and time information between census tracts and zip code for pairs that are located within 5 miles Euclidean distance (centroids of census tracts and long/lat of zip code gazetteer). For the pairs that are farther than 5 miles apart, I proxy the location of each zip code with the closest PUMA centroid, and I extract the travel distance and time between each census tract to the assigned PUMA centroid. It significantly reduces the dimensions required for the data extract.

B.4 Historical travel time

In this section, I describe how I generate the 1990 historical travel time matrix for each MSA. Why estimate historical travel speed? If Google map existed in 1990, I could easily compute the travel time matrix using the historical traffic data. Unfortunately, the Google traffic model is only applicable to today's traffic conditions. One obvious concern of using today's traffic condition is a measurement error problem. But a much bigger concern is that traffic condition is a highly local variable, and it is very likely to be endogenous to location demand. Here is an example of the endogeneity problem. An exogenous demand surge (e.g., amenity shock) for a certain location X makes traffic around location X more congested, which prolongs travel time to and from location X. The long travel time into and out of location X coupled with the observation of a demand surge for location X would lead the model to interpret that the demand surge is caused by people's desire to save on commute time. Using today's traffic model could introduce this "self-fulfiling prophecy" that could introduce a serious endogeneity problem into the estimation of the model. Hence, the historical travel time matrix needs to be traffic information from the past.

To that purpose, I use two sources of data, Google API and the 1995 National Household Travel Survey (NHTS), to impute the historical travel time matrix. I first impute the historical travel speed (using NHTS and Google) for all travel routes within MSAs in 1995 rush hour, and then multiply the historical travel speed with travel distance (from Google) for each route to get expected travel time.

First, I use Google Distance Matrix API to obtain travel time (with traffic model turned off) and travel distance from each census tract to each zip code within each MSA. I make sure that travel time from Google is derived under the condition that the trips take place at midnight so that no traffic is expected. The traffic-free travel time gives me information on the route fixed-effects (such as the slowing-down effect of crossing a bridge, windy road, or dense city blocks with traffic lights).

Second, I use the 1995 NHTS data to fit a simple traffic speed model (Couture 2016) so that I could take the parameters estimated in the model onto the observable neighborhood characteristics in the 1990 Census and predict historical travel speed. I model travel speed as following:

$$\log(\text{speed}_{jnt}) = \beta_{0,t} + \beta_{1,t}\log(\text{distance}_{jn}) + \beta_{2,t}\log(\text{distance}_{jn})^2 + \bar{\mathbf{X}}_{jn}\Gamma_t + d_{jn} + \epsilon_{jnt}$$

j is the origin census tract; n is the destination zip code; t is the year in which the trip is taken. I assume the log speed of the trip is a function of trip distance because longer trips usually have higher speeds as people take the freeway or use the main thoroughfare when the distance is long enough. I assume travel speed is also a function of the average neighborhood characteristics (population density, median income, and percentage of population working) of the origin and destination. Travel speed heavily depends on the types of neighborhoods in which the trips take place. A trip to or from densely populated neighborhoods is expected to experience heavier congestion than another trip which takes place in the suburbs. Additionally, I assume each route admits a time-invariant fixed-effects component, which accounts for the road conditions other than traffic congestion, such as a slowing-down effect of crossing a bridge, windy road, or dense city blocks with traffic lights. I assume these fixed effects do not change over time. The parameters of the model $\beta_{0,t}$, $\beta_{1,t}$, $\beta_{2,t}$, Γ_t governs how location characteristics and trip distance are mapped into travel speed. Since traffic condition evolves over time, these parameters are assumed to be year-specific.

I use the 1995 NHTS data to estimate these parameters to obtain parameters applicable to 1995 traffic conditions. I restrict the trip samples to those that take place Monday to Friday and with departure time between 6:30 to 10:30 am and between 4:30 to 8:30 pm. I also restrict the trips either originate from or destine toward respondents' location of residence. $\bar{\mathbf{X}}_{jn}$ takes the location characteristics of the census tract which respondent lives in (neighborhood characteristics for the other end of the trip is unavailable). Additionally, I use Google API travel time (with traffic model turned off) to estimate the fixed-effects d_{jn} for each route. I impute traffic speed using the following equation:

$$\log(\widehat{\text{speed}}_{jn,1995}) = \hat{\beta}_{0,1995} + \hat{\beta}_{1,1995}\log(\text{distance}_{jn}) + \hat{\beta}_{2,1995}\log(\text{distance}_{jn})^2 + \bar{\mathbf{X}}_{jn}\hat{\Gamma}_{1995} + \hat{d}_{jn}.$$

The travel time is then obtained by multiplying imputed travel speed with travel distance:

$$\text{time}_{jn,1995} = \exp\left(\log(\widehat{\text{speed}}_{jn,1995})\right) \cdot \text{distance}_{jn}.$$

Figure A15 provides two examples of driving times on maps of Chicago and New York.

C Descriptive statistics

C.1 Definition of central city neigbborhoods

As described in the descriptive statistics section of the paper, central city neighborhoods are defined as census tracts that fall within the 5-mile pin of downtown (defined by Holian and Kahn (2015)). In Figure A1, I show the maps of a few cities as examples. In the map, the pin is defined as the point of downtown. The smaller circle represents the 3-mile radius, and the larger circle represents the 5-mile radius. The definition of central city neighborhoods in the main manuscript of the paper is given by the 5-mile radius of the downtown pin.

C.2 Neighborhood change on the map (Chicago and New York)

The first descriptive fact that I show in the paper (Figure 1) is that the income ratio between the central city and suburban neighborhoods declined precipitously and reversed dramatically after 1980. The reversal of fortune in the central cities after 1980 is the main subject of this paper.

Therefore, to build intuition for such change after the 1980s, I demonstrate the neighborhood changes on maps for two prominent cities in the United States: Chicago and New York. I rank census tracts by income quintile within Chicago's MSA and New York's MSA, then plot the income quintile by the census tract's distance to downtown for the four decades from 1980 to 2010. Figure A6 shows that central city neighborhoods in Chicago are overwhelmingly low-income relative to the overall MSA income level in 1980, but after several decades of increase, central city neighborhood income levels are well above the overall MSA income level. A similar pattern can be observed in New York's central city neighborhoods in Figure A6. To various degrees, most major MSAs in the U.S. exhibit a similar pattern of income reversal between central cities and suburbs.

Furthermore, in Figure A7, I plot the census tract income quintile by distance to downtown for Chicago and New York. One can clearly see that the census tracts near downtown experienced a dramatic increase in their rank since 1980.

C.3 Central city population

The terms "gentrification" or "urban revival" may give the impression that central neighborhoods are now seeing faster overall population growth than the suburbs. However, while central neighborhoods may be gaining absolute population, they have not gained in terms of shares of overall MSA population since population growth in the suburbs continues to outpace that in central cities. Overall, American cities were still suburbanizing as recently as from 2000 to 2010, but at a much slower pace. Figure A8 shows the share of central neighborhoods' population as a percentage of the total metropolitan population in the 25 most populous MSAs. The revived demand for central neighborhoods comes primarily from high-income workers and not all workers.

C.4 Change in work hours and commute time by wage decile

In the paper, I highlight the fact that high-wage workers experienced a rising prevalence of working long hours and a slower growth of commute time between 1980 and 2010, which coincides with the episode of gentrification. If I zoom in, I find that the sharp increase in the prevalence of working long hours occurred mainly before 2000. Coincidentally, the strong negative relationship between the growth in commute time and wage decile also mainly occurred before 2000. After 2000, both high- and low-skilled workers actually were less likely to work long hours (although low-skilled workers' probability of working long hours decreased much more). Also, after 2000, the negative relationship between growth in commute time and wage decile disappears. In fact, the workers in the top wage decile actually experience a weakly stronger growth in commute time than workers in lower wage deciles do.

Figure A3 shows the changing probability of working long hours by wage decile for two different periods: 1. 1980 - 2000; 2. 2000 - 2010. Figure A5 shows the growth in commute time by wage decile for the same two periods. These facts are suggestive evidence that the rising value of time provided the *initial* force that attracted high-skilled workers into the central cities. Once the endogenous amenity process started, many high-skilled workers started to move into the city due to amenities rather than shorter commute. As amenity change evolved, the role of amenities started to overwhelm the role of shorter commute time. In fact, many high-skilled workers live in the central cities for the amenities even though they work in the suburbs. This explains why the high-wage workers experience slightly higher growth in commute time between 2000 and 2010. This evidence is also consistent with Couture and Handbury (2020)'s results in which reverse commuting became more prevalent after 2000.

Furthermore, Table A2 shows that model-predicted gentrification (by only changing the value of time) correlates much better with the sorting pattern during the 1990-2000 period than with 2000-2010, as shown by the R-squared. This is indirect evidence that the rising value of time may matter more in the first decade. Nevertheless, I do not take an explicit stance on the timing of gentrification, as my model framework is not dynamic.

D Estimation

D.1 Potential biases in estimating long-hour premium

The long-hour premium is measured off the cross-sectional relationship between weekly log earnings and weekly hours worked. One potential reason for a biased estimate for LHP (long-hour premium) is that weekly hours worked could result from workers' labor supply choice. Therefore the variable of hours worked may be endogenous.

In the context of my estimation, the variation that I use to identify the spatial equilibrium model is the differential *change* in the long-hour premium. While the endogeneity of the hours variable may overstate the size of the static estimate of long-hour premium, the real threat to identification is if the change in the estimated LHP within occupations is driven by the *changing* degree of sorting on earnings and hours described above.

To fix the idea, consider the case of financial workers. Over time, it is possible that high-ability financial workers increasingly supply longer hours and receive higher earnings, relative to the low-ability counterparts. Their increasing supply of long-hour may simply due to a *preference change*. Meanwhile, they receive higher earnings due to their *high abilities*. As a result of this increasing selection by abilities, I would observe an increasing association between high earnings and high work hours among financial workers. Such association may not be driven by the increasing payoff of working long hours.

If I observe workers' true abilities, I would re-estimate the long-hour premium controlling for the levels of ability and see whether controlling for abilities would change the estimate for LHP. The difference between LHP estimates with and without control for levels of ability indicates the degree of selection on workers' abilities. If the degree of selection on ability increases over time, it would raise suspicion that long-hour premium estimate may be driven by increasing selection effect.

Since I do not observe workers' unobservable abilities, I conduct a similar test on observable abilities: reported education levels. I assume that if there is increasing selection on the unobservable abilities, I should see the same increase in selection on the observable abilities, such as education levels (Altonji, Elder, and Taber (2005)).

To do that, I re-estimate the long-hour premium for several key occupations with and without controlling for education levels and compare LHP estimates.

Figure A12 shows the degree of selection on the observable skills for estimates of LHP in 1990 and 2010. The degree of selection is computed as the difference between the LHP estimates without education control and estimates with education control. Two observations can be made here: 1. there is a selection effect on the observable skill levels for almost all occupations for the level estimates of LHP in both 1990 and 2010; 2. the selection effect is larger for occupations with more skill content. These observations suggest that the level estimates of LHP are likely partly driven by the selection effect on the unobservable skill levels.

However, in Figure A13, I show the *change* in the degree of selection by observable skills. The selection effect, on average, is not increasing. Occupations are equally likely to see increasing and decreasing selection effects. In addition, the change in selection effect is not correlated with the skill content of occupations at all. Since the estimates for LHP is not driven by selection by the observables, the *change* in LHP estimated from cross-sectional data is unlikely to be driven by the *changing* degree of selection by the unobservables.

Table A1 in the appendix reports some of these estimates. The level estimates are smaller with education control. But the change in estimated LHP does not seem to sustain a substantial effect from adding education control.

D.2 Alternative measures of the value of time

Another variable that tracks the marginal earnings of hours supply could be constructed based on a "tournament scheme" of compensation, in which workers get paid with prizes from tournament competitions within firms or within labor markets (Lazear and Rosen (1981)). A "prize" such as a job promotion or securing lucrative projects is awarded to the workers who outperform their competitors. Under this scheme, increasing work effort can increase the chance of winning such a prize. If the reward of a "prize" is very high, the payoff of effort is thus likely very high, since even narrowly losing the "tournament" means missing the prize entirely. Therefore, the effort level is an increasing function of the prize *spread* between winning and losing. Since work hour is a crucial input of worker's effort level, the marginal earnings of hours supply would rise if the reward spread between winning and losing the "tournaments" becomes higher (Bell and Freeman (2001)). A measurement of log earnings dispersion within the same occupation could track the size of the "spread" of the financial reward for workers in the occupation. Therefore, I use the Census data and compute the standard deviation of the residual log earnings for each occupation, after controlling for the individual characteristics. I use it as an alternative measurement for the value of time to check the robustness of the main results (See Table 7 Column 4).

D.3 Lasso regression using O*NET occupation characteristics

I also assess the variation in the long-hour premium by projecting the change in long-hour premium onto the 57 occupation characteristics from O*NET (U.S. Department of Labor (2017)). O*NET classifies occupation with Stanford Occupational Classification (SOC), but IPUMS uses Census Occupation Code (OCC2010). I use SOC-OCC2010 crosswalk provided by the Census Bureau to link the two classification regimes (U.S. Census Bureau (2017b)). I standardize the occupation characteristics by their respective mean and standard deviation so that the variation in each variable is not confounded by the scale of each characteristic. I also standardize the outcome variables (change in long-hour premium and change in earnings dispersion).

The lasso coefficients is chosen by solving the following constrained minimization problem:

$$\min_{\beta_1 \cdots \beta_{57}} \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \beta_1 x_{1i} - \ldots - \beta_{57} x_{57i}\right)^2$$

$$\text{subject to } \sum_{j=1}^{57} \left|\beta_j\right| \leq t$$

y_i is the outcome variable, which is the change in long-hour premium. x_{ji} where $j = 1, \ldots 57$ are the 57 standardized occupation characteristics from O*NET. t is some size constraint for the norm of the coefficients. There is no intercept in the regression because standardized variables are centered around zero.

One could rewrite the minimization problem with a single equation and a Lagrange multiplier λ.

$$\min_{\beta_1 \cdots \beta_{57}} \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \beta_1 x_{1i} - \ldots - \beta_{57} x_{57i}\right)^2 + \lambda \sum_{j=1}^{57} \left|\beta_j\right|$$

λ is the weight that the regression gives to the norm of all the regression coefficients. When λ is zero, the lasso regression coefficients are identical to those estimated from OLS regression. With a large value of λ, I penalize large values of the coefficients on any of the explanatory variable, which forces the regression coefficients to drop out and become zero if the corresponding variables do not perform as well in predicting the variation in outcome variable and minimizing the mean squared residual. Therefore, with different levels of λ, regression coefficients would be different. A useful exercise to do would be to raise the size of λ incrementally and observe which explanatory variables drop out and which remain. Those that remain with a large size of λ tend to be those with the best explanatory power.

Finally, I use the variable selection and coefficients that give the minimum mean squared error under a 5-fold cross-validation.

Figure A14 shows the results of the Lasso analysis. Notably, "Time Pressure" is a key occupational characteristic that predicts the change in long-hour premium.

D.4 Linearization of location demand

To facilitate the estimation procedure, I linearize the location demand equation by evaluating the equation with Taylor approximation around $\omega_{zt}+\mu_z v_{kt}$ at some constant. One can think of $\omega_{zt}+\mu_z v_{kt}$ as the marginal disutility of commute time. I let $\omega_{zt}+\mu_z v_{kt} \approx \phi$, so that commute time is discounted with a constant coefficient ϕ. Taking derivative for $\log\left(\sum_{n'\in J_m} \pi_{n'mkt} \exp\left(-\left(\omega_{zt}+\mu_z v_{kt}\right) c_{jn'm}\right)\right)$ with respect to $\omega_{zt}+\mu_z v_{kt}$, leads to $-\widetilde{\mathrm{E}}_t\left(c_{jmk}\right)$ where $\widetilde{\mathrm{E}}_t\left(c_{jmk}\right)$ is the expected commute on an adjusted probability measure (the adjustment depends on the size of ϕ). Therefore, Taylor expansion around $\omega_{zt}+\mu_z v_{kt}=\phi$ equals the following equation:

$$
\begin{aligned}
\log\left(s_{jmkt}\right) \approx\ & \log\left(\sum_{n'\in J_m} \pi_{n'mkt} \exp\left(-\phi \cdot c_{jn'm}\right)\right)+\tilde{\delta}_{mkt}-\widetilde{\mathrm{E}}_t\left(c_{jmk}\right)\left(\omega_{zt}+\mu_z v_{kt}-\phi\right) \\
& -\beta r_{jmt}+\gamma_z \log\left(\frac{N_{jmt}^H}{N_{jmt}^L}\right)+\theta_{kt} X_{jm}+\xi_{jmkt}
\end{aligned}
$$

The nonlinear term $\log\left(\sum_{n'\in J_m} \pi_{n'mkt} \exp\left(-\phi \cdot c_{jn'm}\right)\right)$ can be approximated by

$$
\begin{aligned}
& \log\left(\sum_{n'\in J_m} \pi_{n'mkt} \exp\left(-\phi \cdot c_{jn'm}\right)\right) \\
\approx \; & \log\left(\sum_{n'\in J_m} \pi_{n'mk,t-1} \exp\left(-\phi \cdot c_{jn'm}\right)\right) \\
& + \frac{\sum_{n'\in J_m} \pi_{n'mkt} \exp\left(-\phi \cdot c_{jn'm}\right) - \sum_{n'\in J_m} \pi_{n'mk,t-1} \exp\left(-\phi \cdot c_{jn'm}\right)}{\sum_{n'\in J_m} \pi_{n'mk,t-1} \exp\left(-\phi \cdot c_{jn'm}\right)} \\
= \; & \log\left(\sum_{n'\in J_m} \pi_{n'mk,t-1} \exp\left(-\phi \cdot c_{jn'm}\right)\right) + \frac{\sum_{n'\in J_m} \pi_{n'mkt} \exp\left(-\phi \cdot c_{jn'm}\right)}{\sum_{n'\in J_m} \pi_{n'mk,t-1} \exp\left(-\phi \cdot c_{jn'm}\right)} - 1 \\
\approx \; & \delta_{jmk}.
\end{aligned}
$$

The term itself varies by j, m, k, t. To simplify, I decompose the term into two parts. The first part is the term evaluated with the initial job location. The second part is the ratio between the expected utilities evaluated with job locations at $t-1$ and job locations at t, *holding the distaste for commuting time constant* (ϕ). For feasibility reason, I assume that the ratio is constant across occupations. If jobs with a rising value of time are not becoming more concentrated in the initial locations, this assumption would not affect my estimation. Under this assumption, the first term becomes a j and k specific constant, which I write as a fixed-effects term δ_{jmk}. After some algebraic arrangement, the location demand equation can be approximated as following:

$$
\log\left(s_{jmkt}\right) \approx \delta_{jmk} + \tilde{\delta}_{mkt} - \left(\phi + \omega_{zt}\right) \widetilde{\mathrm{E}}_t\left(c_{jmk}\right) - \mu_z v_{kt} \widetilde{\mathrm{E}}_t\left(c_{jmk}\right) - \beta_z r_{jmt} + \gamma_z \log\left(\frac{N^H_{jmt}}{N^L_{jmt}}\right) + \theta_{kt} X_{jm} + \xi_{jmkt}.
$$

To provide some intuition for the expected commuting time, I construct the maps of expected commute time for some selected occupations across neighborhoods in Chicago and New York in Figure A16.

D.5 Instrumental variables - predicted change in population

Identifying the preference parameters for amenities and rents γ and β relies on the construction of the predicted change in population for high- and low-skilled workers in each neighborhood j. According to the structural model, the predicted population can be written in the following format:

$$
\hat{N}_{jmkt} = N_{mk,t-1} \cdot \frac{\exp\left(\log\left(s_{jmk,t-1}\right) - \hat{\mu}\Delta\hat{v}_{kt}\widetilde{\mathrm{E}}_{t-1}\left(c_{jmk}\right)\right)}{\sum_{j'\in J_m} \exp\left(\log\left(s_{j'mk,t-1}\right) - \hat{\mu}\Delta\hat{v}_{kt}\widetilde{\mathrm{E}}_{t-1}\left(c_{j'mk}\right)\right)}.
$$

The equation is a function of initial location choice, expected commute time, and the change in the value of time for each occupation. Importantly, the magnitude of the predicted population change also depends on the size of $\hat{\mu}$. While the size of $\hat{\mu}$ is important in determining the magnitude of predicted population change, we show in this section that the size does not change the identifying variation. In other words, regardless of the size of $\hat{\mu}$, the final estimates using the predicted changes in population as IVs should be the same.

First, I discuss the intuition of why $\hat{\mu}$ does not matter in constructing IVs. The key identifying variation in the IVs is the variation in expected commuting time (driven by the spatial location of jobs) and the different changes in the value of time by occupation: $\Delta\hat{v}_{kt}\widetilde{\mathrm{E}}_{t-1}\left(c_{jmk}\right)$. $\hat{\mu}$ is simply a scaling factor.

Let us assume for now that there is only one occupation. The change in local population can thus be written as:

$$\begin{aligned}
\Delta\log\hat{N}_{jmkt} &= \log\left(N_{mk,t-1}\cdot\frac{\exp\left(\log\left(s_{jmk,t-1}\right)-\hat{\mu}\Delta\hat{v}_{kt}\widetilde{\mathrm{E}}_{t-1}\left(c_{jmk}\right)\right)}{\sum_{j'\in J_m}\exp\left(\log\left(s_{j'mk,t-1}\right)-\hat{\mu}\Delta\hat{v}_{kt}\widetilde{\mathrm{E}}_{t-1}\left(c_{j'mk}\right)\right)}\right)-\log\left(N_{jmk,t-1}\right)\\
&= \log N_{mk,t-1}+\log\left(s_{jmk,t-1}\right)-\hat{\mu}\Delta\hat{v}_{kt}\widetilde{\mathrm{E}}_{t-1}\left(c_{jmk}\right)\\
&\quad -\log\left(\sum_{j'\in J_m}\exp\left(\log\left(s_{j'mk,t-1}\right)-\hat{\mu}\Delta\hat{v}_{kt}\widetilde{\mathrm{E}}_{t-1}\left(c_{j'mk}\right)\right)\right)-\log\left(N_{jmk,t-1}\right)\\
&= \log N_{mk,t-1}+\log\left(s_{jmk,t-1}\right)-\hat{\mu}\Delta\hat{v}_{kt}\widetilde{\mathrm{E}}_{t-1}\left(c_{jmk}\right)-\Lambda_{mkt}-\log\left(N_{jmk,t-1}\right)\\
&= \log\left(s_{jmk,t-1}\right)-\hat{\mu}\Delta\hat{v}_{kt}\widetilde{\mathrm{E}}_{t-1}\left(c_{jmk}\right)-\Lambda_{mkt}
\end{aligned}$$

Note that the log change in the local population in occupation k is a linear function of $\hat{\mu}$. This means that the variation in the predicted change in population is preserved regardless of the size of $\hat{\mu}$. Even though Λ_{mkt} is also a function of $\hat{\mu}$, it does not vary by neighborhood. The occupation- and city-level fixed effect will take it out.

In my estimates, however, I calculate the predicted change in population for high- and low-skilled workers, which is a sum of the number of workers across occupations. The algebra is substantially more involved. A simple linear form cannot be derived. Nevertheless, I can show that the IVs' derivatives are largely invariant of $\hat{\mu}$, which means that the identifying variation does not change as I vary the size of $\hat{\mu}$.

I start by writing out the equation of the predicted change in log population of skill type z:

$$\Delta\log\hat{N}^z_{jmt}=\log\left(\sum_k N_{mk,t-1}\cdot\frac{\exp\left(\log\left(s_{jmk,t-1}\right)-\hat{\mu}\Delta\hat{v}_{kt}\widetilde{\mathrm{E}}_{t-1}\left(c_{jmk}\right)\right)}{\sum_{j'\in J_m}\exp\left(\log\left(s_{j'mk,t-1}\right)-\hat{\mu}\Delta\hat{v}_{kt}\widetilde{\mathrm{E}}_{t-1}\left(c_{j'mk}\right)\right)}\right)-\log\left(\sum_k N_{jmk,t-1}\right).$$

I then take the derivative with respect to $\hat{\mu}$:

$$
\begin{aligned}
\frac{\partial \Delta \log \hat{N}^z_{jmt}}{\partial \hat{\mu}} &= \frac{1}{\hat{N}^z_{jmt}} \left(\sum_k N_{mk,t-1} \left(-\Delta \hat{v}_{kt} \widetilde{\mathrm{E}}_{t-1}\left(c_{jmk}\right) s_{jmkt} + s_{jmkt} \sum_{j'} \Delta \hat{v}_{kt} \widetilde{\mathrm{E}}_{t-1}\left(c_{j'mk}\right) s_{j'mkt} \right) \right) \\
&= \frac{1}{\hat{N}^z_{jmt}} \left(\sum_k \hat{N}_{jmkt} \left(-\Delta \hat{v}_{kt} \widetilde{\mathrm{E}}_{t-1}\left(c_{jmk}\right) + \sum_{j'} \Delta \hat{v}_{kt} \widetilde{\mathrm{E}}_{t-1}\left(c_{j'mk}\right) s_{j'mkt} \right) \right) \\
&= \sum_k \underbrace{\hat{\rho}_{jmkt}}_{k\text{'s share in tract } j} \left(-\Delta \hat{v}_{kt} \widetilde{\mathrm{E}}_{t-1}\left(c_{jmk}\right) + \Lambda_{kmt} \right)
\end{aligned}
$$

Through the derivation, one can see that the marginal effect of $\hat{\mu}$ on the change of local population is simply the interaction of the change in the value of time and the expected commute time plus a city-occupation-year level constant Λ_{kmt}, weighted by occupation k's share in tract j, $\hat{\rho}_{jmkt}$. Since $\hat{\mu}$'s impact on $\hat{\rho}_{jmkt}$ and Λ_{kmt} are not first-order, varying the size of $\hat{\mu}$ simply scales the aggregate variation in $-\Delta \hat{v}_{kt} \widetilde{\mathrm{E}}_{t-1}\left(c_{jmk}\right)$.

E Alternative decomposition exercise

Here, I present a set of exercises that try to gauge the overall impact of the changing value of time on gentrification in a framework that does not rely on or at least relies less on the model machinery presented in the paper. I run regressions at city level, exploiting cross-city variations in the changing long-hour premium, changing prevalence of long hours, and central city location choice. I implement a series of exercises in that spirit and put them in Table A9 in the appendix.

The idea of these exercises is that if amenities adjust endogenously due to the changing value of time, we may be able to gauge the size of the full effect of the changing value of time (as opposed to only the direct effect) by looking at overall levels of central city gentrification across cities and how they vary by the cities' aggregate changing value of time.

In column 1, I present the result of regressing the actual change in central city relative log skill ratio on model-predicted change in central city relative log skill ratio across cities. In this regression, I am simple regressing the actual changes in log relative central city skill ratio on model-predicted changes (only due to the change in the value of time) in log relative central city skill ratio:

$$
\Delta \left(\ln \left(\frac{N^H_{C,mt}}{N^L_{C,mt}} \right) - \ln \left(\frac{N^H_{S,mt}}{N^L_{S,mt}} \right) \right) = a_0 + a_1 \Delta \left(\ln \left(\frac{\hat{N}^H_{C,mt}}{\hat{N}^L_{C,mt}} \right) - \ln \left(\frac{\hat{N}^H_{S,mt}}{\hat{N}^L_{S,mt}} \right) \right) + \varepsilon_m
$$

I let m to index city. Result shows that $\hat{a}_1$ is far greater than one, which suggests that the observed change in the relative skill ratios in central cities most likely reflect endogenous amenity changes. Based on the coefficient a_1, I compute the percentage of the observed gentrification that can be attributed to the variation of the model-predicted gentrification:

$$\frac{\hat{a}_1 E\left(\Delta\left(\ln\left(\frac{\hat{N}^H_{C,mt}}{\hat{N}^L_{C,jt}}\right)-\ln\left(\frac{\hat{N}^H_{S,mt}}{\hat{N}^L_{S,mt}}\right)\right)\right)}{E\left(\Delta\left(\ln\left(\frac{N^H_{C,mt}}{N^L_{C,mt}}\right)-\ln\left(\frac{N^H_{S,mt}}{N^L_{S,mt}}\right)\right)\right)}$$

Column 1 result suggests that 69.84% of the observed gentrification can be attributed to the variation due to the changing value of time, suggesting that the full effect of the changing value of time is substantially larger than the direct effect, even larger than the results drawn in Table 8.

Of course, column 1 result may subject to some concerns since it is not entirely model-free. Since the model prediction contains information on city's spatial distribution of jobs, which may correlate with other unobserved factors, $\hat{a}_1$ may be biased. I instrument the model-predicted change in relative skill ratio by the change average long-hour premium for each city (column 2), and separately by college attainment (column 3). The results are slightly smaller with the IVs compared to column 1, but still larger than half of the observed gentrification.

Next, I conduct a completely model-free exercise. I use the change in the percentage of working long hours in each city as the regressor. I use the city's average change in LHP as the IV. In other words, the regression can be written as:

$$\begin{aligned}
\Delta\left(\ln\left(\frac{N^H_{C,mt}}{N^L_{C,mt}}\right)-\ln\left(\frac{N^H_{S,mt}}{N^L_{S,mt}}\right)\right) &= a_0 + a_1\Delta LH_{mt} + \varepsilon_m \\
\text{where, } \Delta LH_m &= b_0 + b_1 \Delta v_{mt} + \omega_m.
\end{aligned}$$

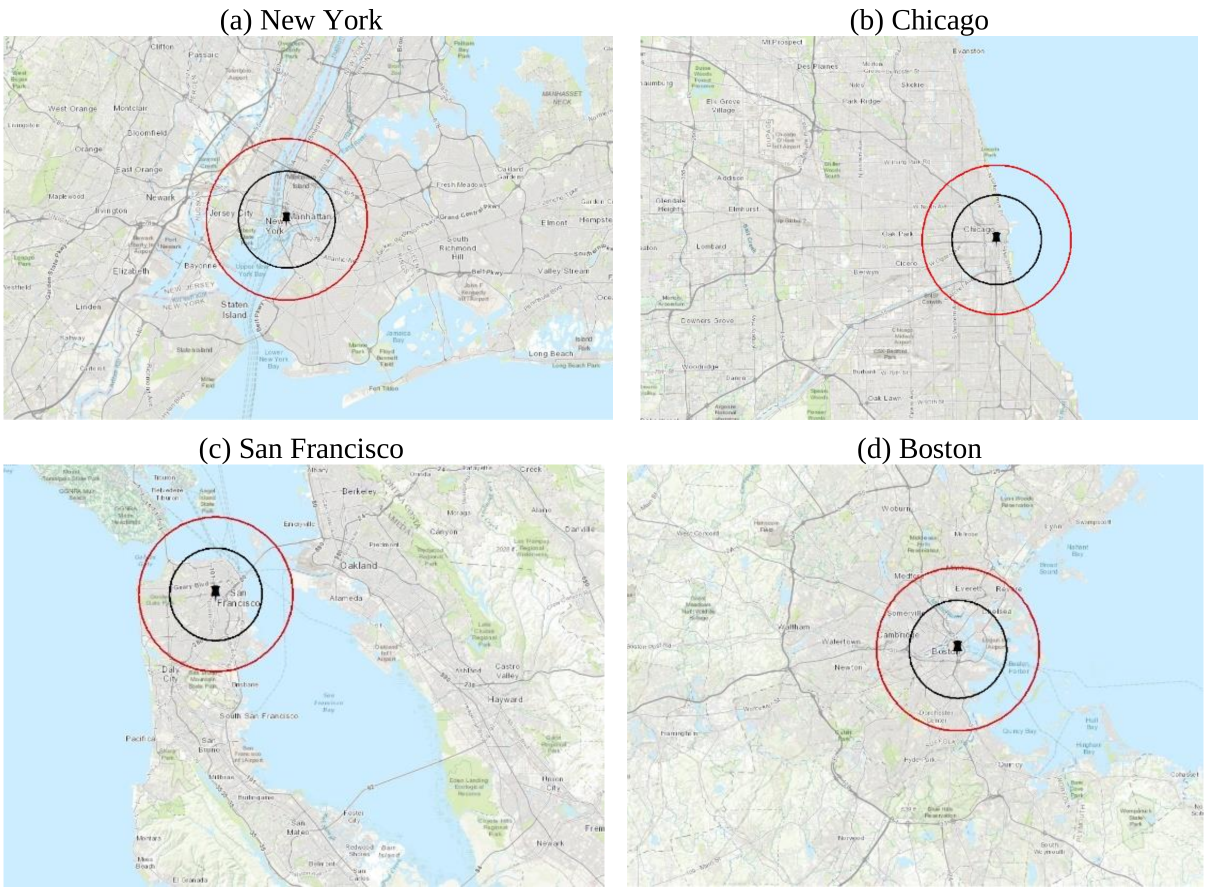
ΔLH_{mt} is the change in average percentage of workers working long hours in city m. Δv_{mt} is the change in the average LHP in city m. With the estimated $\hat{a}_1$, I then compute:

$$\frac{\hat{a}_1 E\left(\Delta LH_{mt}\right)}{E\left(\Delta\left(\ln\left(\frac{N^H_{C,mt}}{N^L_{C,mt}}\right)-\ln\left(\frac{N^H_{S,mt}}{N^L_{S,mt}}\right)\right)\right)}.$$

I present the result of the exercise in column 4. According to the exercise, roughly 44% of the observed variation in central city gentrification can be explained by the variation in the changing prevalence of working long hours. This again suggest that a considerably degree of gentrification can be attributed to the changing hours patterns, but the full effect is far larger than the direct effect, suggesting the endogenous amenity channel is very important.

Appendix: Figures and tables

Figure A1: Map of downtowns and the 3-mile and 5-mile ring in selected MSAs

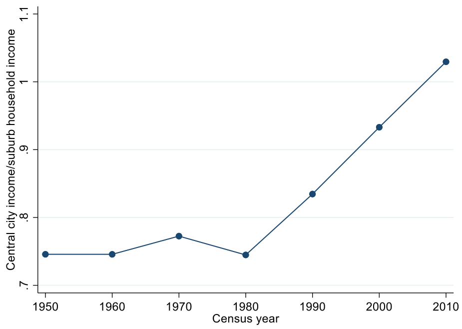


Notes: The longitudes and latitudes of the downtown pins are provided by Holian and Kahn (2015). The smaller circles indicate the 3-mile (Euclidean distance) rings around the indicated downtown pins, and the larger circles indicate the 5-mile rings around the indicated downtown pins.

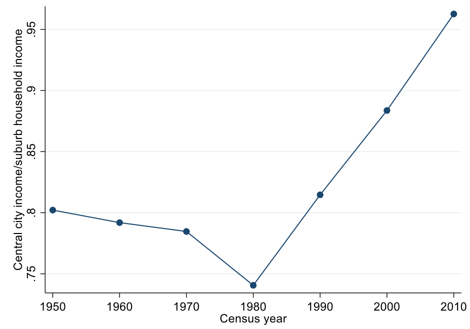
Figure A2: Income ratio between central city and suburban neighborhoods by MSA population ranking

Largest 10 MSAs

a) 3-mile radius

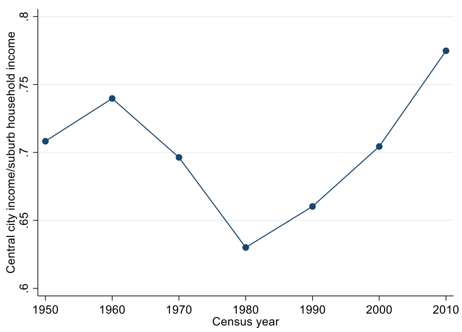


b) 5-mile radius



11th - 25th largest MSAs

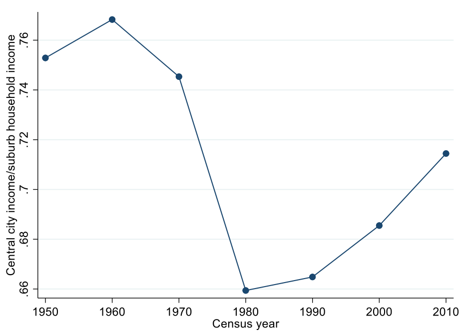
c) 3-mile radius



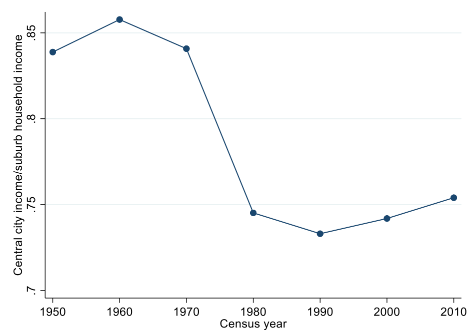
d) 5-mile radius

26th – 50th largest MSAs

e) 3-mile radius

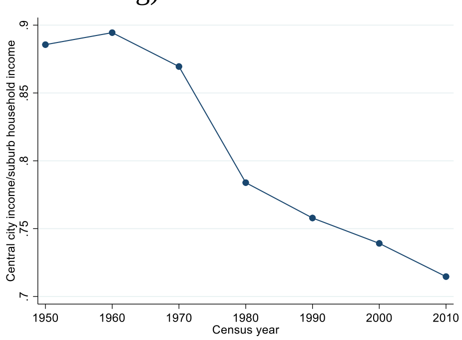


f) 5-mile radius

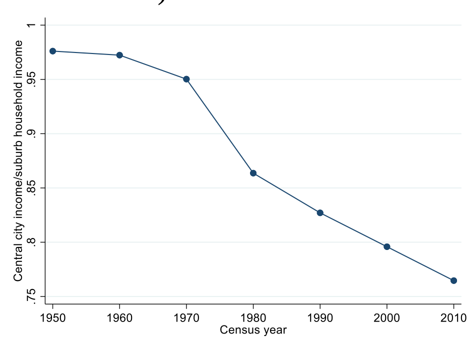


51th + MSAs

g) 3-mile radius

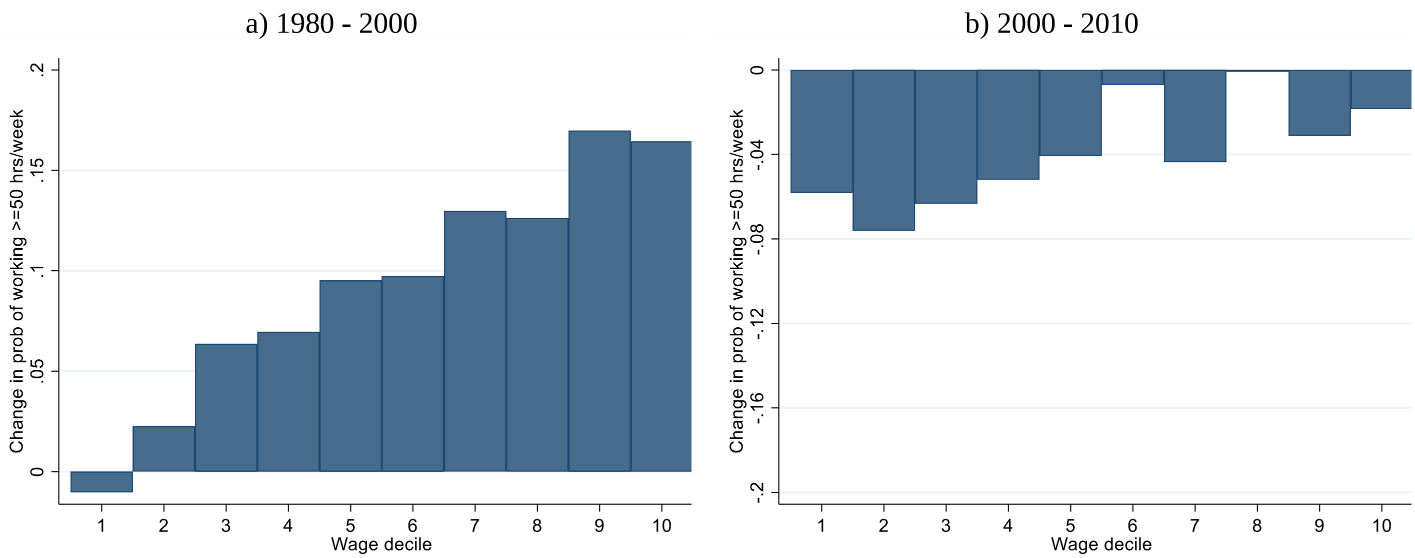


h) 5-mile radius



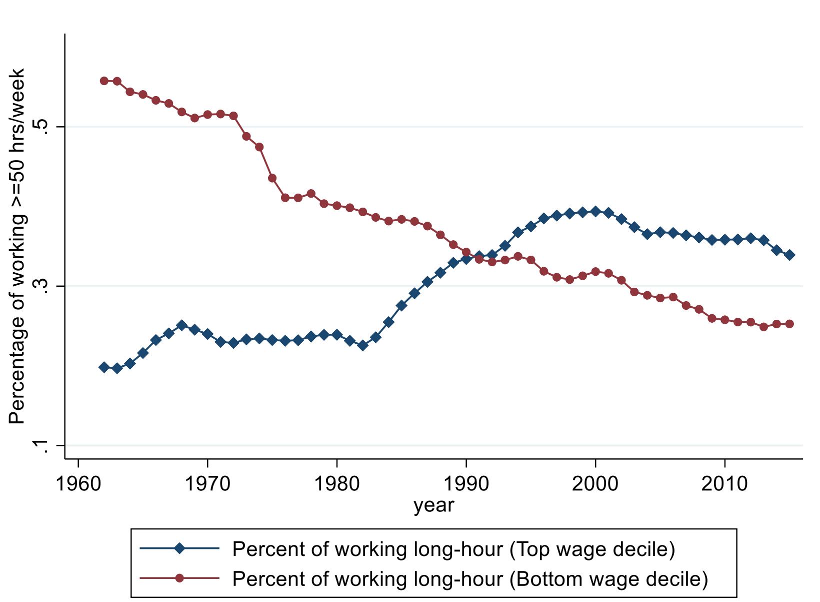
Notes: Central cities in these figures are census tracts that are located within 5 miles of the downtown in the respective MSAs defined in Holian and Kahn (2015). The values plotted are the mean income ratios between the central city census tracts and suburban census tracts with sample of MSAs of different population rankings.

Figure A3: Changing probability of working long hours by wage decile (>=50 hours per week)



Notes: Data come from IPUMS census data in 1980, 2000, 2010 (2007-2011 ACS). To compute the probability of working at least 50 hours per week, the sample I use is workers that are between 25 and 65 of age, males, and working at least 30 hours per week. I include only male in the sample to ensure that the changing female labor force participation does not distortion the statistics. In a), I compute the change in probability of working long hours (>=50 hours per week) from 1980 to 2000. In b), I compute the change in probability of working long hours (>=50 hours per week) from 2000 to 2010.

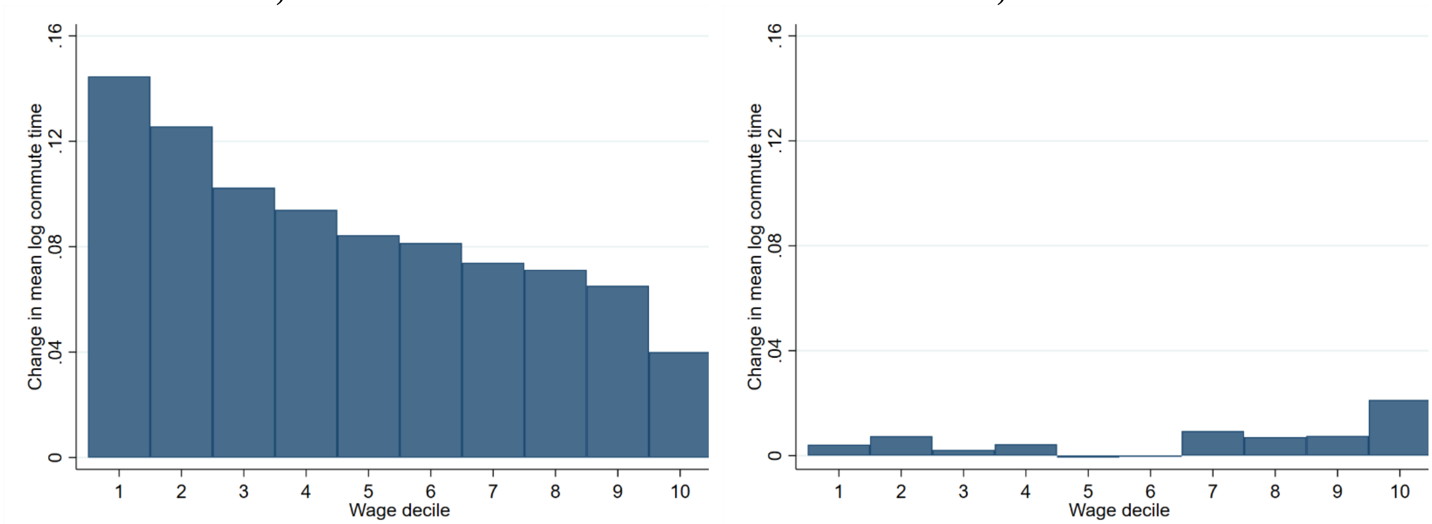
Figure A4: The evolution of long-hour working



Notes: I plot the probability of working at least 50 hours a week using the CPS ASEC data from 1968 to 2016. The sample includes workers that are male, between age 25 and 65 and work at least 30 hours per week. I plot the probability of working long hours for workers in the top wage decile and the bottom wage decile over time. To smooth the plotted curve, each dot represents a three-year moving average.

Figure A5: Growth of commute time by wage decile

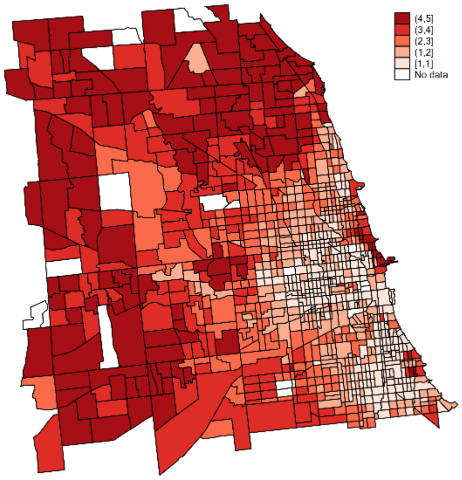
a) 1980 - 2000 b) 2000 - 2010



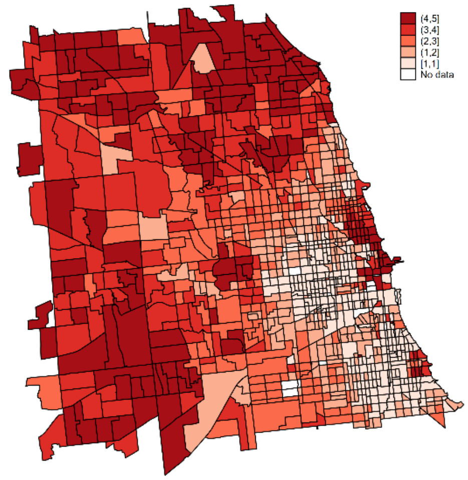
Notes: Data come from IPUMS census data in 1980, 2000, 2010 (2007-2011 ACS). I compute the change in log commute time reported in the Census/ACS data. The sample includes workers that are between 25 and 65 of age, males, working at least 30 hours per week and living in the most populous 25 MSAs in the US. In a), I plot the change in log commute time between year 2000 and 1980. In b), I plot the change in log commute time between year 2010 and 2000.

Figure A6: Income quintile by neighborhood within Chicago MSA (1980 – 2010)

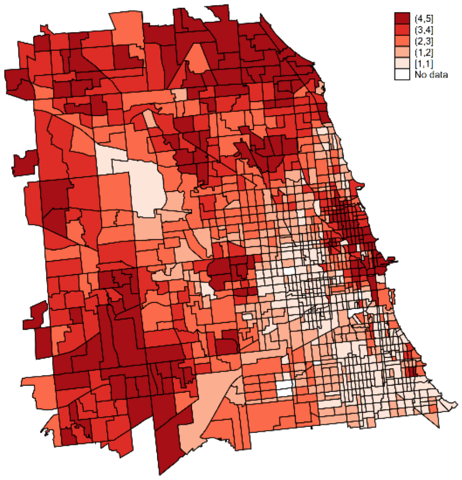
(a). 1980



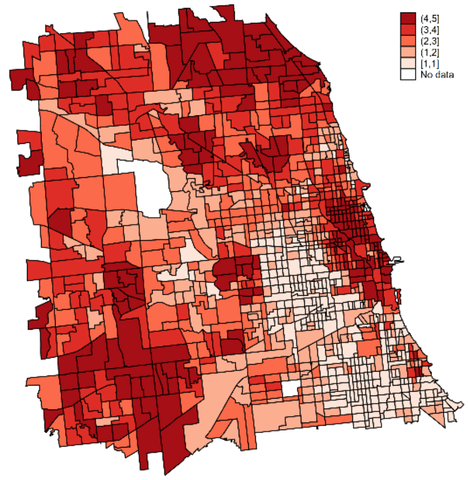
(b). 1990



(c). 2000

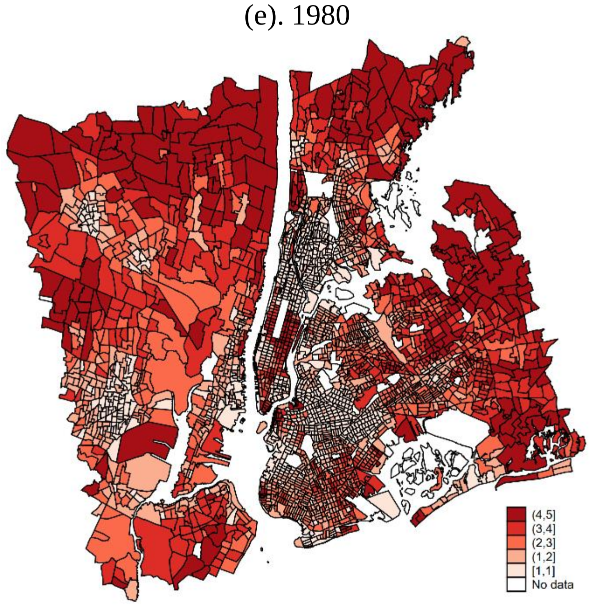


(d). 2010

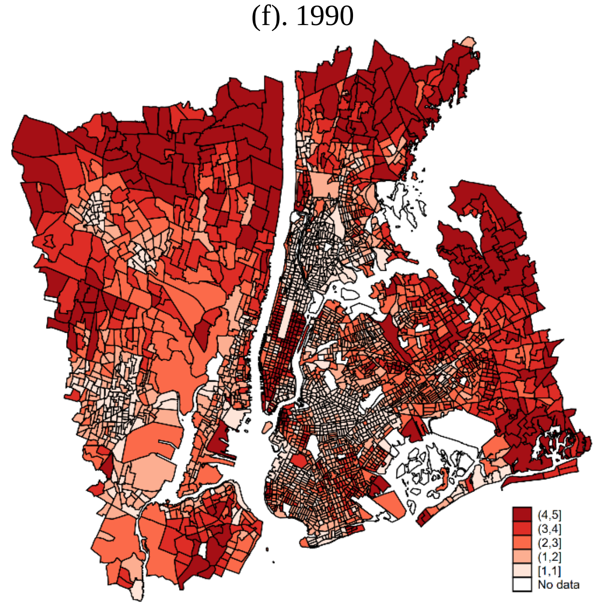


Income quintile by neighborhood within New York MSA (1980 – 2010)

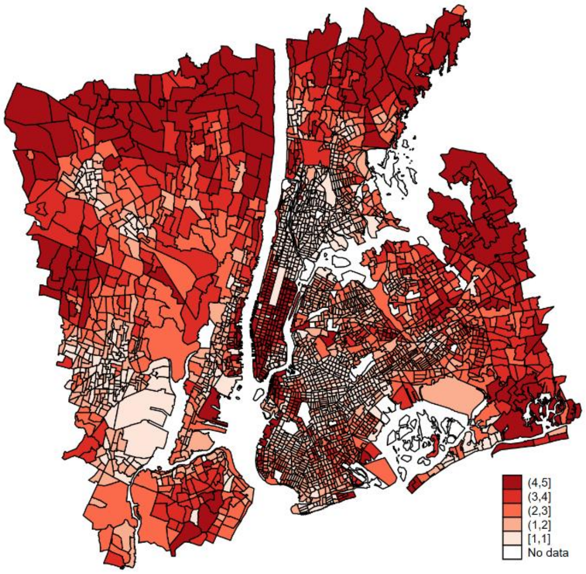
(e). 1980



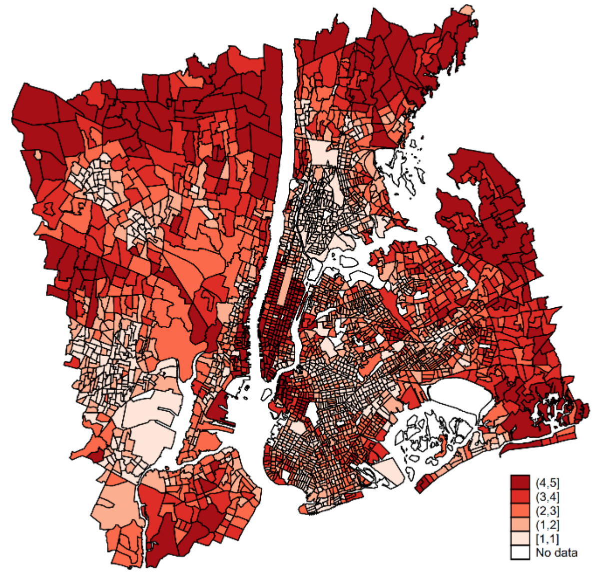
(f). 1990



(g). 2000



(h). 2010

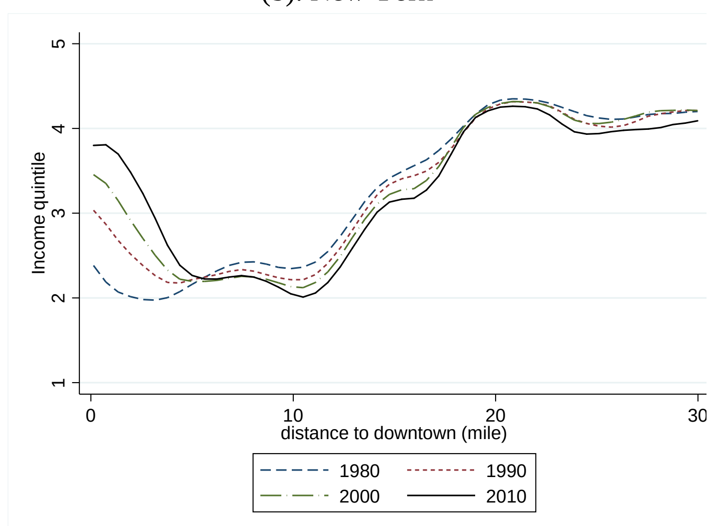


Notes: The plotted values are quintile ranking of census tract level income within the Chicago MSA and New York MSA respectively, from year 1990 to year 2010 using the Census summary statistics (NHGIS). The light color represents lowly ranked census tracts, and the darker red color represents more highly ranked census tracts in each contemporaneous year.

Figure A7: Income quintile by distance to downtown.

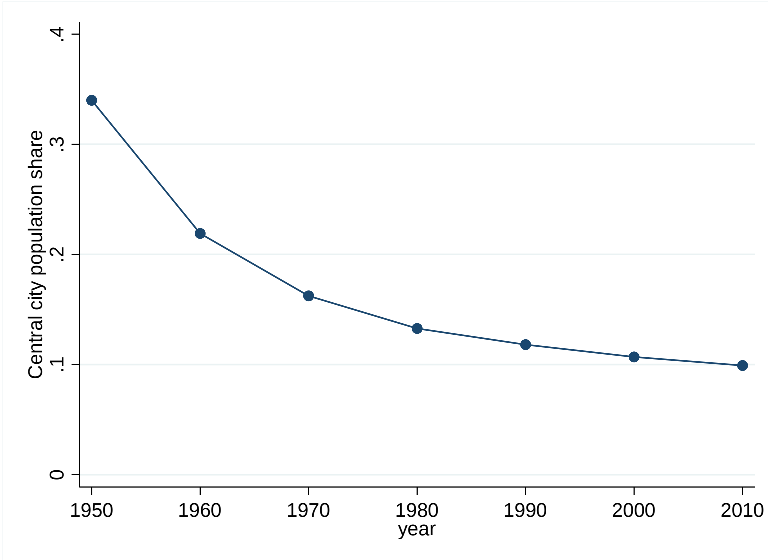
(a). Chicago

(b). New York



Notes: The plotted values are quintile ranking of census tract level income within the Chicago MSA and New York MSA respectively. I plot the census tract income ranking from year 1980 to year 2010 against the distance (in mile) to downtown. The plot is the kernel-weighted local polynomial smoothing curve, and Epanechnikov kernel function.

Figure A8: Central city population percentage among the largest 25 MSAs.



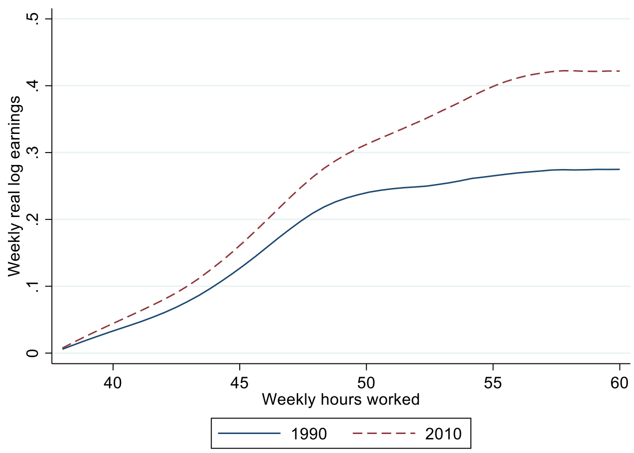
Notes: Central cities in this graph are defined as census tracts that are located within 5 miles of the downtown pin on Google in the respective MSAs. The value plotted in the graph are the population ratio between the population in the census tracts located in the central cities and the total population in the top 25 MSAs (defined by population ranking in 1990). The source of the data is Census and ACS provided by NHGIS.

Figure A9: Work and residential location in early 1990s

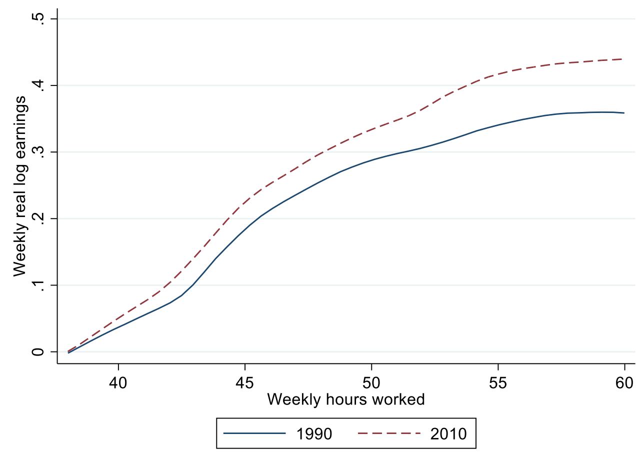
Notes: Residential location data come from both IPUMS and NHGIS Census data. Details are described in the data section. The employment data come from ZCBP at zip code level. Central cities are defined as census tracts and zip codes with centroids within a 5-miles radius of the downtown pin. I use the sample from the most populous 25 MSAs to produce these graphs. The redline is the 45-degree line.

Figure A10: Residual log weekly earnings against weekly hours worked

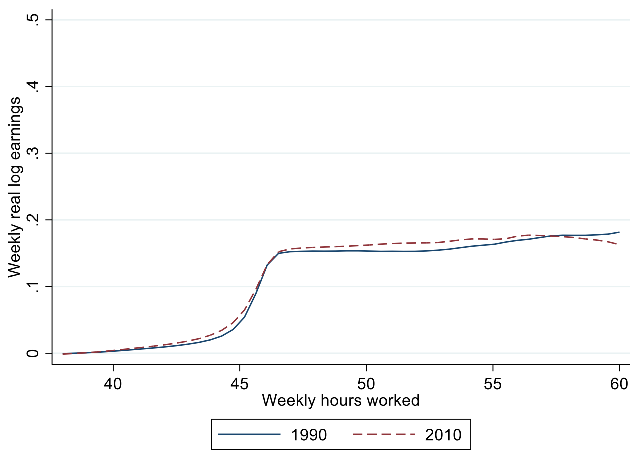
a) Financial specialists



b) Lawyers



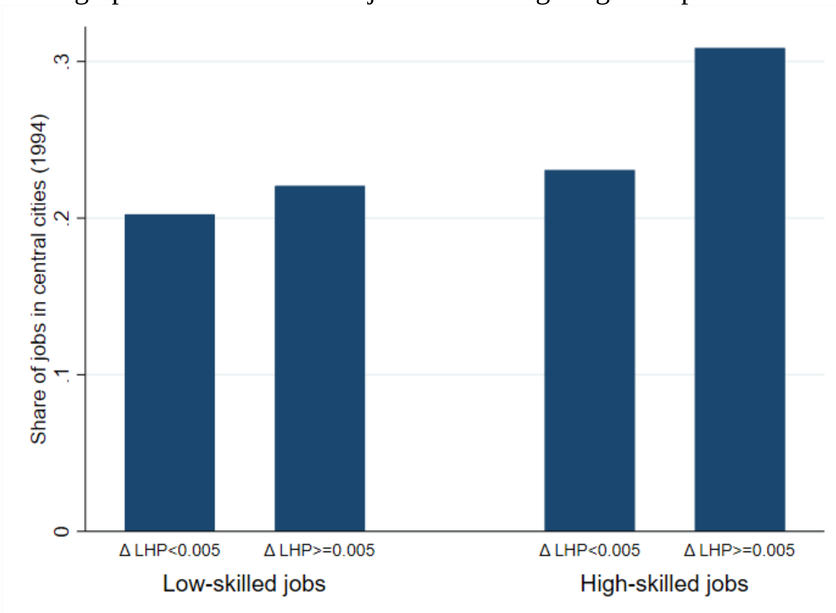
c) Secretaries and administrative assistants



d) Teachers

Notes: All samples come from Census data in IPUMS. ACS 2007-2011 is used for year 2010. The variables used in the plots are residual values after being regressed on individual level control variables (age, sex, race, education, industry code). The residual log earnings are normalized by constants such that the values in 1990 and 2010 start out from zero to help visual contrast. I categorize several occupations into financial specialists. Financial specialists (a) include financial managers (occ2010: 120), accountants and auditors (occ2010: 800), and securities, commodities, and financial services sales agents (occ2010: 4820). Teachers include elementary school teachers (occ2010: 2310) and secondary school teachers (occ2010: 2320). The plot is the kernel-weighted local polynomial smoothing curve, with bandwidth equals 2.5, and Epanechnikov kernel function.

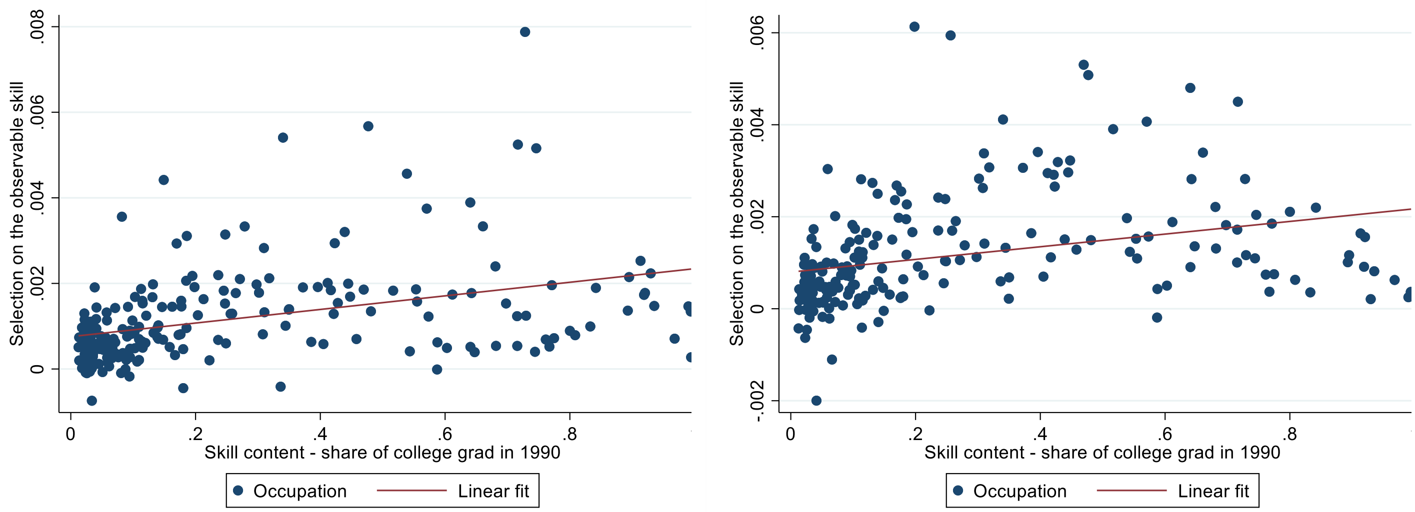
Figure A11: Geographic concentration of jobs with rising long-hour premium in central cities



Notes: The figure shows the shares of jobs located in central cities based on data from the 1994 Zip Code Business Patterns. Central city neighborhoods are defined as census tracts within 5 miles of downtowns. The sample is the most populous 25 MSAs. I divide occupations into low- and high-skilled jobs. Occupations with greater than 40% of the workers having college degree are categorized as high-skilled. Within each skill category, I further categorize occupations based on the change in long-hour premium between 1990 and 2010. 37 occupations included as high-skilled and ΔLHP>=0.005; 47 occupations included as high-skilled and ΔLHP<0.005; 123 occupations included as low-skilled and ΔLHP>=0.005; 134 occupations included as low-skilled and ΔLHP<0.005. The total number of occupations is greater than those included in the model analysis, because some occupations shown in this graph cannot be merged with residential location data used for model analysis.

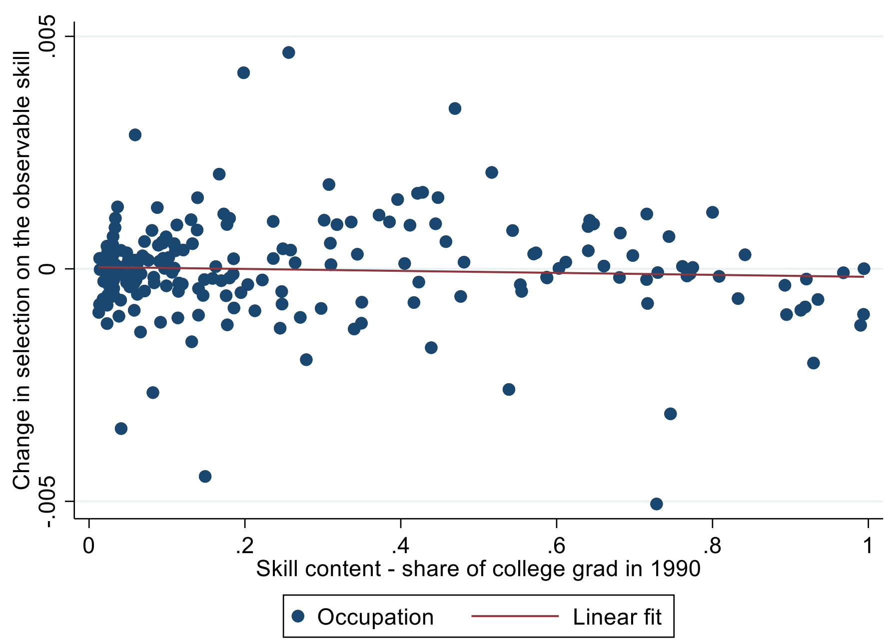
Figure A12: Degree of selection for long-hour premium estimates on observable skills in 1990 and 2010

a) Degree of selection in 1990 b) Degree of selection in 2010



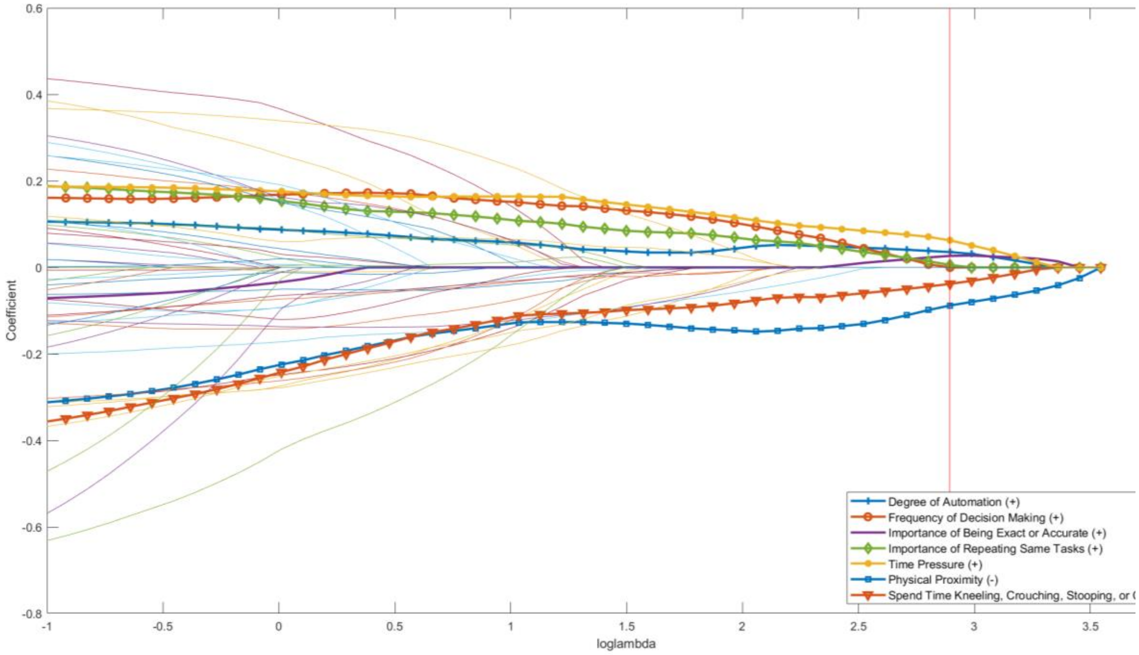
Notes: The y-axis is the difference between the estimates of long-hour premium without controlling for education levels and the estimates controlling for education levels. The difference between the two estimates indicates the degree of selection on the observable skill levels. X-axis is the skill content of each occupation, measured as the share of college graduates in 1990 Census.

Figure A13: Change in the degree of selection for long-hour premium estimates on observable skills



Notes: The y-axis is the difference between the change in long-hour premium estimated without education control and with education control. X-axis is the skill content of each occupation, measured as the share of college graduates in 1990 Census.

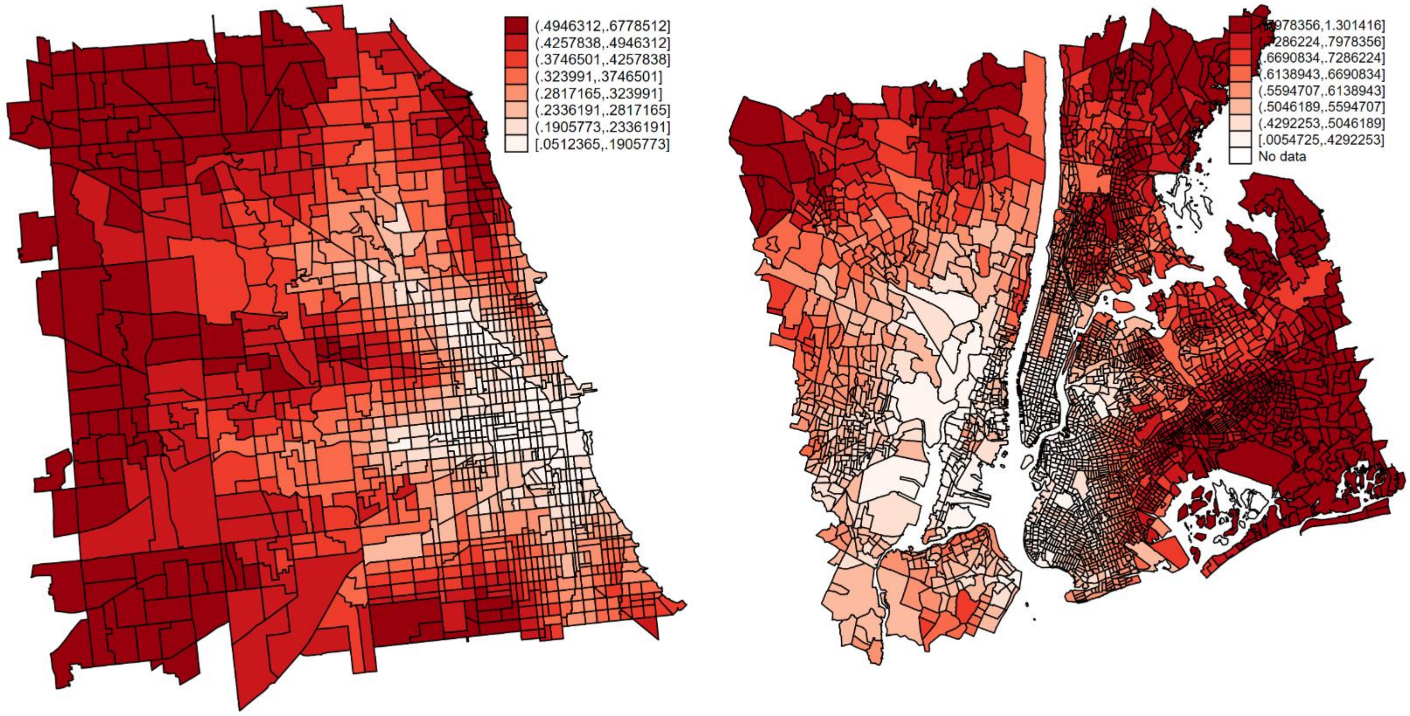
Figure A14: Lasso trace plot of the O*NET characteristics at predicting change in long-hour premium



Notes: I plot the coefficients on each of the 57 O*NET occupation characteristics for different levels of lambda (regularization penalty). The outcome variable is the change in long-hour premium. The red vertical line marks the lambda selected by 10-fold cross-validation. The characteristics that are non-zero at the red line are the non-redundant characteristics.

Figure A15: Imputed 1995 rush-hour driving time

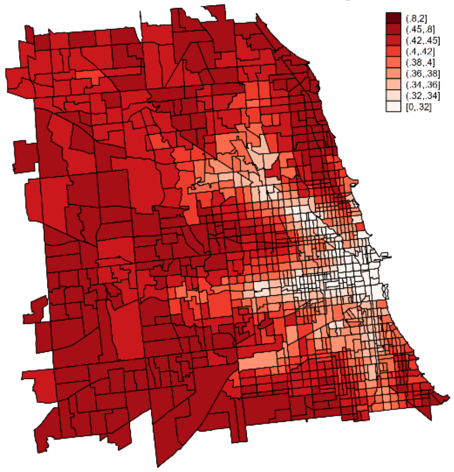
(a) Chicago (zip code: 60605) (b) New York (zip code: 10005)



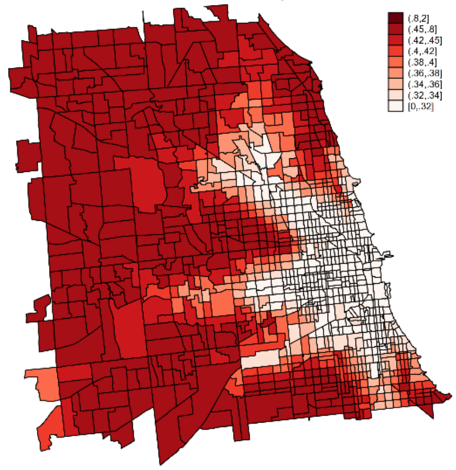
Notes: The above maps plot travel time from each census tract to the downtown of the MSA. I designate the destination for Chicago MSA as zip code 60605 (downtown Chicago) and destination for New York MSA as zip code 10005 (downtown Manhattan). The light color represents census tract with short travel time to the center of the city and dark red color represents long travel time. The maps are shown for the purpose of demonstration. To conduct the model estimation, I impute driving time to every zip code from every census tract in the non-rural counties of the US.

Figure A16: Expected commute time for selected occupations in Chicago MSA.

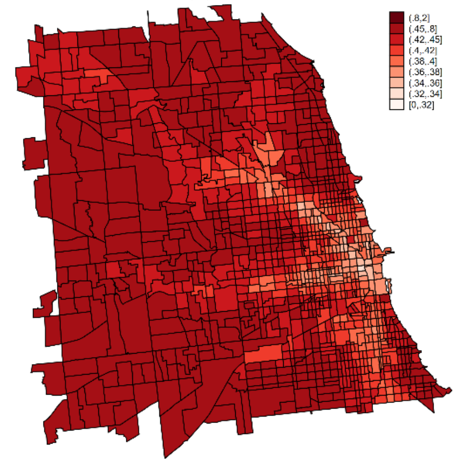
(a) Financial managers



(b) Lawyers

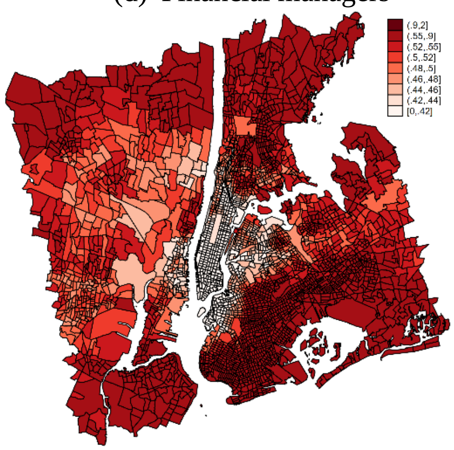


(c) Cashiers

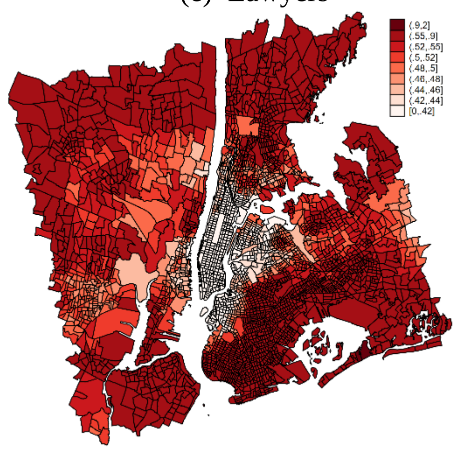


Expected commute time for selected occupations in New York MSA.

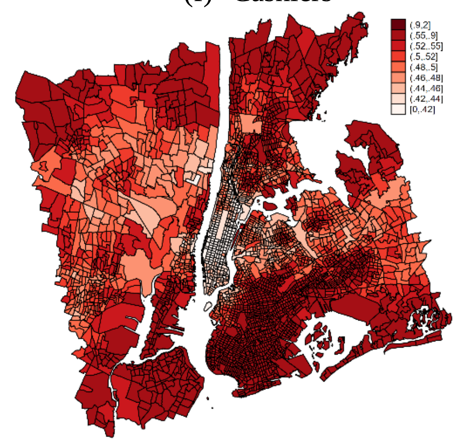
(d) Financial managers



(e) Lawyers



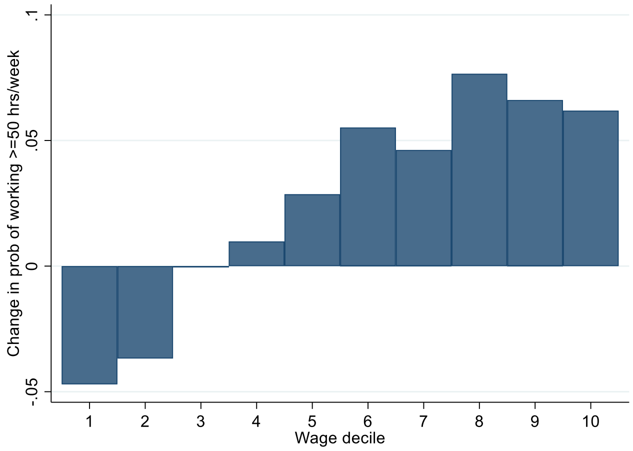
(f) Cashiers



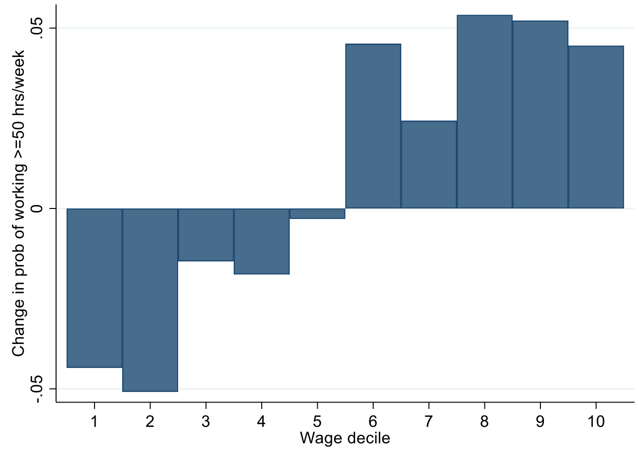
Notes: The above maps are selected demonstrations of the expected commute time computed using employment allocation data (ZCBP data) and travel time matrix. The geographic unit displayed in the graphs is census tract. The color ranges from white to red. The light color represents short commute time, and red color represents long commute time. The color scale is consistent within respective MSA. The purpose of the maps is to show that the expected commute time by census tracts is quite different across different occupations, due to the differential distribution of job locations.

Figure A17: Changing incidence of working long hours by wage decile (1980-2010)

a) Central cities



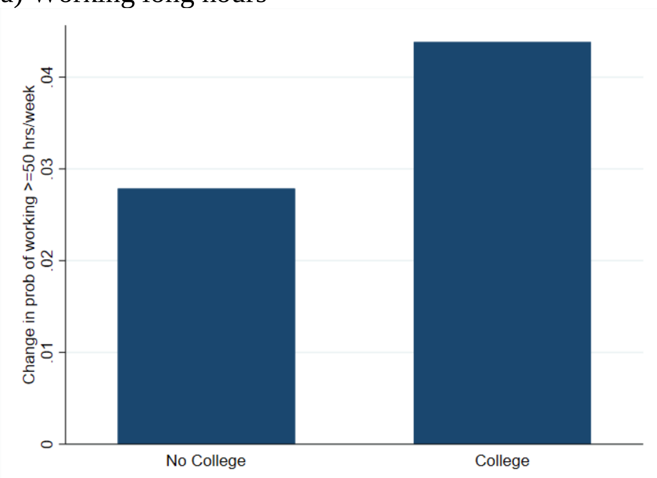
b) Suburbs



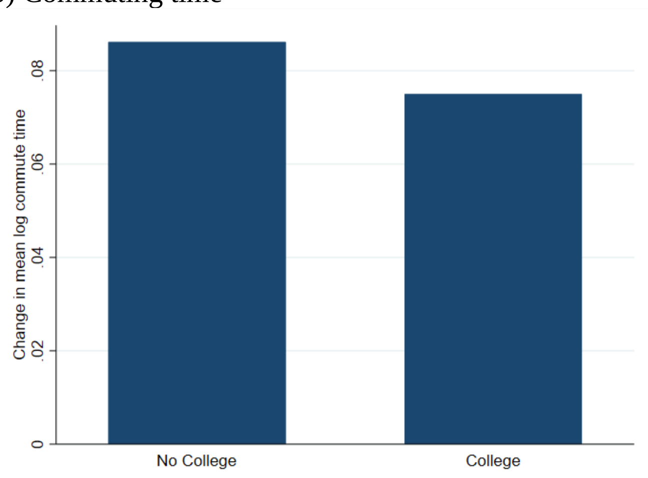
Notes: The data come from IPUMS census data in 1980 and 2010 (2007-2011 ACS). I compute the change in the probability of working at least 50 hours per week by wage decile for workers living in central cities vs. the suburbs. The sample I use includes workers that are between 25 and 65 of age, males, and working at least 30 hours per week. I include only male in the sample to ensure that the changing female labor force participation does not distort the statistics. In a), I include only the sample of workers who live in central city neighborhoods (5 miles within downtowns). In b), I include only the sample of workers who live outside of central city neighborhoods. Wage deciles are assigned by using a national wage ranking.

Figure A18: Changing incidence of working long hours and changing commuting time by college attainment

a) Working long hours



b) Commuting time



Notes: The data come from IPUMS census data in 1980 and 2010 (2007-2011 ACS). In a), I compute the change in probability of working at least 50 hours per week by education attainment. The sample I use includes workers that are between 25 and 65 of age, males, and working at least 30 hours per week. I include only male in the sample to ensure that the changing female labor force participation does not distort the statistics. In b), I compute the change in log commute time reported in the Census/ACS data. The sample includes workers that are between 25 and 65 of age, males, working at least 30 hours per week and living in the most populous 25 MSAs in the US.

Table A1: Estimate of long-hour premium with or without controls for education

Occupation name	code	Without educ. control LRP - 1990	LRP - 2010	Δ in LRP	With educ. control LRP - 1990	LRP - 2010	Δ in LRP
Managers in Marketing, Advertising, and Public Relations	30	0.0173	0.0230	0.0057	0.0161	0.0214	0.0054
Financial Managers	120	0.0217	0.0287	0.0069	0.018	0.0246	0.0066
Accountants and Auditors	800	0.0231	0.0293	0.0062	0.0198	0.0259	0.0062
Computer Scientists and Systems Analysts/Network systems Analysts/Web Developers	1000	0.0108	0.0158	0.0051	0.0102	0.0145	0.0043
Lawyers, and judges, magistrates, and other judicial workers	2100	0.0213	0.0247	0.0034	0.0198	0.0245	0.0046
Secondary School Teachers	2320	0.0065	0.0060	-0.0006	0.0048	0.0051	0.0003
Securities, Commodities, and Financial Services Sales Agents	4820	0.0196	0.0403	0.0206	0.0157	0.0355	0.0197
Secretaries and Administrative Assistants	5700	0.0149	0.0154	0.0005	0.0146	0.0147	0.0000

Notes: The measurements are computed with microdata from Census IPUMS data. To compute the long-hour premium, I restrict the sample to workers between age of 25 and 65, and work at least 40 hours per week but does not work more than 60 hours per week. The results on the left are estimates without education controls, whereas the results on the right are estimates with education controls.

Table A2: Correlations between model predictions and data by decade

Dep var: Actual change in relative log skill ratio	(1) 1990-2010	(2) **1990-2000**	(3) **1990-2000**	(4) **2000-2010**	(5) **2000-2010**
Predicted change in relative log skill ratio (only by the change in the value of time) over 1990-2010	11.1 (1.699)	6.559 (0.826)		4.541 (1.158)	
Predicted change in relative log skill ratio (only by the change in the value of time) over **1990-2000**			10.228 (1.420)		
Predicted change in relative log skill ratio (only by the change in the value of time) over **2000-2010**					5.603 (1.053)
R-squared	0.303	0.314	0.183	0.161	0.143
Observations	221	221	221	221	221

Notes: The table shows the coefficient of regressing the actual change in relative log skill ratio on the change in relative log skill ratio predicted by only the change in the value of time, holding all else equal. Central city is defined as census tracts within 3 miles of downtown. For column 1, the dependent variable is the actual change between year 1990 and 2010. For columns 2 and 3, the dependent variables are the actual changes during 1990-2000. For columns 4 and 5, the dependent variables are the actual changes during 2000-2010.

Table A3: Exogeneity of increasing value of time and incidence of working long hours

Dep var	Δlog(long hours$_{downtown}$)	Δlog(long hours$_{suburbs}$)	Δlog(long hours$_{downtown}$)
	(1)	(2)	(3)
ΔLHP	14.386	13.395	
	(5.200)	(3.771)	
Δlog(long hours$_{suburbs}$)			1.0402
			(0.0822)
R^2	0.0525	0.0834	0.5887
Observation	213	214	213

Notes: Each observation in the regressions represents an occupation. For column 1, the dependent variable is the change in log probability of working long hours with sample of male workers who work at least 30 hours a week and aged 25 to 65 and live in central city neighborhoods (5 miles within downtown). For column 2, the dependent variable is similarly constructed as in column 1, but with sample who live outside of central city neighborhoods. The regressor in both column 1 and 2 is the change in long-hour premium by occupation. For column 1 I regress the dependent variable in column 1 on the dependent variable in column 2.

Table A4: First-stage between actual change in skill ratio and predicted change in skill ratio

	Dep variable: Actual Δ log skill ratio	
	(1)	(2)
Predicted Δ in log skill ratio	0.667	-
	(0.0685)	
Predicted Δ change in high-skilled workers	-	2.0155
		(0.136)
Predicted Δ change in low-skilled workers	-	-0.257
		(0.0648)
MSA fixed-effects	Yes	Yes
Observations	43,246	43,246
F-statistics	94.95	112.61

Notes: Results shown above are OLS regressions. Each observation is a census tract. The first-difference is between 1990 and 2010. I use 2007-2011 ACS for year 2010. Column 1 reports regression result when predicted change in log skill ratio is included as the regressor. Column 1 reports regression result when change in high- and low-skilled workers are included separately as the regressors. The predicted change in log skill ratio is generated by changing only the long-hour premium in the model, assuming $\hat{\mu} = 8.94$, which I obtain by estimating the location demand with only the term of commuting cost and no heterogeneity by skill. The model estimates do not respond to the value of $\hat{\mu}$. Standard errors are clustered at census tract level. MSA fixed effects are used for all regressions.

Table A5: Reduced-form results from regressing the change in location demand on exogenous variables

$\Delta v \times \mathrm{E}(c)$	High-skilled	-11.963
		(0.861)
	Low-skilled	-3.465
		(0.586)
$\Delta \ln(N^H)$	High-skilled	1.894
		(0.967)
	Low-skilled	-7.834
		(0.964)
$\Delta \ln(N^L)$	High-skilled	0.996
		(1.502)
	Low-skilled	8.226
		(1.238)
$\Delta \ln(N^H) \times den$	High-skilled	5.276
		(0.481)
	Low-skilled	4.126
		(0.406)
$\Delta \ln(N^L) \times den$	High-skilled	2.882
		(1.031)
	Low-skilled	0.841
		(0.655)
$\Delta \ln(N^H + N^L)*den$	High-skilled	-9.101
		(0.873)
	Low-skilled	-5.117
		(0.621)
den	High-skilled	-0.00517
		(0.00442)
	Low-skilled	-0.0144
		(0.00339)
Observations		8,755,373

Notes: Results are OLS regressions, using occupation/census tract cell data from 1990 to 2010. Number of cells used is 8,755,373. This is the reduced-form result of the main model estimates. Instead of reporting the coefficients on the endogenous variable, I include the coefficients on the IVs in this regression table. The number of workers in each occupation/MSA in 1990 is used as analytical weight. I control for total expected commute (using expected commute time to jobs unrelated to workers' occupations) and the change in expected commute time, and I allow the coefficients on total expected commute and the change in expected commute time to vary by occupation. Standard errors are clustered at census tract level.

Table A6: Gentrification decomposition – allowing for MSA population adjustments

		Largest 25 MSAs			Largest 50 MSAs		
Dist. To downtown		Δ Value of time	Δ Value of time + endogenous Δ amenity + Δ rent	Δ Value of time + endogenous Δ amenity (no Δ rent)	Δ Value of time	Δ Value of time + endogenous Δ amenity + Δ rent	Δ Value of time + endogenous Δ amenity (no Δ rent)
3 miles	Actual	0.305	0.305	0.305	0.269	0.269	0.269
	Model-predicted	0.0670	0.132	0.140	0.0609	0.120	0.127
	%	21.97%	43.28%	45.90%	22.64%	44.61%	47.21%
5 miles	Actual	0.235	0.235	0.235	0.201	0.201	0.201
	Model-predicted	0.0676	0.128	0.140	0.0622	0.118	0.128
	%	28.76%	54.47%	59.57%	30.95%	58.71%	63.68%

Notes: The results shown in this table show the comparison between actual changes in relative skill ratio and model-predicted changes in relative skill ratio. For the model-predicted changes, I allow the MSA's population by skill to adjust according to the observed change. Relative skill ratio is defined as ratio between skill ratio (residents in high-skilled occupations/residents in low-skilled occupations) in central cities and skill ratio in the suburbs. I use varying definition of central city, and sample from largest 25 MSAs and largest 50 MSAs. Actual changes in relative skill ratio are computed using observed spatial data by occupation. The values shown are the mean change in skill ratio weighted by MSAs' population.

Table A7: Gentrification decomposition – Skilled occupation = college share>30%

		Largest 25 MSAs			Largest 50 MSAs		
Dist. To downtown		Δ Value of time	Δ Value of time + endogenous Δ amenity + Δ rent	Δ Value of time + endogenous Δ amenity (no Δ rent)	Δ Value of time	Δ Value of time + endogenous Δ amenity + Δ rent	Δ Value of time + endogenous Δ amenity (no Δ rent)
3 miles	Actual	0.332	0.332	0.332	0.294	0.294	0.294
	Model-predicted	0.0142	0.1026	0.1113	0.0124	0.0896	0.0973
	%	4.28%	30.90%	33.52%	4.22%	30.48%	33.10%
5 miles	Actual	0.252	0.252	0.252	0.217	0.217	0.217
	Model-predicted	0.0125	0.0865	0.0998	0.0110	0.0764	0.0877
	%	4.96%	34.33%	39.60%	5.07%	35.21%	40.41%

Notes: This table is a reproduction of Table 8 in the main manuscript. Here, I categorize occupations as high-skilled if the occupations' initial share of college graduates in 1990 is greater than 30%. I use the parameters estimated in Table 7 column 7 for this exercise.

Table A8: Gentrification decomposition – Skilled occupation = college share>50%

Dist. To downtown		Largest 25 MSAs: Δ Value of time	Largest 25 MSAs: Δ Value of time + endogenous Δ amenity + Δ rent	Largest 25 MSAs: Δ Value of time + endogenous Δ amenity (no Δ rent)	Largest 50 MSAs: Δ Value of time	Largest 50 MSAs: Δ Value of time + endogenous Δ amenity + Δ rent	Largest 50 MSAs: Δ Value of time + endogenous Δ amenity (no Δ rent)
	Actual	0.246	0.246	0.246	0.215	0.215	0.215
3 miles	Model-predicted	0.0395	0.140	0.148	0.0349	0.123	0.131
	%	16.06%	56.91%	60.16%	16.23%	57.21%	60.93%
	Actual	0.191	0.191	0.191	0.160	0.160	0.160
5 miles	Model-predicted	0.0348	0.120	0.134	0.0310	0.107	0.119
	%	18.22%	62.83%	70.16%	19.38%	66.88%	74.38%

Notes: This table is a reproduction of Table 8 in the main manuscript. Here, I categorize occupations as high-skilled if the occupations' initial share of college graduates in 1990 is greater than 50%. I use the parameters estimated in Table 7 column 8 for this exercise.

Table A9: MSA-level gentrification regressions

Dep var: Observed change in relative log skill ratio over 1990-2010	(1)	(2)	(3)	(4)
Predicted change in relative log skill ratio (only by the change in the value of time) over 1990-2010	11.100 (1.137)	10.297 (2.873)	8.484 (2.912)	
Δ pct in long hours (1990-2010)				9.020 (1.676)
OLS or IV	OLS	IV	IV	IV
IV	N/A	Δ pct in long hours	Δ pct in long hours by skill	Δ mean LHP
Observations	221	221	221	221
% of observed change predicted by the independent variable	69.84%	64.79%	53.38%	43.53%

Notes: Column 1 shows the coefficient of regressing the actual change in relative log skill ratio on the change in relative log skill ratio predicted by only the change in the value of time, holding all else equal. For columns 2 and 3, I instrument the independent variable with the change in the average percentage of working long hours by city (MSA). In column 4, I use the change in the average percentage of working long hours as independent variable, and I instrument it with the change in mean long-hour premium.

Table A10: The Long-hour premium and its change by occupation

Occupation	LHP 1990	LHP 2010	ΔLHP 1990-2010	High-skill indicator
Managers in Marketing, Advertising, and Public Relations	0.0161	0.0214	0.0053	1
Financial Managers	0.0180	0.0246	0.0066	1
Human Resources Managers	0.0145	0.0185	0.0039	1
Purchasing Managers	0.0149	0.0182	0.0033	1
Education Administrators	0.0134	0.0125	-0.0009	1
Medical and Health Services Managers	0.0181	0.0181	0.0001	1
Managers, nec (including Postmasters)	0.0157	0.0154	-0.0003	1
Compliance Officers, Except Agriculture	0.0141	0.0209	0.0068	1
Human Resources, Training, and Labor Relations Specialists	0.0167	0.0235	0.0068	1
Management Analysts	0.0169	0.0168	-0.0002	1
Accountants and Auditors	0.0198	0.0259	0.0062	1
Insurance Underwriters	0.0192	0.0286	0.0094	1
Computer Scientists and Systems Analysts/Network systems Analysts/Web Developers	0.0102	0.0145	0.0043	1
Computer Programmers	0.0091	0.0131	0.0041	1
Operations Research Analysts	0.0108	0.0127	0.0019	1
Architects, Except Naval	0.0130	0.0181	0.0051	1
Aerospace Engineers	0.0133	0.0121	-0.0012	1
Chemical Engineers	0.0078	0.0088	0.0010	1
Civil Engineers	0.0124	0.0119	-0.0005	1
Electrical and Electronics Engineers	0.0101	0.0104	0.0003	1
Industrial Engineers, including Health and Safety	0.0123	0.0121	-0.0002	1
Mechanical Engineers	0.0142	0.0128	-0.0014	1
Engineers, nec	0.0132	0.0117	-0.0015	1
Biological Scientists	0.0067	0.0067	0.0000	1
Chemists and Materials Scientists	0.0108	0.0130	0.0021	1
Environmental Scientists and Geoscientists	0.0113	0.0124	0.0011	1
Psychologists	0.0190	0.0105	-0.0085	1
Counselors	0.0088	0.0127	0.0039	1
Social Workers	0.0088	0.0083	-0.0005	1
Clergy	0.0086	0.0057	-0.0029	1
Religious Workers, nec	0.0107	0.0107	-0.0001	1
Lawyers, and judges, magistrates, and other judicial workers	0.0198	0.0245	0.0046	1
Postsecondary Teachers	0.0128	0.0126	-0.0002	1
Elementary and Middle School Teachers	0.0080	0.0050	-0.0030	1
Secondary School Teachers	0.0048	0.0051	0.0003	1
Other Teachers and Instructors	0.0103	0.0119	0.0016	1
Librarians	0.0117	0.0099	-0.0018	1
Artists and Related Workers	0.0119	0.0115	-0.0004	1
Actors, Producers, and Directors	0.0194	0.0189	-0.0005	1
Athletes, Coaches, Umpires, and Related Workers	0.0208	0.0141	-0.0067	1
Musicians, Singers, and Related Workers	0.0157	0.0139	-0.0019	1
Editors, News Analysts, Reporters, and Correspondents	0.0181	0.0188	0.0007	1
Public Relations Specialists	0.0218	0.0203	-0.0014	1
Technical Writers	0.0115	0.0142	0.0027	1
Writers and Authors	0.0187	0.0164	-0.0024	1
Dentists	0.0077	0.0078	0.0001	1
Dieticians and Nutritionists	0.0128	0.0145	0.0017	1

Pharmacists	0.0096	0.0064	-0.0032	1
Physicians and Surgeons	0.0117	0.0111	-0.0006	1
Registered Nurses	0.0165	0.0137	-0.0028	1
Physical Therapists	0.0243	0.0126	-0.0117	1
Speech Language Pathologists	0.0106	0.0057	-0.0049	1
Therapists, nec	0.0158	0.0080	-0.0078	1
Clinical Laboratory Technologists and Technicians	0.0144	0.0099	-0.0045	1
Advertising Sales Agents	0.0174	0.0230	0.0055	1
Insurance Sales Agents	0.0137	0.0186	0.0049	1
Securities, Commodities, and Financial Services Sales Agents	0.0157	0.0355	0.0197	1
Aircraft Pilots and Flight Engineers	0.0059	0.0037	-0.0023	1
Farmers, Ranchers, and Other Agricultural Managers	0.0087	0.0085	-0.0002	0
Food Service and Lodging Managers	0.0172	0.0126	-0.0047	0
Property, Real Estate, and Community Association Managers	0.0174	0.0144	-0.0029	0
Wholesale and Retail Buyers, Except Farm Products	0.0212	0.0199	-0.0013	0
Purchasing Agents, Except Wholesale, Retail, and Farm Products	0.0149	0.0172	0.0022	0
Claims Adjusters, Appraisers, Examiners, and Investigators	0.0133	0.0111	-0.0022	0
Other Business Operations and Management Specialists	0.0147	0.0235	0.0088	0
Drafters	0.0213	0.0148	-0.0065	0
Engineering Technicians, Except Drafters	0.0167	0.0129	-0.0038	0
Surveying and Mapping Technicians	0.0129	0.0147	0.0018	0
Chemical Technicians	0.0170	0.0206	0.0035	0
Life, Physical, and Social Science Technicians, nec	0.0125	0.0113	-0.0012	0
Paralegals and Legal Assistants	0.0221	0.0169	-0.0051	0
Preschool and Kindergarten Teachers	0.0104	0.0091	-0.0013	0
Teacher Assistants	0.0054	0.0143	0.0089	0
Designers	0.0179	0.0134	-0.0044	0
Photographers	0.0164	0.0069	-0.0096	0
Respiratory Therapists	0.0126	0.0132	0.0006	0
Dental Hygienists	0.0231	0.0006	-0.0225	0
Health Diagnosing and Treating Practitioner Support Technicians	0.0137	0.0165	0.0028	0
Licensed Practical and Licensed Vocational Nurses	0.0186	0.0122	-0.0064	0
Health Technologists and Technicians, nec	0.0180	0.0228	0.0048	0
Dental Assistants	0.0106	0.0079	-0.0028	0
Medical Assistants and Other Healthcare Support Occupations, nec	0.0124	0.0171	0.0047	0
First-Line Supervisors of Police and Detectives	0.0035	0.0066	0.0031	0
Firefighters	0.0044	0.0086	0.0042	0
Sheriffs, Bailiffs, Correctional Officers, and Jailers	0.0105	0.0095	-0.0010	0
Police Officers and Detectives	0.0117	0.0122	0.0005	0
Security Guards and Gaming Surveillance Officers	0.0177	0.0174	-0.0003	0
Crossing Guards	0.0349	0.0111	-0.0238	0
Law enforcement workers, nec	0.0143	0.0155	0.0013	0
Chefs and Cooks	0.0188	0.0174	-0.0014	0
First-Line Supervisors of Food Preparation and Serving Workers	0.0219	0.0211	-0.0008	0
Food Preparation Workers	0.0157	0.0112	-0.0045	0
Bartenders	0.0119	0.0065	-0.0054	0
Counter Attendant, Cafeteria, Food Concession, and Coffee Shop	0.0115	0.0050	-0.0064	0
Waiters and Waitresses	0.0104	0.0069	-0.0034	0
Food preparation and serving related workers, nec	0.0113	0.0082	-0.0031	0
First-Line Supervisors of Housekeeping and Janitorial Workers	0.0137	0.0164	0.0027	0
First-Line Supervisors of Landscaping, Lawn Service, and Groundskeeping Workers	0.0190	0.0134	-0.0056	0
Janitors and Building Cleaners	0.0150	0.0120	-0.0029	0
Maids and Housekeeping Cleaners	0.0050	0.0060	0.0009	0

Grounds Maintenance Workers	0.0180	0.0146	-0.0035	0
First-Line Supervisors of Personal Service Workers	0.0115	0.0075	-0.0040	0
Nonfarm Animal Caretakers	0.0113	0.0036	-0.0077	0
Entertainment Attendants and Related Workers, nec	0.0064	0.0095	0.0031	0
Barbers	0.0180	0.0096	-0.0084	0
Hairdressers, Hairstylists, and Cosmetologists	0.0200	0.0109	-0.0091	0
Childcare Workers	0.0200	0.0124	-0.0076	0
Recreation and Fitness Workers	0.0067	0.0061	-0.0005	0
First-Line Supervisors of Sales Workers	0.0151	0.0184	0.0033	0
Cashiers	0.0170	0.0153	-0.0017	0
Counter and Rental Clerks	0.0169	0.0098	-0.0071	0
Parts Salespersons	0.0115	0.0151	0.0036	0
Retail Salespersons	0.0174	0.0173	-0.0001	0
Sales Representatives, Services, All Other	0.0173	0.0207	0.0034	0
Sales Representatives, Wholesale and Manufacturing	0.0144	0.0182	0.0038	0
Models, Demonstrators, and Product Promoters	0.0256	0.0080	-0.0175	0
Door-to-Door Sales Workers, News and Street Vendors, and Related Workers	0.0149	0.0091	-0.0058	0
Sales and Related Workers, All Other	0.0169	0.0213	0.0044	0
First-Line Supervisors of Office and Administrative Support Workers	0.0133	0.0182	0.0049	0
Telephone Operators	0.0168	0.0093	-0.0076	0
Bill and Account Collectors	0.0181	0.0153	-0.0028	0
Billing and Posting Clerks	0.0153	0.0217	0.0064	0
Bookkeeping, Accounting, and Auditing Clerks	0.0141	0.0180	0.0040	0
Payroll and Timekeeping Clerks	0.0147	0.0140	-0.0007	0
Bank Tellers	0.0162	0.0139	-0.0024	0
File Clerks	0.0082	0.0207	0.0125	0
Hotel, Motel, and Resort Desk Clerks	0.0153	0.0105	-0.0048	0
Interviewers, Except Eligibility and Loan	0.0227	0.0243	0.0016	0
Library Assistants, Clerical	0.0127	-0.0056	-0.0184	0
Loan Interviewers and Clerks	0.0157	0.0216	0.0059	0
Correspondent clerks and order clerks	0.0047	0.0165	0.0118	0
Human Resources Assistants, Except Payroll and Timekeeping	0.0243	0.0197	-0.0047	0
Receptionists and Information Clerks	0.0139	0.0130	-0.0009	0
Reservation and Transportation Ticket Agents and Travel Clerks	0.0121	0.0266	0.0145	0
Information and Record Clerks, All Other	0.0185	0.0116	-0.0069	0
Couriers and Messengers	0.0196	0.0192	-0.0004	0
Dispatchers	0.0134	0.0117	-0.0017	0
Postal Service Clerks	0.0101	0.0140	0.0040	0
Postal Service Mail Carriers	0.0092	0.0137	0.0046	0
Production, Planning, and Expediting Clerks	0.0190	0.0160	-0.0030	0
Shipping, Receiving, and Traffic Clerks	0.0192	0.0184	-0.0008	0
Stock Clerks and Order Fillers	0.0150	0.0139	-0.0010	0
Weighers, Measurers, Checkers, and Samplers, Recordkeeping	0.0211	0.0052	-0.0159	0
Secretaries and Administrative Assistants	0.0146	0.0147	0.0000	0
Computer Operators	0.0190	0.0248	0.0058	0
Data Entry Keyers	0.0136	0.0214	0.0078	0
Word Processors and Typists	0.0237	0.0129	-0.0107	0
Mail Clerks and Mail Machine Operators, Except Postal Service	0.0195	0.0216	0.0021	0
Office Clerks, General	0.0174	0.0165	-0.0008	0
Office Machine Operators, Except Computer	0.0242	0.0291	0.0049	0
Office and administrative support workers, nec	0.0158	0.0197	0.0038	0
Agricultural workers, nec	0.0140	0.0131	-0.0009	0
First-Line Supervisors of Construction Trades and Extraction Workers	0.0129	0.0132	0.0003	0

Brickmasons, Blockmasons, and Stonemasons	0.0120	0.0110	-0.0010	0
Carpenters	0.0110	0.0073	-0.0037	0
Carpet, Floor, and Tile Installers and Finishers	0.0179	0.0132	-0.0047	0
Cement Masons, Concrete Finishers, and Terrazzo Workers	0.0035	0.0128	0.0093	0
Construction Laborers	0.0183	0.0126	-0.0057	0
Construction equipment operators except paving, surfacing, and tamping equipment operators	0.0126	0.0083	-0.0042	0
Drywall Installers, Ceiling Tile Installers, and Tapers	0.0094	0.0069	-0.0025	0
Electricians	0.0131	0.0106	-0.0025	0
Painters, Construction and Maintenance	0.0164	0.0102	-0.0061	0
Pipelayers, Plumbers, Pipefitters, and Steamfitters	0.0073	0.0114	0.0040	0
Roofers	0.0150	0.0091	-0.0059	0
Sheet Metal Workers, metal-working	0.0136	0.0124	-0.0011	0
Structural Iron and Steel Workers	0.0076	0.0108	0.0032	0
Helpers, Construction Trades	0.0124	0.0113	-0.0011	0
Construction and Building Inspectors	0.0092	0.0044	-0.0048	0
First-Line Supervisors of Mechanics, Installers, and Repairers	0.0109	0.0122	0.0013	0
Computer, Automated Teller, and Office Machine Repairers	0.0100	0.0138	0.0038	0
Radio and Telecommunications Equipment Installers and Repairers	0.0153	0.0100	-0.0053	0
Aircraft Mechanics and Service Technicians	0.0118	0.0093	-0.0025	0
Automotive Body and Related Repairers	0.0101	0.0121	0.0020	0
Automotive Service Technicians and Mechanics	0.0129	0.0110	-0.0019	0
Bus and Truck Mechanics and Diesel Engine Specialists	0.0128	0.0125	-0.0003	0
Heavy Vehicle and Mobile Equipment Service Technicians and Mechanics	0.0145	0.0125	-0.0020	0
Heating, Air Conditioning, and Refrigeration Mechanics and Installers	0.0115	0.0096	-0.0018	0
Industrial and Refractory Machinery Mechanics	0.0185	0.0169	-0.0016	0
Maintenance and Repair Workers, General	0.0130	0.0150	0.0020	0
First-Line Supervisors of Production and Operating Workers	0.0145	0.0151	0.0006	0
Electrical, Electronics, and Electromechanical Assemblers	0.0192	0.0197	0.0005	0
Assemblers and Fabricators, nec	0.0172	0.0178	0.0006	0
Bakers	0.0110	0.0158	0.0048	0
Butchers and Other Meat, Poultry, and Fish Processing Workers	0.0097	0.0166	0.0069	0
Cutting, Punching, and Press Machine Setters, Operators, and Tenders, Metal and Plastic	0.0176	0.0212	0.0036	0
Machinists	0.0188	0.0208	0.0020	0
Tool and Die Makers	0.0201	0.0197	-0.0005	0
Welding, Soldering, and Brazing Workers	0.0154	0.0176	0.0022	0
Metal workers and plastic workers, nec	0.0194	0.0216	0.0022	0
Bookbinders, Printing Machine Operators, and Job Printers	0.0196	0.0161	-0.0035	0
Laundry and Dry-Cleaning Workers	0.0167	0.0092	-0.0075	0
Sewing Machine Operators	0.0139	0.0060	-0.0078	0
Tailors, Dressmakers, and Sewers	0.0078	0.0090	0.0012	0
Cabinetmakers and Bench Carpenters	0.0127	0.0165	0.0038	0
Stationary Engineers and Boiler Operators	0.0108	0.0127	0.0018	0
Crushing, Grinding, Polishing, Mixing, and Blending Workers	0.0221	0.0215	-0.0006	0
Cutting Workers	0.0183	0.0260	0.0077	0
Inspectors, Testers, Sorters, Samplers, and Weighers	0.0135	0.0193	0.0059	0
Medical, Dental, and Ophthalmic Laboratory Technicians	0.0218	0.0213	-0.0006	0
Packaging and Filling Machine Operators and Tenders	0.0133	0.0151	0.0018	0
Painting Workers and Dyers	0.0147	0.0197	0.0050	0
Photographic Process Workers and Processing Machine Operators	0.0189	0.0151	-0.0038	0
Other production workers including semiconductor processors and cooling and freezing equipment operators	0.0188	0.0182	-0.0005	0

Supervisors of Transportation and Material Moving Workers	0.0138	0.0151	0.0013	0
Flight Attendants and Transportation Workers and Attendants	0.0054	0.0118	0.0064	0
Bus and Ambulance Drivers and Attendants	0.0139	0.0117	-0.0021	0
Driver/Sales Workers and Truck Drivers	0.0164	0.0155	-0.0009	0
Taxi Drivers and Chauffeurs	0.0104	0.0077	-0.0026	0
Parking Lot Attendants	0.0228	0.0119	-0.0109	0
Crane and Tower Operators	0.0183	0.0167	-0.0016	0
Industrial Truck and Tractor Operators	0.0118	0.0147	0.0030	0
Cleaners of Vehicles and Equipment	0.0162	0.0166	0.0004	0
Laborers and Freight, Stock, and Material Movers, Hand	0.0132	0.0164	0.0032	0
Packers and Packagers, Hand	0.0146	0.0119	-0.0027	0