When Cities Grow: Urban Planning and Segregation in the Prewar US[*]

Sun Kyoung Lee[†]
Yale University

November 2021

Abstract

I construct a historical, longitudinal database that follows individuals to analyze various aspects of urbanization in America's largest city, New York, between 1870 and 1940. This includes occupational changes as well as inter- and intra-city migration during a period of enormous transportation infrastructure improvements and the nation's first comprehensive land- use regulation (called "zoning"). I combine this longitudinal data with newly digitized transaction-level real estate sales records and city- and county-level Census of Manufactures outcomes. I use these data to study how both subway development and zoning have reshaped the internal structure of the city. I find that after the railway construction and zoning implementation, urban land use became highly separated and housing prices increased; additionally, neighborhoods became increasingly segregated both by income and race. Individual panel data reveal that returns to migration from rural areas to urban cities were substantial (with average earning increases of 40%) for native and foreign-born individuals of all races. However, the size of urban premia differs across white and African American (more generally, non-white) men. Although African Americans "fared" better in the big city than in the South, they were largely excluded from the most innovative and high-growth industries: manufacturing and finance.

---

[*] My sincere gratitude goes to Naomi Lamoreaux, Suresh Naidu, Dan O'Flaherty, Rohini Pande and Paul Rhode for their continued invaluable advice and support. My special thanks to Don Davis, Gerald Jaynes, and José Scheinkman for invaluable insights they share with me. I would like to thank Lorenzo Caliendo, José-Antonio Espin-Sanchez, Cecilia Fieler, Sam Kortum, Costas Meghir, Guillermo Noguera, Michael Peters, Luca Perdoni, Peter Schott, Ebonya Washington and seminar participants at Columbia and Yale Economics Department and Economic Growth Center. I am indebted to Arthur Erlendsson for his excellent research assistance. I thank students of Yale's undergraduate economics course "Economics of Cities" who motivated me to think about neighborhoods that we want to live in and need to build. Finally, I want to thank Costas Arkolakis for his years-long enthusiasm and support on our big, ambitious endeavors. Financial support from National Science Foundation (under grant \#1658827, grant \#1831524), Economic History Association, Lincoln Institute, Program for Economic Research at Columbia University, and Kwan Jeong Educational Foundation is gratefully acknowledged. All errors are my own.

[†] sunkyoung.lee@yale.edu

1 Introduction: New York's Rise as the Megacity

What is the role of urban planning during cities' explosive economic and population growth? In the early stages of growth, economic development is characterized by urbanization —a spatial transformation process where the population moves from rural to urban locations (typically cities). This transformation process is accommodated by urban planning interventions that are designed to maximize growth and to reduce congestion forces. Public policy intervention such as subway construction or zoning restrictions have been considered essential: transit infrastructure enables shorter commuting time while zoning prevents activities with large negative externalities (e.g. manufacturing that generate pollutants harmful to health).[1] In this paper, I study the ‘‘unintended'' consequences of urban planning (i.e. transit infrastructure investment and zoning) in America's biggest city during its era of unprecedented economic and population growth (1870-1940). To do so, I use frontier machine learning techniques along with brand new digitized data from original sources.

I establish that after the railway arrival, neighborhoods became increasingly segregated both by income and race (i.e. increased segregation *between* neighborhoods). I also show that the nation's first comprehensive zoning policy (especially residential zoning policy) may have caused African Americans to become extremely ‘‘isolated'' in few minority-neighborhoods (i.e. increased segregation *within* neighborhoods). Based on my findings, I further investigate the impact of residential ‘‘isolation'' on labor market outcomes for African Americans in New York. In my context, residential "isolation" of minorities against the majority population could mean that minorities' job opportunities may be systematically limited largely due to their social interactions. This provides a nineteenth century example of Topa (2001)'s finding that contact process and social interactions could explain labor market outcomes.

Despite the findings on residential segregation, my longitudinal data analyses of rural to urban migrants show that the returns to migration from rural areas to urban cities like New York were substantial for native and foreign-born individuals of all races (with about 37% and 40% earnings increases for white and African American population respectively). Previous works on urbanization that measured urban wage premia during my study period were largely based on cross sectional data due to the data constraints.[2] Instead, using the longitudinal individual panel data that I construct, I show that African Americans migrating from the South ‘‘fared'' better in larger cities. I discuss factors that could have helped them "fare" even better in the absence of certain forces;

---

[1] For example, ‘‘mega cities'' of the world (such as London, Paris and Boston in the United States) had already built rapid commuting transit infrastructure in the late nineteenth century (Heblich, Redding and Sturm (2020)). In terms of regulating the urban land use, at the turn of the century ideas and urban planning tools of regulating urban land were accepted under a favorable climate. This was with the assumption (or hope) that cities could and *should* be developed according to ‘‘rational'' plans (Brook and Rose (2013)).

[2] Boustan, Bunten and Hearey (2018) uses various sources of data including IPUMS and provides an excellent overview of the US's urbanization over for the last two hundred years.

these forces include exclusion from high-growth industries and extreme, residential segregation.[3]

My investigation is based on the collection and harmonization of large, mostly newly-digitized data from original sources. To gain a better understanding of what was going on during the city's rapid growth, I combine complete count population censuses (1870-1940) that captures who lived where; I add ‘‘spatial dimensions'' to the population censuses that capture who lived ‘‘where'' by georeferencing census enumeration district boundary files and harmonizing constantly changing census enumeration district boundaries into time-consistent neighborhoods.[4] Then, to gain an understanding of migrants from rural areas to large urban cities like New York, I use machine-learning methods to construct individual panel data by linking population census data without unique individual identifiers (e.g. social security number) using time-invariant information from the census.[5] Next, to measure the impact of urban planning policies, I collect zoning and subway planning documents and convert those "relics" of maps into Geographic Information System-compatible shape-files for performing spatial analyses on time-consistent neighborhoods. Finally, changes in housing prices amidst public policy interventions and population growth, I digitize transaction-level real estate sales records that amount to approximately 450,000 transactions in New York during 1870 to 1940. This trove of house transaction data at a daily frequency across all neighborhoods in New York City captures how housing prices. This trove of house sales transaction data at daily frequency across all neighborhoods in New York City captures how the housing prices evolved as the city's high productivity increased workers' earnings.[6]

Through the lens of ‘‘tipping'' and the dynamics of segregation (as in Card, Mas and Rothstein (2008)), I establish the connection between housing prices and residential segregation over time. I show the value of ‘‘tipping point'' (i.e. share of minority share in each neighborhood where mixed equilibrium is no longer sustainable) by following the same neighborhoods and document the population inflows and outflows in ‘‘tipping'' neighborhoods. I also show that the housing prices changes of ‘‘tipping'' neighborhoods—housing prices had gradually decreased throughout the dynamic ‘‘tipping'' process. This may be due to a dramatic population loss —I document

---

[3] In other words, as in Topa (2001); Glaeser, Sacerdote and Scheinkman (1996) suggest, social interactions confined by highly segregated minority neighborhoods could have been a contributing factor to African Americans' exclusion from high-growth industries such as manufacturing industries.

[4] Shertzer, Walsh and Logan (2016) makes a similar effort in georeferencing census enumeration district boundary files of northern cities, including New York, from 1900 and 1930. However, they only did this work for Manhattan and Brooklyn. A similar effort for 1880 was made by Logan under the Urban Transition Historical GIS Project.

[5] To investigate migrants' urban wage premia, I construct individual panel data (from 1870 to 1940) to follow the same individuals over time. This enables me to document one's occupation and industries in both periods (i.e. all census information from when the same person lived in a rural area and all census information from when the same person later migrated into a big city: "all census information" includes individual characteristics such as race and age, one's residential neighborhood at different points in time and occupation and industry information). This way, I can measure one's own urban wage premium rather than measuring urban wage premia from the cross-sectional data alone.

[6] Although I did not fully discuss the data here, I also digitized other complementary data sources such as county-city-level by industry growth information from the Census of Manufactures, and industrial location surveys that document the location of industries and jobs.

that inflow of one African American in tipping neighborhoods is associated with outflow of 2.5 white population, leading to the net population loss. Therefore, a dramatic population loss in "tipping" neighborhoods could explain the decrease of housing prices. Considering Boustan (2010)'s finding that each black arrival led to 2.7 white departures in the Northern cities during the postwar suburbanization, my neighborhood-level finding (i.e. each black arrival is associated with 2.5 white departures) seems to imply that the neighborhood-level tipping process was very much underway *well* before WWII.

Other than the papers mentioned above, this paper is related to several strands of literature. First, my work is closely related to residential segregation. Boustan (2010), for example, shows that the distinctive American pattern where blacks live in cities and whites live in suburbs (postwar suburbanization) was triggered by a large black migration from the rural South. Boustan (2010) shows whites responded to black influx from the rural South by leaving cities ("white flight").[7] Cutler, Glaeser and Vigdor (1999) documents changes in racial segregation in the United States over time across cities and finds that the period 1890 to 1940 saw the birth of ghetto in the United States, accompanied by perhaps African American migration from the rural South to the North. However, Logan and Parman (2015) derives a new neighbor-based segregation measure and establishes that the dramatic increase in segregation in the twentieth century was driven by a national increase in racial sorting at the household level (*not* by urbanization, black migratory patterns, or white flight). Related to aforementioned works, my paper finds that residential zoning that was first implemented in 1916 New York (and widely accepted in American cities in the 1920s) may have induced the increased residential racial segregation in the twentieth century.[8][9]

---

[7] Boustan (2016, 2010) that explores the effect of black in-migration between 1940 and 1970. Boustan (2016) shows that the black migration produced "winners and losers" in the black community as the competition between existing black workers and southern black migrant lowered the wages of African American male workers in the North tremendously. Boustan (2016) also finds that despite this competition, mass migration from the South was advantageous to the average black worker which is consistent with my findings from individual panel data that span till WWII.

[8] New York City's 1916 Zoning Resolution was the first land use regulation that puts restrictions on "use" restrictions. Other cities such as Los Angeles and Boston may have had some requirements about height; the 1901 Tenement House Law in New York City had imposed height and lot coverage restrictions on multi-family dwellings, but commercial and industrial buildings were still unregulated (Ward and Zunz (1992)). Up until 1916, no city had attempted to segregate "land uses". New York City's 1916 Zoning Resolution was the first zoning law in the nation that segregated residential uses from commercial and industrial uses, and created exclusive districts for single family houses. New York City's zoning resolution helped launch the rapid spread of zoning laws in other American cities during the 1920s. The 1916 Zoning Resolution was superseded in 1961, but was in effect till the end of my study period.

[9] In addition, relative to Boustan (2016, 2010) and other works on postwar suburbanization, my study period is structurally different from the postwar period in two ways: 1. automobile-based-suburbs and employment centers were not fully built yet for New York (Eli, Hausman and Rhode (2021)), 2. instead of the majority group fleeting to suburbs ("white flight/urban flight"), pre-WWII New York was built with " `separate, but equal`" neighborhoods. Methodologically, while Boustan (2016, 2010) looks at the effect of black in-migration in Northern cities through cross sectional data, I take the individual panel data (that include rural to urban migrants of all races) with higher-resolution data (down to "streets" and "neighborhoods").

Relatedly, there is a growing strand of research that studies impact of of the US government's discriminatory policies and practices. One major strand of research on this measures impacts of the "redlining" in the 1930s whereby the Home Owners' Loan Corporation (HOLC) develops maps that limited the access to credit for neighborhoods that were assigned as "red (risky and undesirable)" grade. Fishback, LaVoice, Shertzer and Walsh (2020) suggests that racial bias in the construction of the HOLC maps can explain at *most* 4 to 20 percent of the observed concentration of black households in the lowest-rated ("redlined") zones. Fishback, LaVoice, Shertzer and Walsh (2020) also suggests that the majority of black households were located in such zones *well* before the federal government's involvement in mortgage markets. My findings that the residential zoning (done almost two decades before "redlining") contributed to "isolating" the majority of African American households are consistent with their findings that HOLC's redlining and its impacts may have been, in fact, relatively smaller than conventionally perceived.

Second, my paper contributes to the literature of social (i.e. non-market) interactions, which are not regulated by price mechanism, among individuals. Scheinkman (2021) defines social interactions as particular forms of externalities, in which the actions of a reference group affect an individual's preferences. Schelling (1971)'s agent-based model, where individual tendencies regarding neighbors could lead to high segregation, was a pioneering work in social interactions. Each person's action changes not only because of the direct change in fundamentals, but also because of the change in the behavior of their peers. All the indirect effects then result in the *social multiplier* (Scheinkman (2021)). Glaeser, Sacerdote and Scheinkman (1996), for example, asks the question of "why observed crime rates across large American cities seem to vary too much" to be explained by the usual factors such as benefits of crime and changes in the exogenous costs of crime; Glaeser, Sacerdote and Scheinkman (1996) shows social interactions among individuals are the key components in creating enough covariance across individuals to explain the seemingly too high variance in high crime rates across American cities. The idea that "when social interactions exist, the impact of exogenous increase in a variable can be quite large" was extensively studied in works including Glaeser and Scheinkman (2000) and Card, Mas and Rothstein (2008). Relative to the big literature on social interactions, my paper shows the dynamic process of racial segregation through a long-horizon, high-resolution dataset covering everyone who lived in those neighborhoods that "tipped" along with the observation of large differences in outcomes (i.e. housing prices by neighborhoods over time) yet lacked a corresponding differences in fundamentals (Scheinkman (2021)).

Third, my paper contributes to the literature on land-use regulation. Gyourko and Molloy (2015) reviews the existing literature on local government regulation of urban land and concludes that most studies have found substantial effects on the housing market. The vast majority of papers on zoning have focused on effects in the housing market.[10] Only a handful of papers

---

[10] For example, most papers on zoning find that zoning raises house prices and reduces construction.

on zoning study effects *beyond* housing market. Kahn, Vaughn and Zasloff (2010) is a notable exemption exemption that makes a connection between residential regulation and sorting. They, for instance, look at changes changes in neighborhood income after a regulation is enacted and show that after the regulation on construction near the California coastline, household income of census tracts inside the zone rose faster than outside the zone. This finding implies that housing supply regulation could influence the location choices of households of different races and incomes, and therefore induce a higher level of residential sorting. My paper, therefore, speaks to both strands of research on land-use regulation—my digitized house price data reveals that zoning increased house prices; and residential zoning might have led to household sorting by race.

My paper speaks to a large, growing literature that studies how the provision of expansive public transport infrastructure investments affect a city's population, internal structure of the city, and its economic growth (Duranton and Puga (2014); Redding and Turner (2015)).[11] Relative to most papers that study the role of transit infrastructure investments, my paper focuses on the "unintended" consequences of them —segregating neighborhoods by income and race. Related to measuring "unintended" impacts of transit infrastructure, Tsivanidis (2019) investigates welfare distribution across workers with different skills when the world's largest bus rapid transit; Tsivanidis (2019) finds little impact on inequality across low- and high-skilled workers. Several papers connect the transit infrastructure and economic activities. For example, Heblich, Redding and Sturm (2020) uses the invention of steam railways in the 19th century London to document the role of separating the workplace and residence in supporting concentrations of economic activity. Baum-Snow (2007) demonstrates that the construction of new limited-access highways caused central city population decline. Relatedly, Glaeser and Kahn (2004) views automobile prevalence as the single most important driver of urban sprawl.

Finally, my paper also generates novel data from original sources in various formats.[12] For

---

[11] Urban historians including Bairoch (1991) argue that the cost of moving residents within cities is a key impediment to urban growth.

[12] For example, I use the digitized city- and county-level Census of Manufactures from 1860 to 1940 covering more than 100 major American cities and corresponding counties. This captures the different types of industry growth across location at the very time period of the nation's second industrial revolution. Also, I digitize transaction-level house sales records in New York starting from 1870 to 1940, totaling approximately 450,000 observations of real estate transactions across locations over a long horizon. I also create a big longitudinal individual database by linking historical individual census records using "machine learning." Similar efforts have been pioneered by Goeken, Huynh, Lynch and Vick (2011) that create the IPUMS linked samples; Feigenbaum (2015) links individual records in the 1915 Iowa State Census to their adult-selves in the 1940 US Federal Demographic Census records; Abramitzky, Boustan, Eriksson, Feigenbaum and Perez (2021) also evaluate different automated methods for record linkage and provide longitudinal datasets from historical US census records. Price, Buckles, Van Leeuwen and Riley (2019); Bailey, Cole, Henderson and Massey (2017); Nix and Qian (2015) make methodological contributions to the record linking methodologies as well.

Relative to the mentioned work, my machine-learning record linking that integrates the family information such as parents and/or spouse. Information such as birthplaces and names could further reduce the false positives and link more individuals as my methodology is more likely to link individuals that otherwise had to be dropped. Since people with common characteristics such as common first and last names may be systematically under-represented in linked datasets, better linking may reduce selection bias coming from the common characteristics.

example, I generate a wealth of spatial data that previously existed in "relics" (originally existed in papers and scanned images), converting them into sources that researchers can take to the spatial analyses. As an example, the impact of zoning has been studied primarily in modern contexts due to data constraints. However, as I connect various dimensions of high-resolution spatial data and population censuses, data that was in the past studied on its own can be taken together to tackle a set of questions that seemed almost unattainable. For instance, residential segregation was studied very thoroughly on its own (as in Logan and Parman (2015)). However, as now I combine racial segregation data and public policies with a spatial dimension (i.e. neighborhood-variations of zoning), I could establish the impact of residential zoning on residential segregation. Similarly, I can now quantify the impact of land-use regulation on housing prices, and investigate how housing prices evolved when the neighborhoods "tipped". Moreover, by connecting the spatial distribution of races prior to the 1940s with segregation today, I can measure long-run "unintended" impacts of public policies *well* before redlining.

The remainder of the paper is structured as follows. Section 2 discusses the relevant background of the study, and Section 3 discusses the data and methodology. Section 4 discusses a theoretical framework of the dynamics of residential segregation. Section 5 discusses the reduced-form evidence, whereas Section 6 specifically investigates people of different classes and races lived at a time of city's unprecedented urban transition and economic growth. Finally, Section 7 concludes.

2 New York City Background: 1870-1940

2.1 Population Growth

The total population of New York City increased from 1.48 million to 7.5 million during my study period (1870-1940);[13] the total city population experienced an astonishing growth with its peak population growth rate being 39% over a decade. However, beginning in the early twentieth century, Manhattan experienced a dramatic loss in population while all outer boroughs were gaining population at an unprecedented rate (for example, between 1920 and 1930, Manhattan lost 18% of its population when the population in Queens and Bronx grew by 130% and 73% respectively).

2.2 Labor Market Trends in New York

New York was growing in skill during the study period, as well as in population, and this growth in skill was occurring among almost all demographic groups. However, skill growth in the city was nowhere near as fast as population growth, and in some decades it faltered slightly. New York was more skilled than the rest of the nation during the study period, but its advantage was eroding.

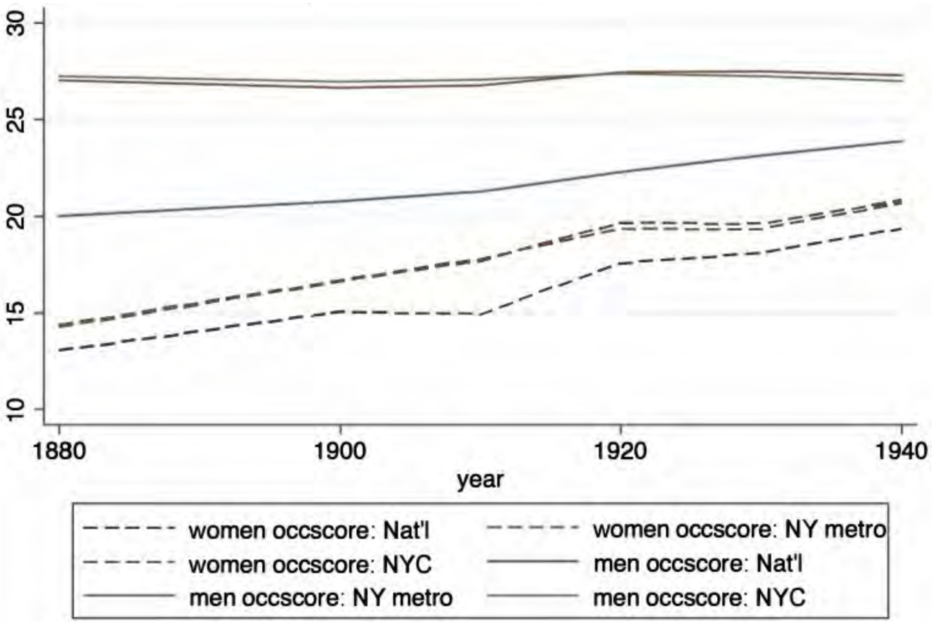
---

[13] In 1898, through the consolidation of NYC, outer boroughs (Brooklyn, Bronx, Queens, and Staten Island) were incorporated into New York City. For my analysis, I *always* define the city as five boroughs throughout the study period.

This aggregate skill growth matters for my analysis because it implies that *growth in skill in one neighborhood did not have to come at the expense of a reduction in skill in others.*

Data reveal that men in New York City and the New York metro area had significantly higher mean occupational income than the rest of the country in 1880, but that income converged to the rest of the country over the next 60 years. A similar pattern was observed for women but at a much smaller magnitude. Figure 1 shows mean occupational income trends for all employed men and women aged between 16-60 at varying geographic scopes.

Figure 1: Mean Occupational Income Trend: Men and Women



Note: Solid lines indicate men, whereas dotted lines indicate women; in terms of geography, the national average is in Blue, New York metro area average is in Red, and New York City average is in Green.

Source: The US complete-count census records. All observations are aged between 16-60 with reported occupations.

2.3 Industry Growth Trends in New York

Using the digitized Census of Manufactures, I document the nominal wage differences across space. Specifically, I document the nominal wages by industry in New York, relative to the other US locations over the study period. To my knowledge, there exists no rigorous, systematic study of nominal wage differences across time and space primarily due to the data constraints. Using the digitized data, I document how nominal wages for the given industry varied across space; especially, I document how wages in New York may have been systematically different relative to elsewhere; finally, I also document the ‘‘type’’ of industries that existed in only a few number of cities including New York. Documenting the type of industries that existed in ‘‘higher-order’’ cities but not in ‘‘lower-order’’ cities is necessary to understand New York's major economic transition from *"a previously second-order city to the unparalleled first-order city of the United States* (Fujita, Krugman and Venables (2001))."

2.4 Major Urban Planning I: The Transportation Revolution in NYC

Transit infrastructure improved dramatically at both the intra- and inter-city level during the study period. In particular, during the subway construction period between 1904 and 1920, the total number of stations grew by 200% and 113% in the Bronx, 87% and 105% in Brooklyn and 50% and 133% in Queens. Inter-city transit infrastructure improvements occurred at an unprecedented scale during the study period as well. For example, electrification of railroads in 1907 and 1914 greatly improved the efficiency and speed of railways. The Hudson Tubes connected NYC to New Jersey in 1908, and Penn Station's opening linked NYC to the rest of the country through inter-city railways in 1910.

Intra-city transit access by subways and the elevated railways construction Before the introduction of the subway in 1904, New York City had a large central business district in lower Manhattan and a smaller business district in downtown Brooklyn. These districts were served by elevated railways and ferries and most of the services were operating in Manhattan. Manhattan was the only borough with rapid mass transit commuting infrastructure before the introduction of the subway in 1904; most outer boroughs (i.e. Queens, Staten Island, and the Bronx) did not have transit network until the 1910s and were semi-rural and underdeveloped. The first decade of subway construction mostly served Manhattan and Brooklyn, while parts of Bronx, Queens and South Brooklyn received more subway constructions in the 1910s under the Dual Contracts. However, the rapid growth of the system largely was over by 1940.[14]

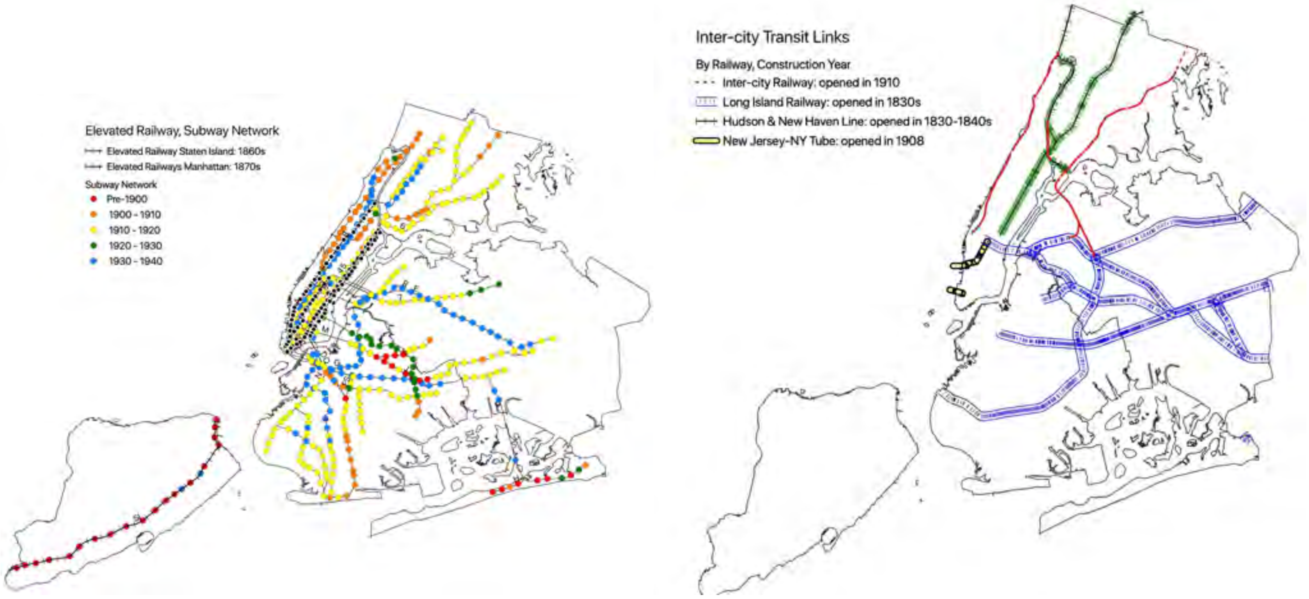
Inter-city transit access by railroad By the Year 1910, access to infrastructure-based inter-city transit in New York City experienced an unprecedented, spectacular growth. Inter-city transit infrastructure was largely concentrated in Midtown Manhattan, and the combination of both inter- and intra-city transit infrastructure improvements grew faster in Midtown than in Lower Manhattan. For example, inter-city railways connected NYC to the rest of the country with the opening of Penn Station in 1910 (Red line in the Figure 2).[15] As seen in Figure 2, inter-city transit

---

[14] The first underground line of the subway opened in 1904, almost 40 years after the opening of the first elevated railway in Manhattan. New York City's subways were built by two private companies (the Brooklyn Rapid Transit Company (BRT, later Brooklyn–Manhattan Transit Corporation, BMT) and the Interborough Rapid Transit Company (IRT)) as well as one city-owned company (Independent Subway System (IND)). In 1940, the city bought the two private systems and consolidated the transit network.

[15] Steam railroad began in 1830s New York with the Harlem Railroad (Green line); by the 1840s, the same line served central Westchester county; Long Island Railroad (LIRR)-based commuter service was established largely by the 1860s (Blue line); the Hudson Tubes, which became Port Authority Trans-Hudson (PATH) opened in 1908

Figure 2: Evolution of Spatial Links by the Intra- and Inter-city Transit Network



Note: The above figures show the evolution of within and between city spatial links in terms of the elevated railways and subways (for within-city transit network, Left figure) and railroad network (for between-city transit network, Right figure) over the study period. For within-city transit network (Left) different color dots denote the opening decades of transit links. For between-city transit network (Right) different color lines denote different inter-city transit links.

Source: Author's creation using New York City Department of City Planning's data called "LION" GIS data which is a base map representing the city's geographic features. I augment this GIS data with transit station opening year information from archival documents in the New York Transit Museum Archive. In investigating now-razed elevated railway stations, data from Historical Urban Ecological Data (Costa and Fogel (2015)) was extensively referenced. For railroad network and its construction year, I referenced information provided by the New York City Transit Authority and related books (`http://www.mta.info/`).

infrastructure improvements over the study period had triggered the city's transit hubs to expand from Downtown Manhattan to Midtown Manhattan. The extreme growth of Midtown Manhattan since 1910 was partly due to inter-city railway infrastructure that connected NYC to the *rest* of the country.

2.5 Major Urban Planning II: Nation's First Comprehensive Land-Use Regulation

Brook and Rose (2013) states that zoning as a general form of land use control came to the United States by way of Germany, where this type of regulation had been in use since the 1890s. In the United States, the zoning concept came along as practical means to meet the demand for rationally controlled urban growth. Especially after Chicago's 1871 fire, the idea that cities *could* and *should* be developed according to rational plans became more widely accepted. Urban planners and the city government of New York held the view that the subways did *nothing* to relieve the problem of crowding. They saw that the subways not only encouraged speculation in real estate to unrealistic levels but also increased instability and crowding, making city living less comfortable and less bearable. Municipal government viewed that it had no way of meeting the problems generated by population increases and subway expansion (Makielski (1966))

Therefore, at the turn of the century, the idea of zoning was "fresh and new and full of promise" (Brook and Rose (2013)). In 1916, New York City implemented the nation's first comprehensive, city-wide land use regulation called "the 1916 Zoning Resolution." This resolution was designed to regulate and limit the height and bulk of buildings, and to regulate and restrict the location of industries and the location of buildings designed for specified uses, and to establish the boundaries of districts for these purposes.[16] This nation's first comprehensive land use regulation is believed to be *"the most important step in the development of New York City since the construction of the subway."* (*City Fixes Limit On Tall Buildings* (1916); W.Dunlap (2016))

New York City's 1916 zoning had three layers of regulations —Height, Use, and Area. The *Height restrictions* regulated the height of the new buildings by considering the width of street that the building faces; the *Use restrictions* assigned the use of new buildings into three types: residential, business and unrestricted; the *Area restrictions* prevented new buildings from covering

---

(Yellow line). Most of these inter-city railways may have not been used for daily commuting purposes. For example, Jackson (1985) argues the first railroads were designed for long-distance rather than local travel or commuting. However, as railroad companies sought revenues, they built stations whenever their lines passed through rural villages on the outskirts of larger cities. Jackson (1985) argues that since inter-city railway fares were considered too high for most wage earners, such suburbanization was only for the "well-to-do."

[16] The Reformers felt strongly about overcrowding. Basset — "The Father of City Planning"—commented "My interest in zoning was largely based on sunlight." The other pivotal urban planner in the zoning commission, McAneny also stated "High buildings robbed their neighbors of light and air and filled the streets with the density of the moving population." A commission on building heights expressed grave public concerns — overcrowding, lack of light, fire safety risk due to large-building-to-narrow-street ratio. The full reports are available here: `https://archive.org/details/reportofheightso00newy/page/n91/mode/2up`

their entire sites, mandating open spaces at the rear and sides of the structure — the taller the building, the more space required on all sides (Makielski (1966); Ward and Zunz (1992)).

However, despite these ideals and promises, Brook and Rose (2013) also argues that zoning ordinances, in practice, turned out to have served considerably more mundane goals: preserving a preexisting status quo in already-developed areas, and defending the single family home from the "lower" uses of multiple dwellings, commerce, and manufacturing in that descending order. To quantify the impacts of the nation's first land use regulation, I collect and georeference all 1916 original maps of New York Zoning Ordinances, thereby creating GIS-compatible shapefiles which convert the "relics" into something compatible with 21st-century spatial analyses. Details of these processes and descriptions are available in Section 3.4 and Figure 23, 24, and 26(b).

3 Data on New York Economy and its Transition

I construct new, spatially high-resolution datasets on New York's economy for the period 1870-1940. My main source of data for NYC is the US federal population census and real estate data that I digitize from the archives. I augment other sources of data, and this augmentation includes the construction of panel data that follow the same individuals over time. Further details are discussed in the Appendix.

3.1 People

Residential Population Data in New York: 1870-1940 I use the restricted-access IPUMS complete count population census records from the US Federal Demographic Census from 1870 to 1940 (Ruggles, Flood, Goeken, Grover, Meyer, Pacas and Sobek (2019)) . These individual-level census records provide rich socioeconomic and demographic information such as one's occupation, industry, race, and family characteristics along with the residential location.[17]

- Neighborhood changes from repeated cross-sectional data

Using the population census, I document how neighborhoods changed in terms of resident composition using time-consistent neighborhood boundaries. However, as datasets including population censuses have different spatial units and/or the boundaries of spatial units constantly change, I take time-consistent neighborhood boundaries and create spatial crosswalks from historical locations

---

[17] I use the 1950 Census Bureau occupational classification system (henceforth, OCC1950)-based occupational measures of income and education to enhance comparability across the years. Ruggles, Flood, Goeken, Grover, Meyer, Pacas and Sobek (2019) coded occupation-based values according to the 1950 Census Bureau's classification. Throughout the analysis, I use OCC1950-based occupational income score (called "OCCSCORE") as measures of occupational standing. This approach controls for inflation and is widely used in the literature to measure individuals' skills.

in various data sources to time-consistent boundaries.[18] See the Appendix Section B.2 for details regarding this procedure.

- New longitudinal database and dynamic changes

However, complete-count population censuses only exist in a cross-sectional format and they do not have time-invariant individual identifier(s). The longitudinal tracking of individ- uals is essential in understanding the city's growth and neighborhood changes: if one observes a city or neighborhood at two different times, then one can observe only how the aggregates changed. Given any sequence of aggregates, there are a huge number of different individual sequences that can produce them, and those different collections of in- dividual sequences have different welfare interpretations. Therefore, to truly investigate the city's economic growth and people's migration decisions behind this, I follow individuals with different skills (or incomes) across locations over time.

Similar efforts have been pioneered by Goeken, Huynh, Lynch and Vick (2011) which creates the IPUMS linked samples. Feigenbaum (2015) links individual records in the 1915 Iowa State Census to their adult-selves in the 1940 US Federal Demographic Census records. Abramitzky, Boustan, Eriksson, Feigenbaum and Perez (2021) evaluates different automated methods for record linking, performing a series of comparisons across methods and against hand linking. Although I largely follow the standard machine-learning record linking methodology suggested by Goeken, Huynh, Lynch and Vick (2011), I have extended the techniques of Goeken, Huynh, Lynch and Vick (2011) by inventing a two-step machine learning matching methodology. See the Appendix Section A for details of the census record linking.

3.2 Economic Growth

Economic Activities from Digitized Census of Manufactures (1870-1940)

I digitize the county and/or city-level Census of Manufactures from 1870 to 1940 for every decade. These county- and city- aggregates are available by industry at about SIC 3-digit level every decade.[19] I digitize these tables to collect the average nominal wage, the number of workers, amount of capital invested, and quantity and value of materials by industry and county/city over time. Through this table, we can not only see how cities of different size began specializing and growing in certain industries, but also how the nominal wages differed across both cities and

---

[18] The primary geographic unit of the analysis is the "Neighborhood Tabulation Areas" (hereafter, NTAs), with a minimum population of 15,000 (there are 195 NTAs (neighborhoods) within the city). New York City Department of City Planning defines this geographic boundary. More details can be found here: `https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-nynta.page`

[19] This manufactures schedule was originally called the "industry schedule" and it collects information about manufacturers with an annual gross product of $500 or more.

industries over time. This information is essential in capturing the rise of mega cities including New York as the hub of manufacturing.

3.3 Urban Land Use

Location of jobs by industries over time

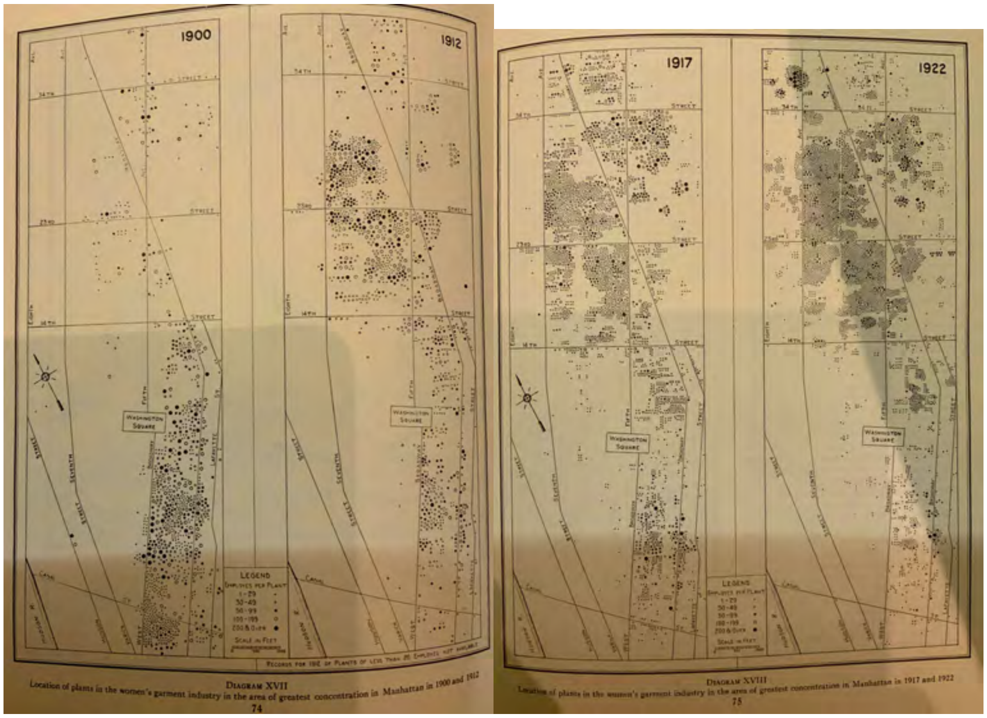
Although the population census provides an excellent information about individuals including one's industry and occupation, it does *not* list the job location. Therefore, to gain an idea of where jobs were located, I turn to `the Regional Survey of New York and its Environs in 1929`. From these industrial surveys and published book series, I digitize a set of key information including plant locations for each major industry in the city over time. This extensive industrial survey gathered and analyzed the data and documented (1) where jobs related to NYC's major industries were located, (2) changes of industrial locations which have recently been taking place (at the time of publication), and (3) the forces which are causing those changes. This industrial survey relies on the records of the factory inspection departments of New York, New Jersey, and Connecticut for 1900, 1912, 1917 and 1922.[20]

I digitize maps that feature the location of jobs by industry to accurately document the primary location of jobs in the city. Figure 3, for example, shows the location of women's clothing industry from 1900 and 1922. These maps reveal that the women's garment industry tended to concentrate in one central district, and the tendency toward concentration did not change despite the investments in commuting infrastructure. Similar tendencies for concentration in other industries such as textile, printing, wholesale markets, retail shopping and financial districts were observed throughout my study period.

---

[20] Under this survey, it documents the character, location, and the number of employees in each industrial establishment; and industries were selected on the basis of their size and their importance. For example, clothing and metals were selected as they employed roughly a quarter of a million workers in the city. Considering the industry character, it studied men's clothing separately from women's clothing. Altogether, the nine industrial studies covered 72% of all of the plants and 79.5% of all of the employees listed by the factory inspectors for the tri-state areas.

Figure 3: Location of Women's Clothing Industry (1900 - 1922)



Note: The above figure shows the location of plants in women's garment industry in New York in 1900, 1912, 1917, and 1922. It reveals that the women's garment industry tends to concentrate in one central district. The industrial survey states that this tendency toward concentration is due in part to the highly competitive and fickle nature of the industry (Haig (1926);Haig (1927); of New York and Environs (1928);of New York and Environs (1927)). I digitize these maps from their original publication (of New York and Environs (1927))

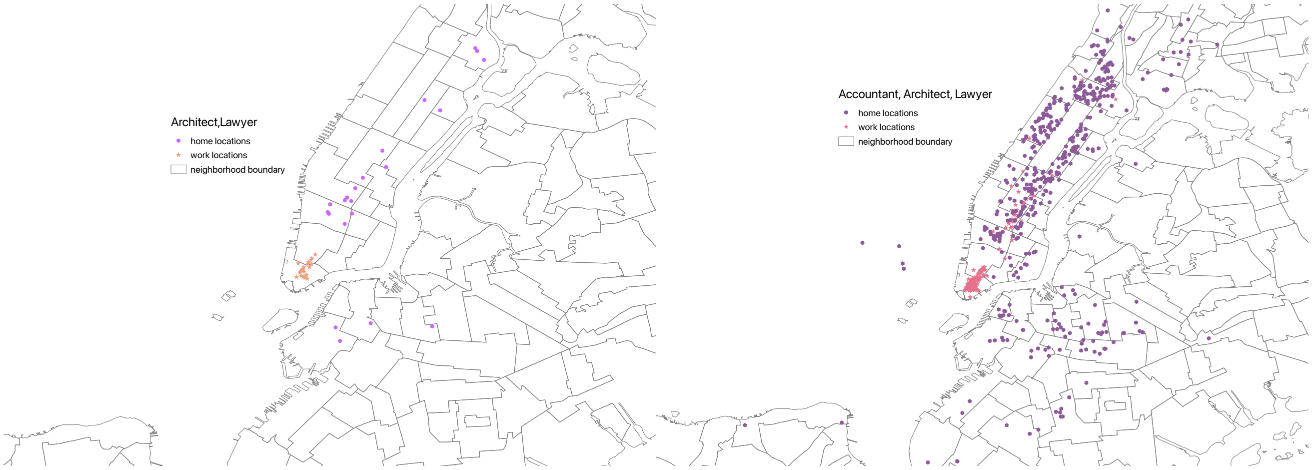
Job Locations and Commuting patterns over time

As the population census does *not* list one's job location, I turn to alternative sources to investigate the commuting pattern of the city's residents. For example, `Trow's City Directory` typically lists one's name, occupation, home and work locations during my period, and I use this source to document commuting pattern changes. For example, Figure 4 shows that work locations of accountants, architects and lawyers stayed in the same location near the Financial District, whereas home locations moved much further away from job locations as transit infrastructure enabled workers to commute to their work places easily.[21]

---

[21] Barr (2016) generously shared with me the digitized data from `the Trow's City Directory`.

Figure 4: Home and Work Locations from City Directory



Note: The above figures show the commuting patterns by plotting home and work locations for the same category of jobs (i.e. accountants, architects and lawyers) in 1879 (Left) and 1906 (Right). In 1879, in the absence of rapid commuting transit infrastructure, home and work locations were not entirely separated. Work locations were concentrated near Downtown in 1879 and people lived mostly near their work locations (captured by purple dots in the map on the left). However, by 1906, we see that work locations are still the same in Downtown (pink dots); however, people's home locations have chaned dramatically and many moved toward northern Manhattan (captured by purple dots) mostly along the rapid transit commuting infrastructure network.

Digitized Real Estate Transaction Records

I digitize the entire run of the *Real Estate Record and Builders' Guide* from 1870 to 1940 in order to investigate how the provision of urban transportation infrastructure and land-use regulation affected property values. Quantifying the effect of transit infrastructure and zoning on property values has proven difficult as some of the largest and most interesting public policies were undertaken long before the advent of computers, and therefore the necessary data has been locked up in disparate paper records.

Nicholas and Scherbina (2011) pioneers the digitization and construction of the hedonic house sales price index using the same source, *the Real Estate Record and Builders' Guide*, a weekly publication of real estate transactions. Relative to their works where they randomly hand-collect 30 transactions per month for Manhattan between 1920 and 1939 (totaling 7,538 observations), I collect *every* listed transaction which covers entire properties in the city and its surrounding areas (such as Westchester and Long Island, NY) from 1870 to 1940, totaling approximately 450,000 observations. For the final analysis, I excluded transactions without actual transacted price information.[22]

As discussed in Grebler (1955), *the Real Estate Record and Builders' Guide* has published data on the actual consideration and the assessed value of properties in New York City that were

---

[22] For example, I do not include transactions where transacted price was recorded as "OC & 100" meaning "Other Things Considered and $100" , or "nom" meaning "nominal"), or bundled properties as I cannot perform hedonic regressions on bundled transacted properties with one price.

conveyed in bona fide sales.[23] As I am particularly interested in capturing market transactions and property value changes, in constructing the hedonic house sales index, I only include observations where either the actual considerations are reported or assessed values for each property are available for my analysis. Typically, the transaction records list the address, building characteristics, assessed value and actual consideration price, mortgage and foreclosure if applicable, and finally buyer and seller information. Then, I take the digitized property addresses to get its geocoordinates and map each particular transaction into a neighborhood. Details and methodologies are available in the Appendix Section C.2.

From the digitized real estate sales records, I record the each property's location, property characteristics (e.g. total square footage, building materials, the number of stories, irregular shape of the property), date of transactions, along with buyer and seller information. Figure 5 shows typical transaction records. Here, in the first entry on the left figure (starting with Attorney St), I have the address of the property with block and lot number are in parentheses. "`e.s`" describes its orientation (east south side of the street) 75 feet south of Rivington street, the size of the lot is 25 feet by 50 feet, and this property was "house and lot `(h&l)`"; and Samuel Phillips, as the seller, sold this property to Frederick Hoch, the buyer, on Oct 31, 1871 at $8000.[24] Starting 1905, the records also began listing assessed values for each property. As Figure 5 on the right shows, the seller took a mortgage amount of $18,500 to purchase the property at $34,000 when the assessed values for this property ranged from $20,000 to $32,000.

I also systematically classify whether the type of transacted properties (e.g. house and lot, tenement, or plots of land and so on). Using the digitized real estate price records, I construct hedonic house sales price indexes. A Hedonic Price Index views a real estate property as a bundle of characteristics, and this collection of priced characteristics sums up to the transacted market price. Given the overall availability of house characteristics, I construct house prices using the hedonic price index.[25]

Finally, I add the additional property-level public-policy variables when constructing hedonic home sales prices indexes. To elaborate, I overlay digitized properties' exact geocoordinates and georeferenced 1916 zoning maps (use, height, area) using GIS software so that I can exactly attach

---

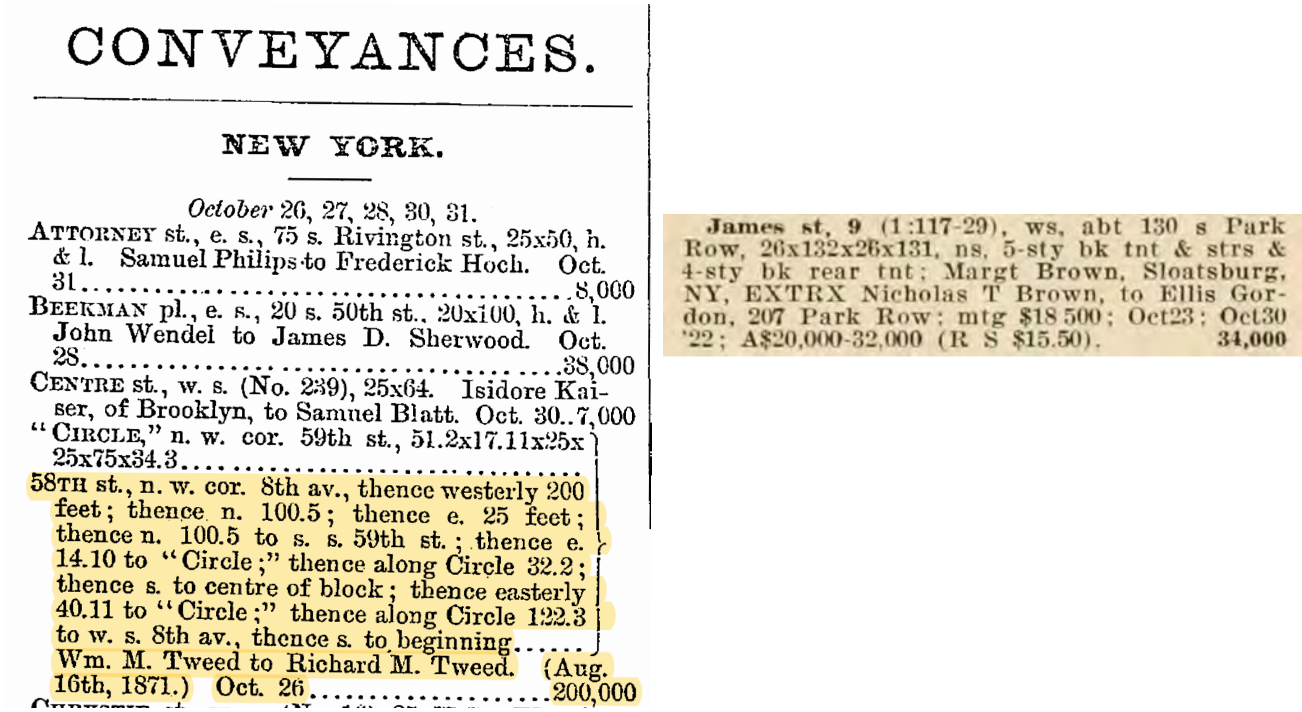
[23] Considering assessed values are not necessarily identical with market prices, the availability of actual transacted market values along with assessed value can provide insights that may be difficult to obtain otherwise. Grebler (1955) states that "the data in *the Real Estate Record and Builders' Guide* cover a varying percentage of total bona fide sales, namely those for which it was possible to obtain confidential information on actual consideration, as distinguished from newspaper announcements."

[24] There were some speculative activities as highlighted in the same figure. For example, Wm. M. Tweed, who was known as one of the most corrupt politician of Tammany Hall and the third-largest landowner in New York City, appeared frequently in my digitized transaction records.

[25] The most simple attempt would be an average or median prices by neighborhoods and year over all transactions without controlling for heterogeneous characteristics of transacted properties. Another commonly used index is the repeat sales index commonly known as "Case-Shiller" index. However, as the street names and house numbers have changed quite a lot over my study period (over 70 years), doing the repeat sales index was *not* ideal. See New York Times article on streets and building numbers change: `https://www.nytimes.com/interactive/2021/01/27/nyregion/brooklyn-streets-numbers-renaming.html`

the zoning regulations that were imposed to the transacted properties. I do the same procedure of attaching "redlining" grades drawn in the 1930s which restricted access to credit to affected neighborhoods.[26] See Section C.2 for more details on real estate records.

Figure 5: Real Estate Transaction Record Examples (1871 and 1922)



CONVEYANCES.

NEW YORK.

October 26, 27, 28, 30, 31.

ATTORNEY st., e. s., 75 s. Rivington st., 25x50, h. & l. Samuel Philips to Frederick Hoch. Oct. 31 8,000

BEEKMAN pl., e. s., 20 s. 50th st., 20x100, h. & l. John Wendel to James D. Sherwood. Oct. 28 38,000

CENTRE st., w. s. (No. 239), 25x64. Isidore Kaiser, of Brooklyn, to Samuel Blatt. Oct. 30..7,000

"CIRCLE," n. w. cor. 59th st., 51.2x17.11x25x 25x75x34.3..................................

58TH st., n. w. cor. 8th av., thence westerly 200 feet; thence n. 100.5; thence e. 25 feet; thence n. 100.5 to s. s. 59th st.; thence e. 14.10 to "Circle;" thence along Circle 32.2; thence s. to centre of block; thence easterly 40.11 to "Circle;" thence along Circle 122.3 to w. s. 8th av., thence s. to beginning...... Wm. M. Tweed to Richard M. Tweed. (Aug. 16th, 1871.) Oct. 26 200,000

James st, 9 (1:117-29), ws, abt 130 s Park Row, 26x132x26x131, ns, 5-sty bk tnt & strs & 4-sty bk rear tnt; Margt Brown, Sloatsburg, NY, EXTRX Nicholas T Brown, to Ellis Gordon, 207 Park Row; mtg $18 500; Oct23; Oct30 '22; A$20,000-32,000 (R S $15.50). **34,000**

Note: *The Real Estate Record and Builders' Guide* also began listing assessed values (minimum and maximum) of each property along with mortgage and foreclosure information starting 1905. Therefore, I digitize these additional entries as well, and use some of them to identify the financially distressed properties. I also use the assessed values as a way to check that the transcribed transacted price was "reasonable".

3.4 Major Public Policies

Transit Network

I have collected various subway and elevated railway datasets, including the data on each station in the existing New York transit system. The year each station has opened was determined in

---

[26] Rothstein (2017) argues that the origin of "redlining" come from government homeownership programs that were created as part of the 1930s-era New Deal. The government color-coded maps ranking the loan worthiness of neighborhoods in cities including New York. Neighborhoods were ranked from least risky (or "A" in color green) to most risky ("D" in color red).

order to estimate the subway opening, network, and station effects. Based on the compiled dataset, the evolution of the subway and the elevated train network from 1870 to 1940 is documented.

Land Use Regulation called "zoning"

I have collected the 1916 Zoning Resolution-related information. Specifically, I collect images of the 1916 zoning maps that specify the 1. Area, 2. Height, and 3. Land Use (i.e. residential/business/unrestricted restrictions). Georeferencing of the entire run of 1916 NYC zoning maps and creation of shapefiles enabled me to document zoning-regulations that were in place down to the street and building level. See the Appendix Paragraph B.2 for details.

4 Theoretical Framework of Tipping

In this section, I illustrate how the model of tipping from Card, Mas and Rothstein (2008) can explain the data pattern and dynamic patterns segregation from my analyses. This is a partial equilibrium model where endogenous housing supply is not considered. Consider a neighborhood with a homogenous housing stock of measure one and two groups of potential buyers: white (w) and minority (m).

Let $b^g(n^g, m)$ where $g \in \{w, m\}$ denote the inverse demand functions of the two groups for homes in the neighborhood when it has minority share (m). There are n^g families from group g who are willing to pay at least $b^g(n^g, m)$ to live there. Card, Mas and Rothstein (2008) also assumes that $\frac{\partial b^w}{\partial n^w}$ and $\frac{\partial b^m}{\partial n^m}$ are weakly negative (i.e. as there are more families from group g, the marginal bidder from group g will have a (weakly) lower willingness to bid than the first bidder from group g.

$\frac{\partial b^w}{\partial m}$ and $\frac{\partial b^m}{\partial m}$ represent social interaction effects on the bid-rent functions. Suppose there exist a threshold where (let's call that m^*) if the minority share in the neighborhood is above this threshold (i.e. $m > m^*$), $\frac{\partial b^w(n^w, m)}{\partial m} < 0$.

If the mth highest minority bidder has the same willingness to pay as the $(1 - m)$th highest white bidder, we have an integrated equilibrium where $b^m(m, m) = b^w(1 - m, m)$.

Taking the derivative of the white bid function ($b^w(n^w, m)$) with respect to the neighborhood minority share (m):

$$\frac{\partial b^w(1-m, m)}{\partial m} = -\frac{\partial b^w}{\partial n^w} + \frac{\partial b^w}{\partial m} \tag{1}$$

By construction, $\frac{\partial b^w}{\partial n^w} < 0$ and therefore the first term $-\frac{\partial b^w}{\partial n^w}$ is positive. The social interaction effect $\frac{\partial b^w}{\partial m}$is small when m is small (close to 0), leading $\frac{\partial b^w(1-m, m)}{\partial m}$ to be positive for small values of m.

However, as m increases, we observe how previously all-white neighborhood neighborhood could "tip". To elaborate, Card, Mas and Rothstein (2008) show through a illustrative figure (Figure 6) that low levels of minority share m result in a stable equilibrium. Then, as b^m shifts upward (for example, the Great Migration Era shifts b^m upwardly), housing prices begin to rise and a few minority families displace white families with the lowest willingness to pay.[27] Card, Mas and Rothstein (2008) shows that this mixed equilibrium (at low levels of m) is stable. However, if there exist further increases in the relative demand of minorities that push the minority's bidding curve (b^m) even upward, then eventually, m will rise until b^m is tangent to b^w.

Once the minority share reaches m^*, the level of minority share at its tangency, integrated mixed equilibrium disappears and it will move toward 100% minority equilibrium (i.e. $m = 1$) where minority population is perfectly "isolated", interacting solely with the minority population itself. Once this tipping process is underway, even if there the minority demand function (b^m) shifts downward, it will not reverse the tipping process —m will continue converging to the 100% minority equilibrium as long as m lies right to the unstable equilibrium.

Finally, Card, Mas and Rothstein (2008) notes that the value of the "tipping point" (m^*) depends on the white/majority population's tolerance for growing minority neighbors. If the majority's aversion toward minority is "strong" (captured by the sharp negative slope of $\frac{\partial b^w}{\partial m}$), then even at a relatively low level of minority share (m), the neighborhood could undergo the tipping process. Relatedly, Card, Mas and Rothstein (2008) discusses housing prices prediction. Rents at the 100% minority equilibrium (i.e. $m = 1$) could be higher or lower than the tipping point, depending both on the minority demand function (curvature of b^m) and shifts in minority demand once the tipping is underway.

4.1 Contextualizing Social Interaction and Residential Segregation

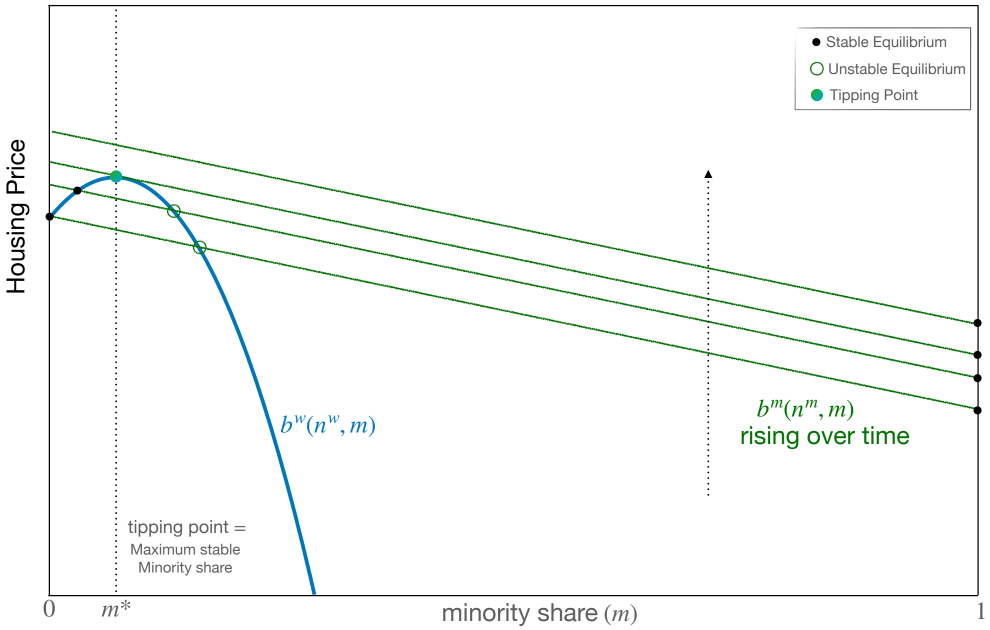
Through Card, Mas and Rothstein (2008)'s theoretical framework of a dynamic segregation process, I now take the digitized data that I construct (i.e. the composition of minority and majority population, and house prices changes over time-consistent boundaries) to the dynamics of segregation and neighborhoods' tipping process in America's largest city.

Cutler, Glaeser and Vigdor (1999) states that "Segregation rose dramatically with the influx of southern blacks, particularly in the industrial North. (...) To a great extent, the modern spatial distribution of races in American cities was established by 1940." In this sense, studying social interaction and residential racial residential segregation as a consequence of social interaction, and finally, connecting these two to the housing prices and labor market outcomes of minorities living in segregated neighborhoods as outcome variables (Section 5 and 6) is a natural step.

---

[27] In the context of my study, East Harlem was a historical Italian American community and by the 1920s, approximately 100,000 Southern Italians lived in Harlem. However, they were replaced by Puerto Rican immigrants, and the same location that was once called Italian Harlem is now called "Spanish Harlem *(El Barrio)*".

Figure 6: Tipping and the Dynamics of Neighborhood Segregation



Note: I reproduce the Figure III Rising Demand Leads to a Tipping Point from Card, Mas and Rothstein. (2008). This figure shows a series of equilibria for the neighborhood. Card, Mas and Rothstein (2008) notes that the location of tipping point depends on the strength of white/majority distaste for minority neighbors (i.e. $\frac{\partial b^w}{\partial m}$)

5 Reduced-Form Evidence

5.1 Event-Study Difference-in-Differences Specification

I investigate the effects of intra-city transit infrastructure on the geographic distribution of economic activity. Here, I follow the baseline specification from Heblich, Redding and Sturm (2020). I provide various aspects of neighborhood changes relative to the railway network expansion using a difference-in-differences specification. I use my spatially disaggregated neighborhood-level data for New York from 1870 to 1940. The major identification challenge in estimating the intracity regression is non-random assignment of transit infrastructure. Due to this non-random assignment of transit infrastructure, ordinary least squares (OLS) regressions that compare the treated and untreated locations in terms of transport infrastructure are unlikely to yield the unbiased estimates that capture the causal effects of the transport improvement investment (Redding and Turner (2015), Heblich, Redding and Sturm (2020), Baum-Snow (2007)). I address this identification challenge by including a neighborhood fixed effect, neighborhood time-trend, and the timing of railway network expansion.

Through this specification, I examine the changes in population growth, land-use specialization (commercial and residential), and segregation as it relates to the arrival of transit infrastructure and zoning. The baseline specification is as follows:

$$\log Y_{it} = \alpha_i + \sum_{\tau=-40}^{\tau=40} \beta_\tau \left(\mathbb{S}_i \times \mathbb{I}_{i\tau}\right) + \left(\mu_i \times Year_t\right) + d_t + \epsilon_{it} \tag{2}$$

Let t index the census year, and $i \in N$ index a set of neighborhoods in the city. Let Y_{it} denote an outcome of interest for neighborhood i at time t including population density, house price index, total area of new construction of buildings with different land-uses (i.e. residential/commercial), and measures of segregation; α_i is a neighborhood fixed effect; $\mathbb{S}_i$ is an indicator variable that equals 1 if a neighborhood i has an overground or underground railway station in at least one census year during the study period; τ is a treatment year indicator, which equals the census year minus the last census year that a neighborhood had no railway connection; $\mathbb{I}_{i\tau}$ being an indicator variable that equals 1 in treatment τ in neighborhood i and 0 otherwise; $Year_t$ is a census-year trend; μ_i is a neighborhood-specific coefficient on this census-year trend; d_t being a census-year dummy variable; and finally, ϵ_{it} is an unobserved heterogeneity in neighborhood i at the census year t. In my baseline specification, I cluster the standard errors on boroughs.

This specification allows me to control for certain neighborhoods to have higher level of outcome (Y_{it}) in all years through neighborhood fixed effects. The census-year dummies (d_t) controls for time-trends that apply to all neighborhoods (e.g. population changes in all neighborhoods in certain census year(s)). The neighborhood-specific census-year trends ($\mu_i \times Year_t$) allow neighborhoods treated with transit infrastructure investment and/or zoning to have higher outcomes of interest (Y_{it}) than other neighborhoods in all years. Finally, the key coefficients of interests (β_τ) are the interaction terms between the transit network indicator ($\mathbb{S}_i$) and treatment year indicator ($\mathbb{I}_{i\tau}$) that capture the treatment effect of the railway arrival on outcome (Y_{it}) in neighborhood (i) in treatment year (τ). They (β_τ) capture the deviations from the neighborhood-specific census-year trends where the first difference captures treated and untreated neighborhoods, whereas the second difference yields the timing (before and after) the railway network.

$$\log Y_{it} = \alpha_i + \sum_{\tau=-40}^{\tau=40} \beta_\tau \left(\mathbb{S}_i \times \mathbb{I}_{i\tau}\right) + \sum_{\tau=-40}^{\tau=40} \gamma_\tau \left(\mathbb{S}_i \times \mathbb{I}_{i\tau} \times \mathbb{I}_i^{center}\right) + \left(\mu_i \times Year_t\right) + d_t + \epsilon_{it} \tag{3}$$

However, another key prediction of the economic mechanism of specialization is that the treatment effect of the railway network should be heterogeneous, depending on whether neighborhoods are located in the city core or city periphery (Heblich, Redding and Sturm (2020)). Therefore, beyond the baseline specification where the coefficients of interest (β_τ) (equation 2) capture a common average treatment effect for the entire New York, I also estimate the heterogeneous treatment effects of the railway network (γ_τ from equation 3). This heterogeneity comes from whether neighborhoods are located in the city core, or city periphery. γ_τ from equation 3 separately estimates the heterogeneous treatment effect of the railway network on residential population, depending on neighborhoods being located in "central" locations or not.

5.2 Population Decentralization

I show that the population decentralized in central locations in New York, and the central locations themselves became more specialized in production activities (measured by the construction of new buildings for commercial purposes). Relatedly, I document the population gains in suburbs as the railway network expands. In Section D.1, I report difference-in-differences event-study specifications (i.e. a "dynamic" diff-in-diff) that establish the connection between the transit network and the spatial distribution of population. Figure 7 also shows the estimated treatment effects (β_τ) from equation 2 and the 95% confidence intervals, with the full regression results in Appendix Table 1. I find positive and significant deviations in log population from the neighborhood-specific time trends immediately after the railway arrival, and no evidence of significant deviations from these trends before the railway arrival. This results also shows that the population growth occured in close connection with the transit network, and the neighborhood fixed effects and time-trends largely control for the nonrandom assignment of the transit network (Redding and Turner (2015); Heblich, Redding and Sturm (2020); Duranton and Turner (2012)).

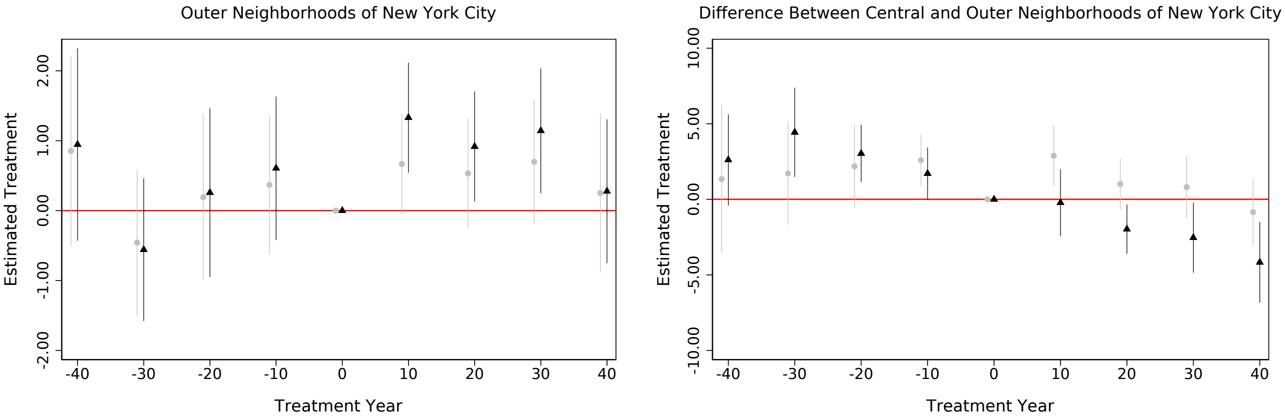
Next, I implement a regression specification to allow for heterogeneous treatment effects, as population trends between neighborhoods in the city core and the city periphery could be distinctively different. Equation 3 now estimates the heterogeneous treatment effects (γ_τ). $\mathbb{I}_i^{center}$ in an indicator variable that takes a value of 1 for neighborhoods located in the city core, and 0 elsewhere. This way, when neighborhoods in the city core received the railway treatment, its treatment effect would be captured by ($\beta_\tau + \gamma_\tau$) whereas neighborhoods in the city periphery would have railway treatment effect as (β_τ).

Finally, I also discuss how I define "city core" (where $\mathbb{I}_i^{center}$ takes a value of 1). One generic definition of "city core" is based on the county boundary of New York (i.e. "Manhattan"), whereas another definition of "city core" revolves around the transit network and central business district where jobs were highly concentrated as in Figure 3. For the latter definition, I take centroids of each neighborhood within 5 kilometers of the Battery, formerly known as Battery Park which is located at the southern tip of Manhattan Island in New York City. Neighborhoods within 5 kilometers of the Battery include the Financial District, the Midtown, and downtown Brooklyn where jobs were highly concentrated throughout my study period (See Section 3.3 regarding job and industry locations in the city during my study period. I show that all major industries in New York City covering 72% of all of plants and 79.5% of all of employees listed by factory inspectors for the city were largely concentrated in areas within 5 kilometers of the Battery).

As in Figure 7, , the heterogeneous treatment effects (γ_τ from equation 4) between neighborhoods in the city core and city periphery reveal that central locations (especially, the latter definition of the "city core" where neighborhoods are located within 5km distance from the Battery) have pre- and post-railway arrival population trends that are quite distinctively different. In particular, before the railway arrival, these central locations were the places where residential population was highly concentrated. However, after the railway arrival, I find negative and significant deviations in log population from the neighborhood-specific time trends shortly after the railway arrival which is an exact antithesis from the trends before the railway arrival.

$$\log \text{Population}_{it} = \alpha_i + \sum_{\tau=-40}^{\tau=40} \beta_\tau \left(\mathbb{S}_i \times \mathbb{I}_{i\tau}\right) + \sum_{\tau=-40}^{\tau=40} \gamma_\tau \left(\mathbb{S}_i \times \mathbb{I}_{i\tau} \times \mathbb{I}_i^{center}\right) + \left(\mu_i \times Year_t\right) + d_t + \epsilon_{it} \quad (4)$$

Figure 7: Population Changes



Note: Estimated treatment effects fromthe railway station arrival on log neighborhood population; the sample includes 195 time-consistent neighborhoods in New York City (made up of 5 boroughs) and comes from population censuses from 1870 to 1940 (every 10 years, except for 1890 where the original census was lost due to fire); all specifications include neighborhood fixed effects, year fixed effects, and neighborhood-specific time trends (the estimated coefficients and standard errors are reported in Appendix Table 1).

The left figure shows estimated treatment effects (β_τ from equation 4) for neighborhoods outside the city core; the right figure shows the heterogeneous treatment effects (γ_τ from equation 4) between neighborhoods in the city core and city periphery. The vertical lines in both left and right figure show the estimated 95% confidence intervals. Standard errors are clustered on boroughs.
IAs seen in the left and right figures, the first specification estimates equation 4 (estimates are illustrated in gray circle) by using "Manhattan" as the definition of city-core, whereas the second specification estimates equation 4 (estimates are illustrated in black triangle) by using "5km from the Battery" as the definition of the city-core.

5.3 Segregation

Stratification of neighborhoods by income and race

I show that there is a greater degree of residential "sorting" by income and race after the railway arrival. I measure the mean earnings of residents who were aged between 16 and 60 and in the labor force in the city during the study period. To enhance comparability, I take time-invariant neighborhood boundaries to consistently measure the mean earnings (based on one's primary occupation) and racial composition of residents by neighborhoods over time.

There are several measures of segregation. In this analysis, I implement **the index of dissimilarity** that measures the degree of unevenness in terms of the spatial distribution of different groups. The dissimilarity index for two groups in the city, White and African American, for example, is measured as follows (Massey and Denton (1998)):

$$\text{Dissimilarity Index}_{it} = \frac{1}{2}\sum_{i=1}^{n}\left|\frac{w_i}{W_T} - \frac{b_i}{B_T}\right|$$

where n equals the number of neighborhoods within the city; w_i denotes the number of white population in neighborhood i; W_T denotes the total white population in the city; b_i denotes the number of African American population in neighborhood i; B_T denotes the total African American population in the city.

O'Flaherty (2005) explains that the index of dissimilarity starts with the idea of perfect integration, or no segregation. If a city is perfectly integrated, then every neighborhood has the same proportion of minorities as every other neighborhood. The index of dissimilarity measures the distance of a city from perfect integration by figuring the size of the "shake-up" that would be needed to achieve perfect integration; the bigger the shake-up needed, the greater the degree of segregation.

In Section D.2, I report difference-in-differences event-study specifications that establish the connection between the transit network and neighborhood stratification (by income and race). Figure 8 shows the estimated treatment effects (β_τ) from equation 5 and the 95% confidence intervals, with the full regression results in Appendix Table 2. I find positive and significant deviations in log dissimilarity index from the neighborhood-specific time trends immediately after the railway arrival, and no evidence of significant deviations from these trends before the railway arrival. This result also shows that the stratification of neighborhoods occur in close connection with the transit network, and the neighborhood fixed effects and time-trends largely controls for the nonrandom assignment of transit network (Redding and Turner (2015); Duranton and Turner (2012)).

$$\log \text{Dissimilarity Index}_{it} = \alpha_i + \sum_{\tau=-40}^{\tau=40} \beta_\tau \left(\mathbb{S}_i \times \mathbb{I}_{i\tau}\right) + \sum_{\tau=-40}^{\tau=40} \gamma_\tau \left(\mathbb{S}_i \times \mathbb{I}_{i\tau} \times \mathbb{I}_i^{center}\right) + (\mu_i \times Year_t) + d_t + \epsilon_{it} \tag{5}$$

The regression specification (equation 5) also allows for heterogeneous treatment effects, as segregation trends between neighborhoods in city core and city periphery could be distinctively different. As in equation 4, $\mathbb{I}_i^{center}$ in an indicator variable that takes a value of 1 for neighborhoods located in the city core, and 0 elsewhere.[28] This way, when neighborhoods in city core received the railway treatment, its treatment effect would be captured by $(\beta_\tau + \gamma_\tau)$ whereas neighborhoods in city periphery would have their railway treatment effect captured by (β_τ).

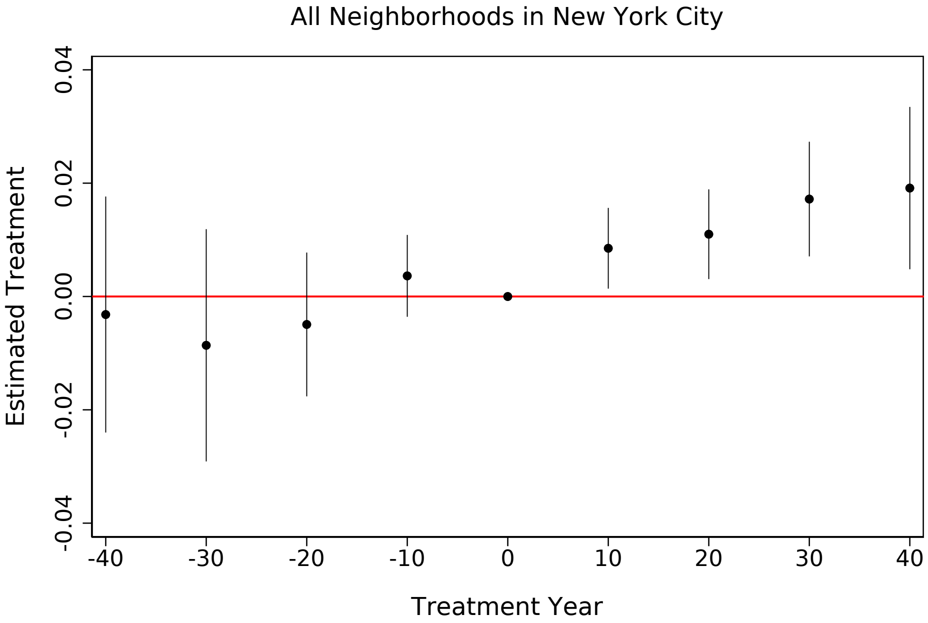
Unlike population trends where heterogeneous treatment effects (γ_τ) from the railway arrival were positive and significant, segregation trends do not exhibit heterogeneous treatment effects from the railway arrival (the full regression results in Appendix Table 2). This shows that segregation trends in city core were *not* quite distinctively different from the pre- and post-railway arrival.

As heterogeneous treatment effects of railway arrival on residential segregation (γ_τ) is mostly statistically insignificant, I also report difference-in-differences event-study specifications that establish the connection between the transit network and zoning as specified in equation 6. Specifically, I show that people with different levels of income and race sort into different neighborhoods when neighborhoods became zoned as either residential or business. In residence districts, only new buildings for residential purposes (such as homes, apartments, hotels) could be built, and all other business and manufacturing activities were prohibited. And the 1916 zoning designated about two-fifths of Manhattan and about two-thirds of the city as residential areas (See Figure 23 for details; different colors indicate different types of land-use designations under 1916 NYC Zoning Ordinances). Buildings in business-zoned districts could house all forms of business and industry *except* those generating objectionable odors and by-products—like asphalt, paint, soap manufacturing, metal or stone works, or anything involving ammonia or sulfuric acid. The ordinances allowed other types of manufacturing in business districts but confined them to 25 percent of the total floor space of the building. Finally, in unrestricted areas, buildings could be used for residential or business and industrial purposes.

---

[28] I define the city center in two ways as before —one being "Manhattan" or not; and the other being "within 5km from the Battery".

Figure 8: Estimated Treatment Effects (β_τ) of Railway Arrival on Neighborhood Segregation



Note: Estimated treatment effects (β_τ from equation 5) from railway station arrival on log neighborhood dissimilarity index; the sample includes 195 time-consistent neighborhoods in New York City (made up with 5 boroughs) and comes from population censuses from 1870 to 1940 (every 10 years, except for 1890 where the original census was lost due to fire); all specifications include neighborhood fixed effects, year fixed effects, and neighborhood-specific time trends; the estimated coefficients and standard errors are reported in Appendix Table 2.
In the figure above, the horizontal axis shows the treatment year τ, and τ is defined as census year minus the last census year in which the neighborhood did not have railway network. Therefore, positive values of τ capture the post-treatment years, whereas negative values of τcapture the pre-treatment years. The figure shows estimated treatment effects (β_τ from equation 5) for neighborhoods. The vertical lines in the figure show the estimated 95% confidence intervals. Standard errors are clustered on boroughs.

The full regression results in Appendix Table 2 reveal that when neighborhoods became zoned as residence districts, the dissimilarity index increases about 2.6%. Note that among business zoning, a subcategory called "Business District Restricted Against Industry", where all types of manufacturing are banned, is associated with a higher degree of segregation relative to "business district" that *does* allow some types of manufacturing: up to 25 percent of the total floor space of the building.

The height restrictions tend to be positively correlated with neighborhood segregation. For example, a height restriction of "1.25" established a relationship between street width and building height: the building could rise up to 1.25 times the width of the street.[29] Lower restrictions such as 1 or 1.25 were applied to large sections of city periphery. Figure 24 shows that most underdeveloped

---

[29] After that, the building could rise an additional 2 feet for every one foot from the street that it was set back.

sections of the city were zoned for either 1 or 1.25, indicating that the city planners of the zoning ordinances did not want other areas to be built up as densely.[30]

Finally, the ordinances specified area districts which prevented new buildings from covering their entire sites, mandating open spaces at the rear and sides of the structure —the taller the building, the more space required on all sides (Makielski (1966); Ward and Zunz (1992)). These restrictions resulted in the construction of rather small structures. For example, Figure 26(a) shows the type of structures that arose when Area restrictions were coupled with Land-Use regulation (especially, "Residence" districts). If certain area restrictions were implemented in residence districts (as in Upper Left and Upper Right of Figure 26(a)), then these resulted in small structures suited for upscale residential sections of the city (Makielski (1966)).

Relatedly, a regression that estimates equation 6 reveals that once I account for all types of land-use regulations, the estimated treatment effects (β_τ) from the railway arrival becomes positive and significant only for long-term treatment periods. This establishes the tighter connection between land-use regulation and residential sorting by income and race.

$$\log \text{Dissimilarity Index}_{it} = \alpha_i + \sum_{\tau=-40}^{\tau=40} \beta_\tau \left(\mathbb{S}_i \times \mathbb{I}_{i\tau}\right) + \left(\mu_i \times Year_t\right) + d_t + \epsilon_{it} \tag{6}$$

"Isolation" of minorities

The other measure of segregation is called the **index of "isolation**". This index equals the proportion of minorities who live in the neighborhood that the average minority group member lives in; that is, the probability that a neighbor of a minority group member is also a member of that minority group. With perfect integration, this index is the proportion of that minority in the whole city, and with perfect segregation, the index equals one (as every minority group member lives in an all-minority neighborhood) (O'Flaherty (2005)). The isolation index for two groups in the city, White and African American, for example, is measured as follows (Cutler, Glaeser and Vigdor (1999)):

$$\text{Index of Isolation} = \frac{\sum_{i=1}^{n}\left(\frac{black_i}{black_{total}} \cdot \frac{black_i}{person_i}\right) - \left(\frac{black_{total}}{persons_{total}}\right)}{min\left(\frac{black_{total}}{person_i}, 1\right) - \left(\frac{black_{total}}{persons_{total}}\right)}$$

where $\sum_{i=1}^{n}\left(\frac{black_i}{black_{total}} \cdot \frac{black_i}{person_i}\right)$ is the percentage black of the area occupied by the average black, and $person_i$ refers to the total population of tract i; Cutler, Glaeser and Vigdor (1999).

---

[30] My spatial analysis confirms that Brook and Rose (2013)'s claim of "zoning ordinances, in practice, turned out to have served considerably more mundane goals: reserving a preexisting status quo in already-developed areas, and defending the single family home from the "lower" uses of multiple dwellings, commerce, and manufacturing in that descending order" holds true.

$$\log \text{Isolation Index}_{it} = \alpha_i + \sum_{\tau=-40}^{\tau=40} \beta_\tau \left(\mathbb{S}_i \times \mathbb{I}_{i\tau}\right) + \left(\mu_i \times Year_t\right) + d_t + \epsilon_{it} \tag{7}$$

In Section D.2, I report difference-in-differences event-study specifications that establish the connection between the transit network and isolation of African Americans (with the full regression results in Appendix Table 4). Unlike residential sorting (captured by dissimilarity index), I do not find significant deviations in log isolation index from the neighborhood-specific time trends immediately after the railway arrival, and no evidence of significant deviations from these trends before the railway arrival either.

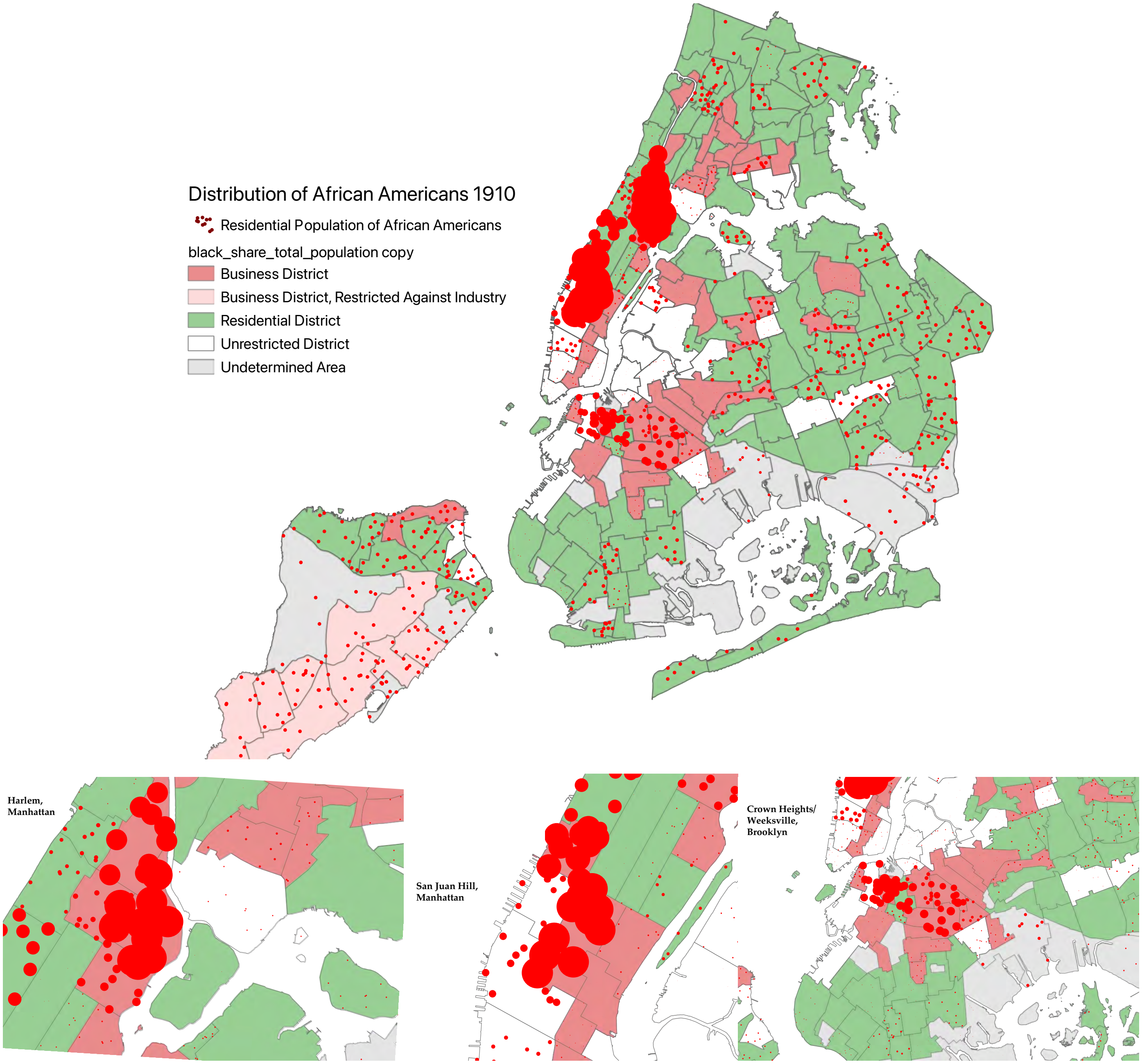
Instead, the results establish a tight connection between zoning (especially, ‘‘residence district’’ land-use regulation) and segregation of minorities. In other words, land use regulations such as Residence District and Business District Restricted Against Industry (which does not allow any manufacturing activities) makes African Americans especially isolated in neighborhoods where manufacturing and industrial activities were allowed. Relatedly, ‘‘Area’’ regulations that prevented new buildings from covering their entire sites lead to the construction of a more ‘‘upscale’’ housing structure, therefore resulting in more isolated African Americans. As mentioned above, certain area restrictions coupled with residence districts (as in Upper Left and Upper Right of Figure 26(a)) resulted in small structures suited for upscale residential sections of the city. Regression results support that creation of such ‘‘upscale’’ structures may have induced certain African Americans to disproportionately live in few minority neighborhoods in ‘‘isolation.’’

Figure 9 depicts the spatial distribution of African Americans in 1910 before zoning (i.e. six years before zoning was implemented). The salient features from Figure 9 are as follows: 1. Even before zoning, African Americans lived in fairly large numbers (captured by the size of red bubbles) in a smaller number of neighborhoods; 2. Neighborhoods where African Americans resided in high numbers were disproportionately more likely to be zoned as Business districts (in pink shade) where manufacturing and industrial activities *were* allowed. Neighborhoods where African Americans lived in very small numbers (presented by small or non-existing red bubbles in the figures above) were almost always zoned for residence districts where any type of manufacturing or industrial activities was strictly prohibited.

Figure 10 now depicts the spatial distribution of African Americans in 1930 after zoning. The salient features from Figure 9 are as follows: 1. By 1930, almost no African Americans lived in white neighborhoods; 2. In 1910 (before zoning), there still were African Americans living in mostly white neighborhoods (e.g. see neighborhoods in State Island or Queens where we still see scattered ‘‘dots’’), but by 1930 (post-zoning) almost no African Americans live in non-African American neighborhoods; In other words, exposure of the African American population toward the majority population dropped (and therefore the ‘‘isolation’’ index of minorities increases); 3. Neighborhoods where African-Americans resided in high numbers were almost always zoned as Business districts (in pink shade) where manufacturing and industrial activities *were* allowed. Considering that two thirds of outerboroughs were zoned for residence districts, it is hard to deny that there exists a link between the composition of minority residents and the different types of zonings and consequences that came with them.

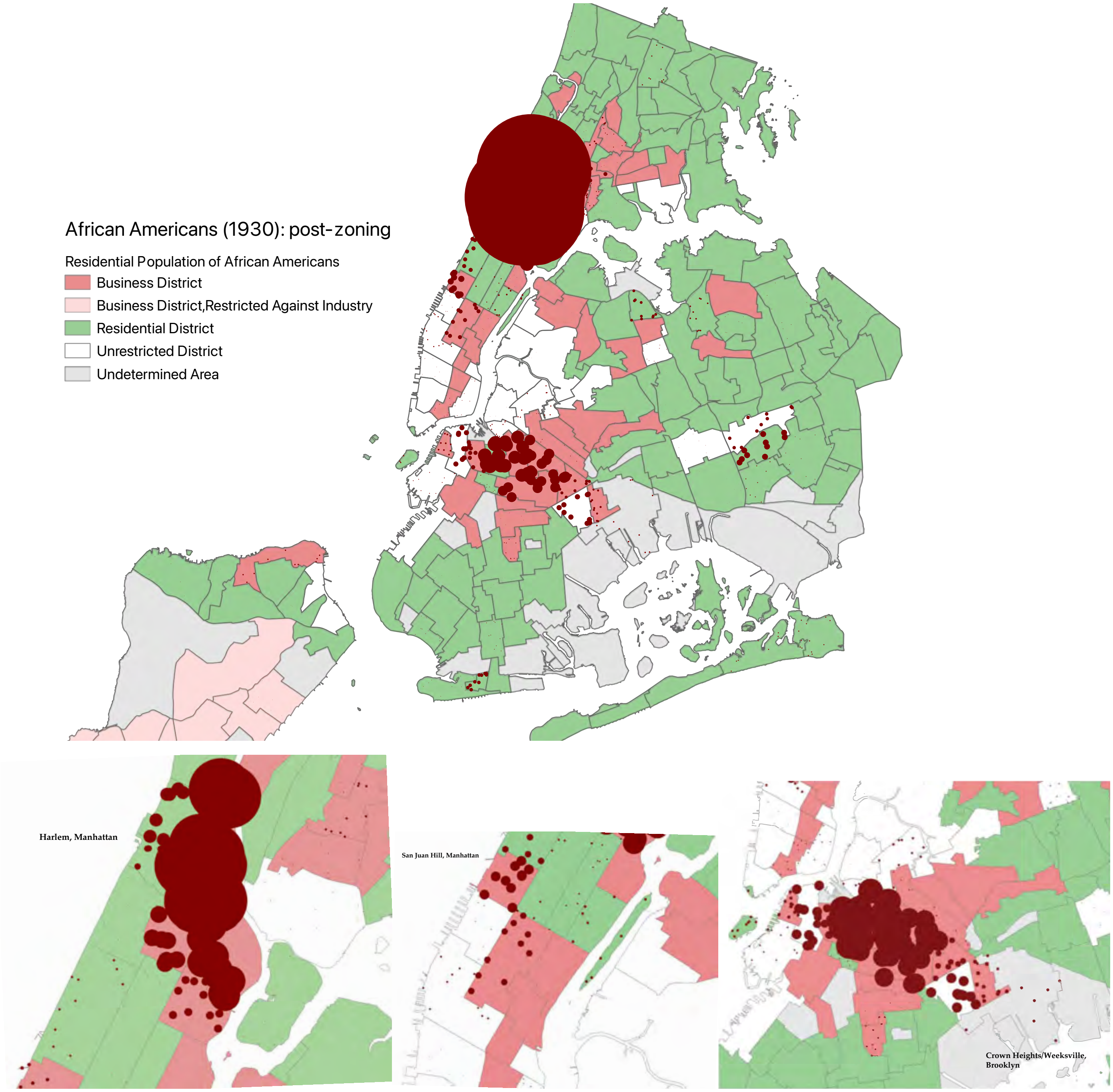
These findings themselves carry three major public policy implications. First, it shows that racially-discriminatory policies were underway well before ‘‘redlining’’–the point that there were extremely segregated neighborhoods where African Americans were ‘‘isolated’’ (surrounded only by its own minority population) well before redlining holds true. Second, there could be a bigger welfare disparity between racial groups in terms of neighborhood amenities (e.g. noise, clean air) and long-term health consequences such as long-term exposure to pollutants or hazardous materials (See, for example, Heblich, Trew and Zylberberg (2021).). Third, in minority neighborhoods, the wealth accumulation channel of home ownership was practically impossible, which I discuss in detail in Section 6.

Figure 9: Spatial Distribution of African Americans (1910) : pre-zoning (1916)



Note: The above figures show the spatial distribution of African Americans in 1910 (6 years before 1916 Zoning Ordinances); Bigger red bubbles indicate a higher number of African Americans. Areas where African-Americans resided in high numbers (presented by big red bubbles in the figures above) were disproportionately more likely to be zoned as Business districts (in pink shade) where manufacturing activities were allowed. Neighborhoods where African Americans lived in very small numbers (presented by small or non-existing red bubbles in the figures above) were zoned for residence districts where any type of manufacturing or industrial activities was strictly prohibited.

Figure 10: Spatial Distribution of African Americans (1930) : post-zoning (1916)



Note: The above figures show the spatial distribution of African Americans in 1930 (14 years after 1916 Zoning Ordinances); Bigger red bubbles indicate a higher number of African Americans. One can see that the spatial ‘‘concentration’’ of minorities have become even more severe.

5.4 Housing Prices Change

Hedonic House Sales Index

From the digitized real estate conveyance records, I document how neighborhood-level housing prices had evolved in relation to transit access and zoning over a long-time horizon (See Section C.2 for details on hedonic price index construction and descriptions of the digitized transaction-level house data for 70 years). I show that after the railway arrival, house prices increased substantially (about 50%) for all neighborhoods in New York. I also show that house prices (measured by the hedonic house sales index) in non-central locations dramatically increased after the railway arrival.

In Section D.1, I report difference-in-differences event-study specifications that establish the connection between the transit network and house values. Figure 11 also shows the estimated treatment effects (β_τ) from equation 8 and the 95% confidence intervals, with the full regression results in Appendix Table 5. I find positive and significant deviations in log house sales price index from the neighborhood-specific time trends immediately after the railway arrival, and mostly no significant deviations from these trends before the railway arrival.

$$\log \text{House Price Index}_{it} = \alpha_i + \sum_{\tau=-50}^{\tau=50} \beta_\tau \left(\mathbb{S}_i \times \mathbb{I}_{i\tau}\right) + \left(\mu_i \times Year_t\right) + d_t + \epsilon_{it} \tag{8}$$

As in Section 5.2 and 5.3, I implement a regression specification to allow for heterogeneous treatment effects of transit infrastructure, as house price trends between neighborhoods in city core and city periphery could be distinctively different. Same as before, I define ‘‘city core'' (where $\mathbb{I}_i^{center}$ takes a value of 1)—1. a generic definition of ‘‘city core'' is ‘‘Manhattan''; 2. a Central Business District-oriented definition is locations within 5 kilometers of the Battery.[31]

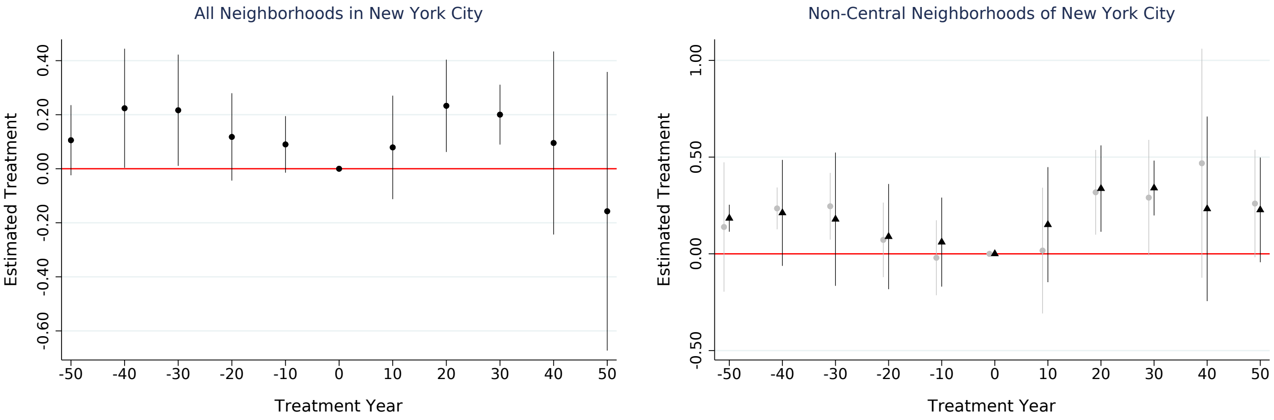
$$\log \text{House Price Index}_{it} = \alpha_i + \sum_{\tau=-50}^{\tau=50} \beta_\tau \left(\mathbb{S}_i \times \mathbb{I}_{i\tau}\right) + \sum_{\tau=-50}^{\tau=50} \gamma_\tau \left(\mathbb{S}_i \times \mathbb{I}_{i\tau} \times \mathbb{I}_i^{center}\right) + \left(\mu_i \times Year_t\right) + d_t + \epsilon_{it} \tag{9}$$

As in Figure 27, the heterogeneous treatment effects (γ_τ from equation 9) between neighborhoods in the city core and city periphery reveal that central locations (especially with Manhattan being the ‘‘city core'') have house price trends that are quite distinctively different in terms of pre-railway (treatment) period. Digitized housing price data reveals that central (i.e. ‘‘Manhattan'') locations had higher growth in housing prices. However, after the railway arrival, I no longer see significant deviations in log house price from Manhattan neighborhoods. Relatedly, before the railway arrival, high house price growth is observed for locations within 5 kilometer distance from the Battery but most of them are statistically insignificant. However, after the railway arrival, locations in CBD exhibits sharp price increase of house sales price.

---

[31] See Section 3.3 regarding job and industry locations in the city during my study period.

Figure 11: House Price Changes



Note: Estimated treatment effects from railway station arrival on log hedonic house sales price index (by neighborhood); the sample includes 195 time-consistent neighborhoods in New York City. The digitized house sales data, at the transaction-level with a daily frequency for 70 years with approximately 450,000 observations, were mapped to time-consistent boundaries through geocoordinates. All specifications include neighborhood fixed effects, year fixed effects, and neighborhood-specific time trends (the estimated coefficients and standard errors are reported in Appendix Table 5; 6).

The left figure shows estimated treatment effects (β_τ from equation 8) for all neighborhoods in the city. The vertical lines in both left and right figure show the estimated 95% confidence intervals. Standard errors are clustered on boroughs. In the right figure, the specification estimates equation 9 by using "Manhattan" as the definition of city-core (estimates are illustrated in gray circle), whereas the second specification estimates equation 9by using "5 km from the Battery" as the definition of city-core (estimates are illustrated in black triangle).

Housing Prices of "Tipped" Neighborhoods

As discussed in Section 4, Card, Mas and Rothstein (2008) notes that the value of the "tipping point" (m^*) depends on the white/majority population's tolerance for growing minority neighborhoods. If the majority's aversion towards minority is "strong" (captured by the sharp negative slope of $\frac{\partial b^w}{\partial m}$), then even at a relatively low level of minority share (m), the neighborhood could undergo the tipping process. Relatedly, Card, Mas and Rothstein (2008) discusses housing price predictions. Rents at the 100% minority equilibrium (i.e. $m = 1$) could be higher or lower than the tipping point, depending both on the minority demand function (curvature of b^m) and shifts in minority demand once the tipping is underway.

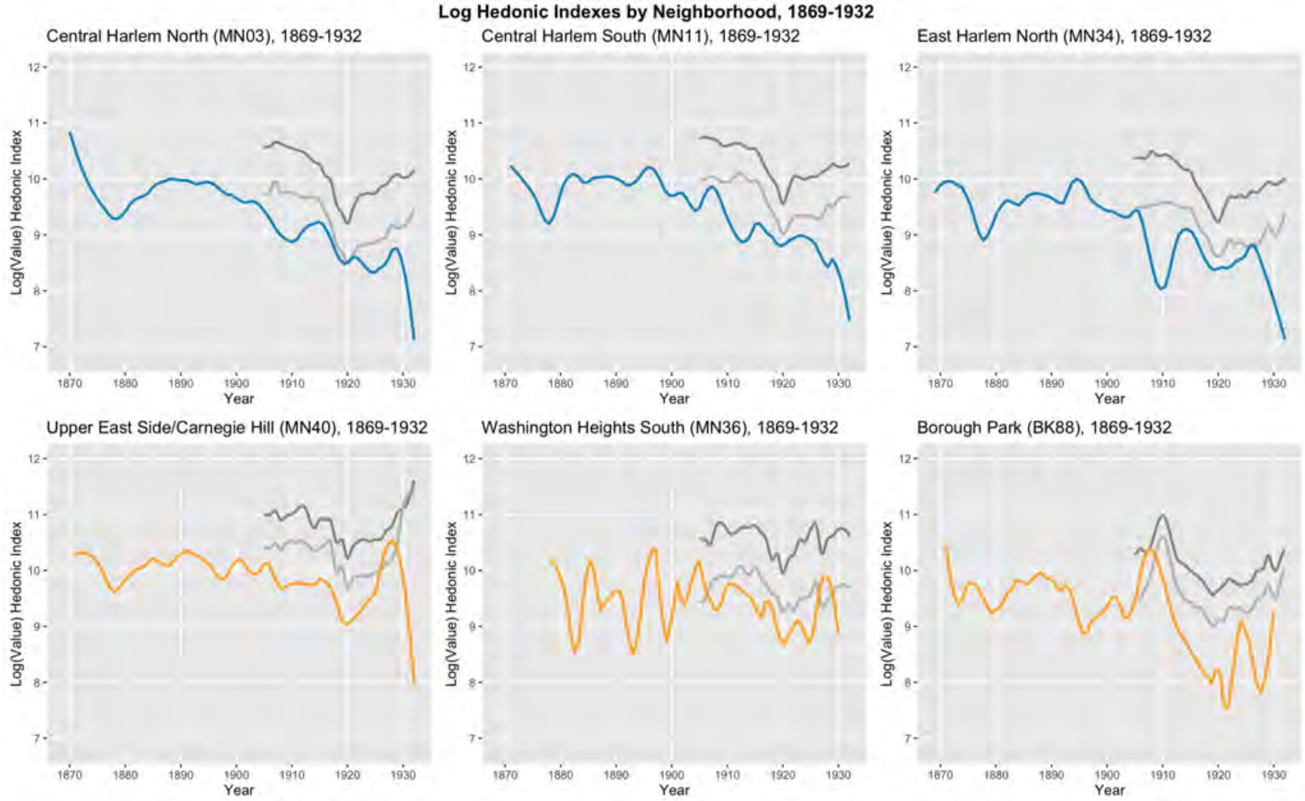
Figure 14 plots housing prices indexes by neighborhoods against the share of African American population in the neighborhood. Each dot represents a neighborhood, and the share of African American population was calculated from each federal population census (i.e. 1910, 1920, 1930). For housing prices, I plot each neighborhood's hedonic housing price index in years that correspond to the census year. Each dot represents a time-consistent neighborhood and only neighborhoods where the housing price index in those census years was available were plotted. There were primarily 3 neighborhoods in Harlem that underwent the "tipping process" during my study period

—the non-tipped neighborhoods are in grey circles, whereas ‘‘tipping’’ neighborhoods (i.e. all neighborhoods in Harlem) are in blue, green and orangecircles. One can follow the dynamic process of ‘‘tipping’’ by following the same-color dots over time in Figure 14. For example, considering Blue circles, in 1910, Central Harlem North had about a 7% African American population (this share was the highest among all other neighborhoods in 1910). Then, by 1920, the share of black population increased from 7% to 50% in just 10 years; by 1930, the same neighborhood reached an almost ‘‘perfectly’’ segregated equilibrium where approximately 90% of the population in the neighborhood was African American (although, African Americans still made up less than 5% of the city’s total population). We can see a similar tipping process in other neighborhoods in Harlem where the share of minorities was less than 5%, but in just one decade the share more than doubles.

Then, in Figure 12, I plot housing price movements of ‘‘tipped’’ neighborhoods in Harlem (in Blue line) over the period 1870-1940. The dark grey line is the hedonic price index based on assessed valuations of properties that include lot-only sales and building structures; whereas the light grey line is the hedonic assessed-valuation based price index based on real estate properties that are lot only; the blue line is the hedonic price index based on transacted real estate properties that include all types of real estate properties such as lot only, house and lot, or some other building structures such as apartments or tenements with stores. Although specific timing and trends may somewhat vary, hedonic price indexes in these ‘‘Tipping’’ neighborhoods in Harlem were decreasing quite dramatically, especially since 1920 (after 1916 NYC Zoning Ordinances and the huge influx of Black Migration during the Great Migration Era).

To gain an understanding of why housing prices were decreasing in ‘‘tipped’’ neighborhoods, I plot the trends of population changes in Figure 13. Figure 13 shows that for East Harlem between 1910 and 1920 when there was an inflow of 9000 African Americans, there was an outflow of 22,000 white population. In the same neighborhood between 1920 and 1930, an inflow of 23,000 African Americans is associated with an outflow of 57,000 white population. Taking these numbers in perspective, each black arrival led to 2.5 white departures during the a neighborhood’s ‘‘tipping’’ process. Boustan (2010) shows that the white population responded to black migration from the rural South by leaving cities; she finds that each black arrival led to 2.7 white departures. Although I am studying prewar suburbanization at the ‘‘neighborhood’’-level, the inflow and outflow of population dynamics is truly salient in a neighborhood’s dynamic segregation process.

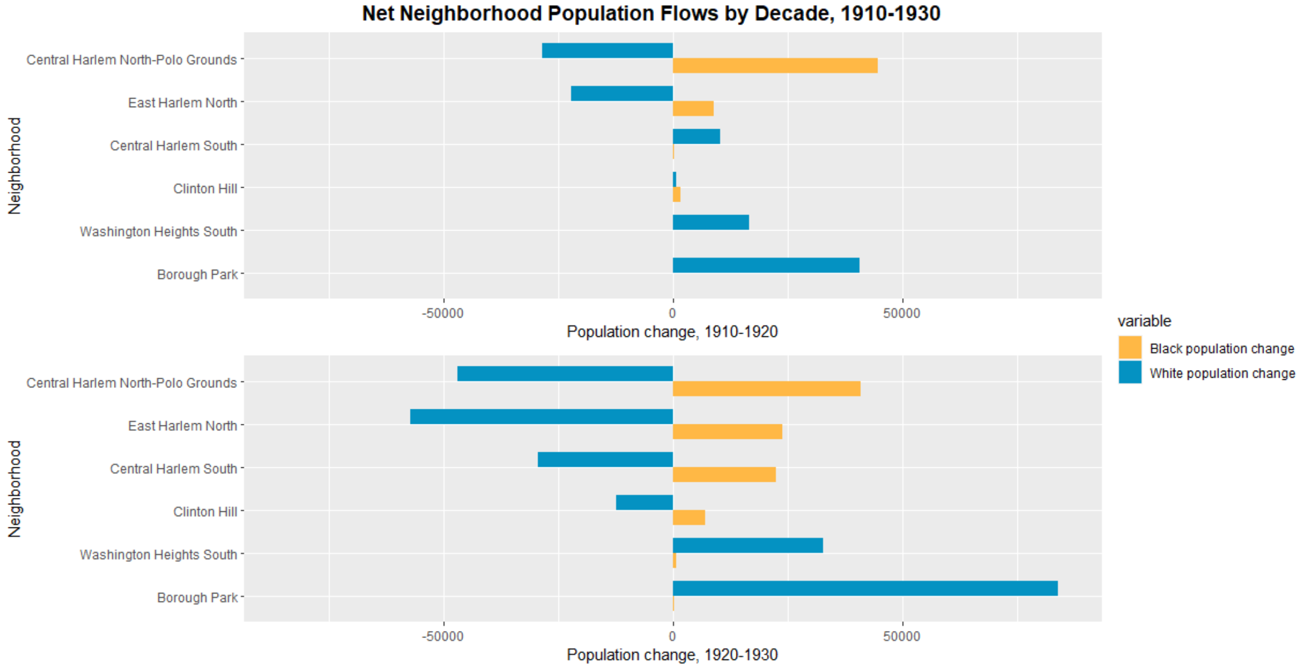
Figure 12: Housing Prices of “Tipped” Neighborhoods



Note: I plot hedonic house price indexes by neighborhood over 70 years. In all subfigures here, the dark grey line plots the hedonic assessed-valuations(maximum assessed values)-based price index where the underlying transaction-data include both lot-only properties and buildings; the light grey line plots the hedonic assessed-valuations(maximum assessed values)-based price index where the underlying transaction-data include lot-only properties (i.e. no structures on transacted land/lot).

The three subfigures on the first row (with blue lines) plot the housing prices for neighborhoods that underwent the dynamic process of residential racial segregation. The three subfigures on the second row (with orange lines) plot the housing prices for neighborhoods that stayed ‘‘all-white’’ throughout my study period. In understanding housing price dynamics during the dynamic process of residential racial segregation, the hedonic price indexes between the period 1910 and 1930 are extremely important. Compared to their pre-trends before 1910 and fundamental values ‘‘measured by the assessed values’’ (in grey lines), housing prices in ‘‘tipping’’ neighborhoods were declining.

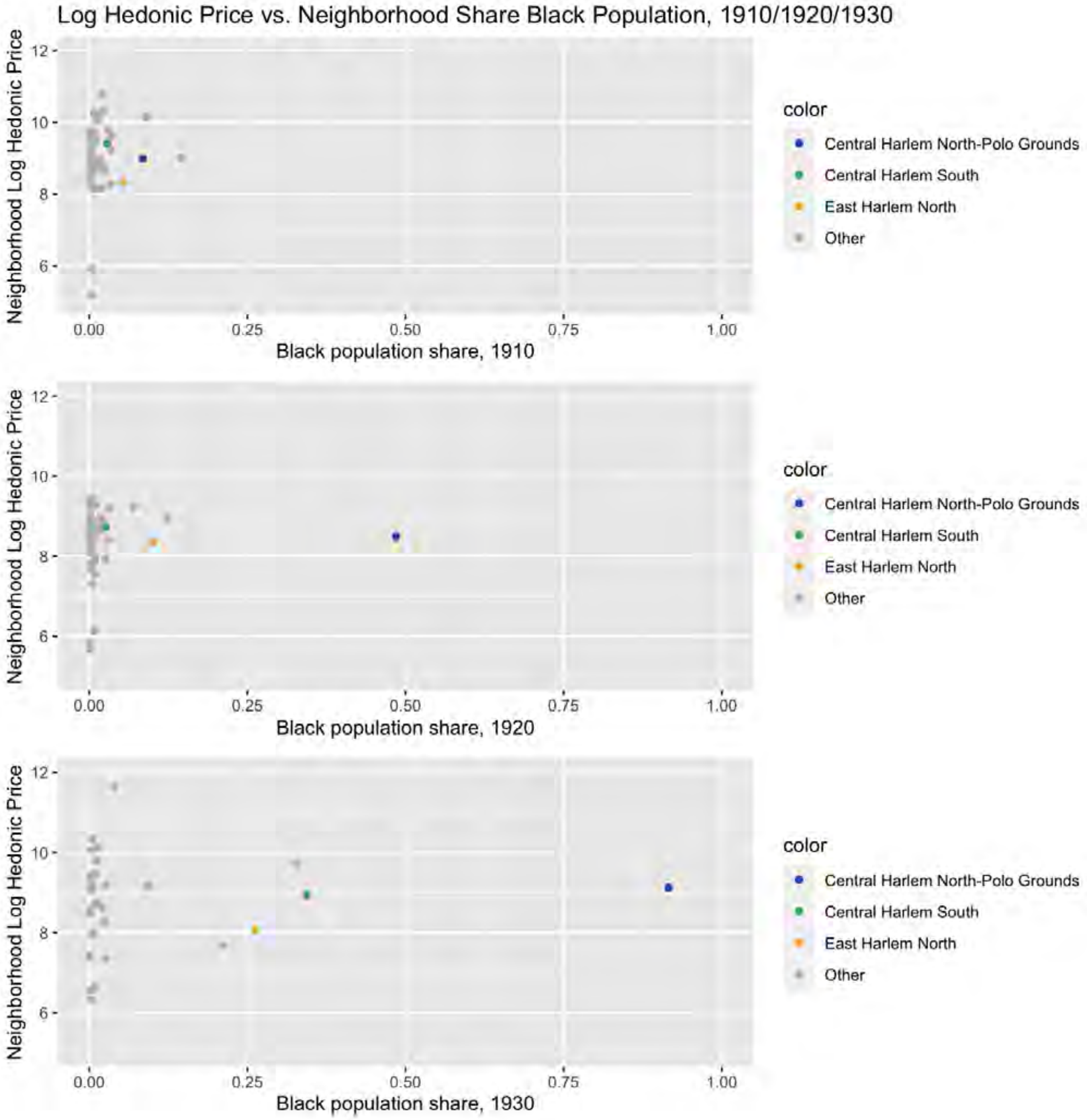
Figure 13: Population Inflow and Outflow of “Tipped” and other Neighborhoods



Note: I plot net neighborhood population flows by decade for neighborhoods that “tipped” (Central Harlem North and South, and East Harlem) between 1910 and 1930 (1st-3rd row in the figure). Clinton Hill/San Juan Hill area (4th row) was an area with a high number of African Americans till 1910; however, with the rise of Midtown, the African American population declined between 1910 and 1930. Finally, I also plot some “non-white” neighborhoods that remained “100% white” throughout my study period.

Net population analyses reveals that arrival of one African American is associated with 2.5 white departures for “tipping” neighborhoods. Out of 196 neighborhoods in the city, only about 5 neighborhoods were going through this residential racial segregation, and all other “100% white” 190 neighborhoods were gaining population between 1910 and 1930.

Figure 14: Hedonic Housing Prices of "Tipped" and "Not-Tipped" Neighborhoods



6 Catching the Opportunities in the *Metropolis*

In this section, I look at how people of different classes and races lived in the city during a time of unprecedented urban transition and economic growth.[32] Boustan, Bunten and Hearey (2018)

---

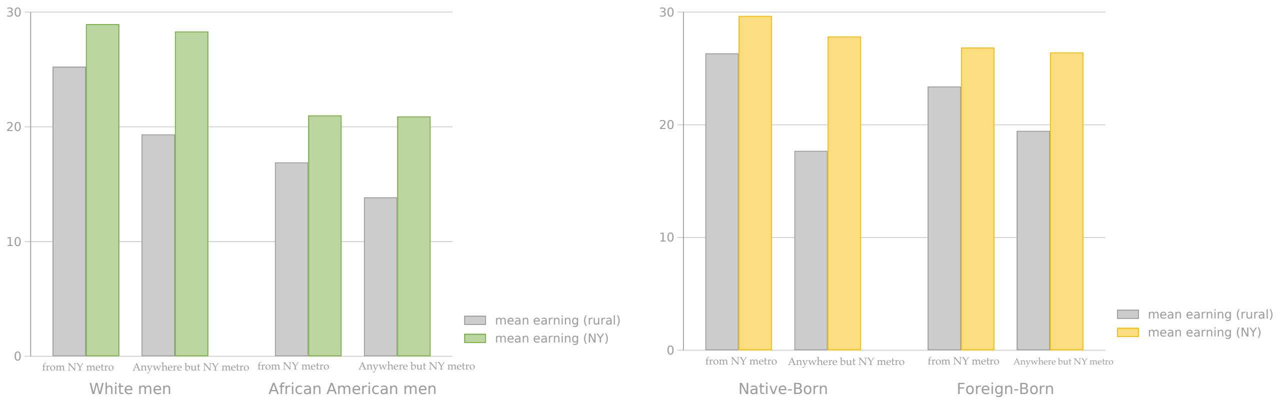
[32] This attempt is quite the opposite of Cronon (1991)'s work *Nature's Metropolis* that explores the economic changes that made Chicago one of America's biggest and most dynamic city. Cronon (1991) himself notes that *"there are almost no people in Nature's Metropolis. And almost no lived, textured reality of classed, gendered,*

documents that the urban wage premium in the US was remarkably stable over the past two centuries, ranging between 15 to 40 percent.[33] Relative to Boustan, Bunten and Hearey (2018), I use individual panel data to measure urban wage premium during this period of second industrial revolution in America's largest city. Especially, I investigate the migrants who moved from rural to urban areas (i.e. New York) and measure how their earnings have changed.[34]

6.1 How did migrants' lives change?

From 1870 to 1940, migrants from rural areas to New York enjoyed a substantial urban wage premium, resulting in about a 36% earnings increases for white men and a 40% earnings increase for African American men. Figure 15 reveals that migrants from rural areas in metro New York to New York City still had urban wage premium of 15% for both white and African American men; the magnitude of the urban wage premium was much higher for migrants from rural areas in the South than it was for migrants from some rural areas near New York. Although African Americans "fared" better in the big city than in the South, their mean earnings in New York were still significantly lower (about 32%) than those of white men.

Figure 15: Income Changes: Pre- and Post- Urban Migration



Note:

Figure 16 reveals why African American men's urban wage premium was lower than that of

---

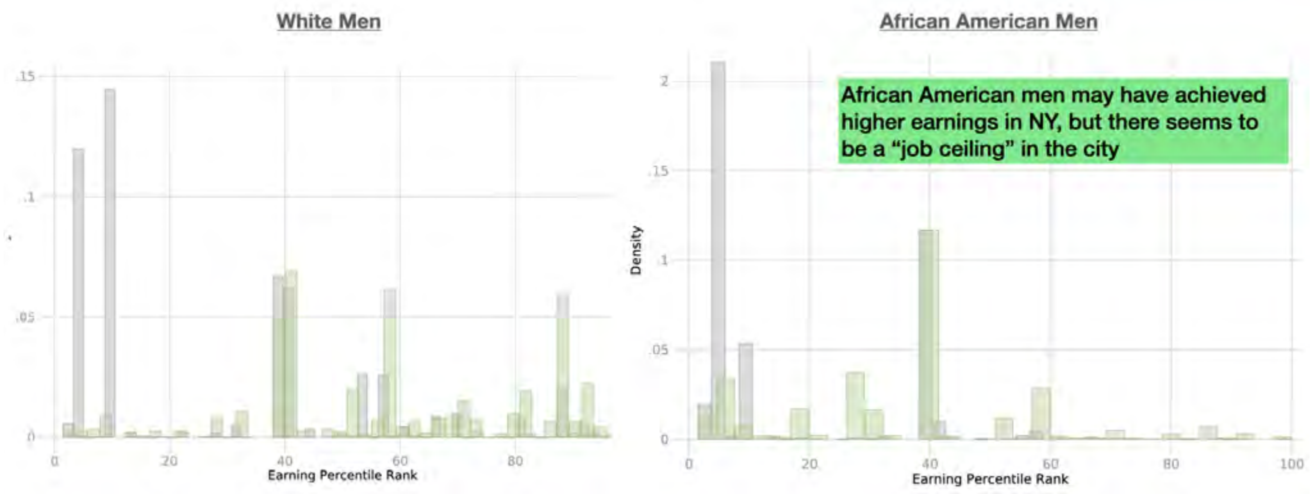
raced people.''

[33] Boustan, Bunten and Hearey (2018) show steadily rising wages in both urban and rural areas, and a *rising* urban wage premium, which increased from 17 percent to 37 percent. Through the Rosen-Roback framework, Boustan, Bunten and Hearey (2018) conclude that considering workers' continued migration to the cities during this period, the rising urban wage premium implies cities were becoming much more productive as production centers during the second industrial revolution.

[34] In measuring the urban wage premium, Boustan, Bunten and Hearey (2018) relies on wage data for three occupation categories (unskilled, blue collar and white collar) from Williamson and Lindert (2016). However, I use wage data from the Census of Manufactures and IPUMS (Ruggles, Flood, Goeken, Grover, Meyer, Pacas and Sobek (2019))'s occupation-based measures that follows the 1950 Census Bureau's occupation classification. See Appendix Section C.1 for details.

white men. The distribution of occupation-based earning percentile rank reveals the following: 1. white men who migrated from rural areas to New York transitioned from low-earning occupations to occupations that were mostly above the median earning percentile rank, 2. African American men who migrated from rural areas to New York also transitioned from low-earning occupations to occupations that are higher than before but mostly below median earning percentile rank.

Figure 16: Distribution of Earned Income Percentile Rank Changes of Rural to Urban Migrants: by Race



White Men

African American Men

African American men may have achieved higher earnings in NY, but there seems to be a "job ceiling" in the city

Note: The figure on the left plots the distribution of earning percentile rank of white migrants from rural areas to New York. Their prior occupation-based earnings distribution are depicted in grey, whereas theirearning distribution in New York is depicted in green. Higher concentration of grey bars in rural areas imply that their earnings were concentrated in low-earning occupations in rural areas; once they migrated to an urban locations like New York, their earning distribution transitions from mostly low-paying jobs to above-median-earning jobs.

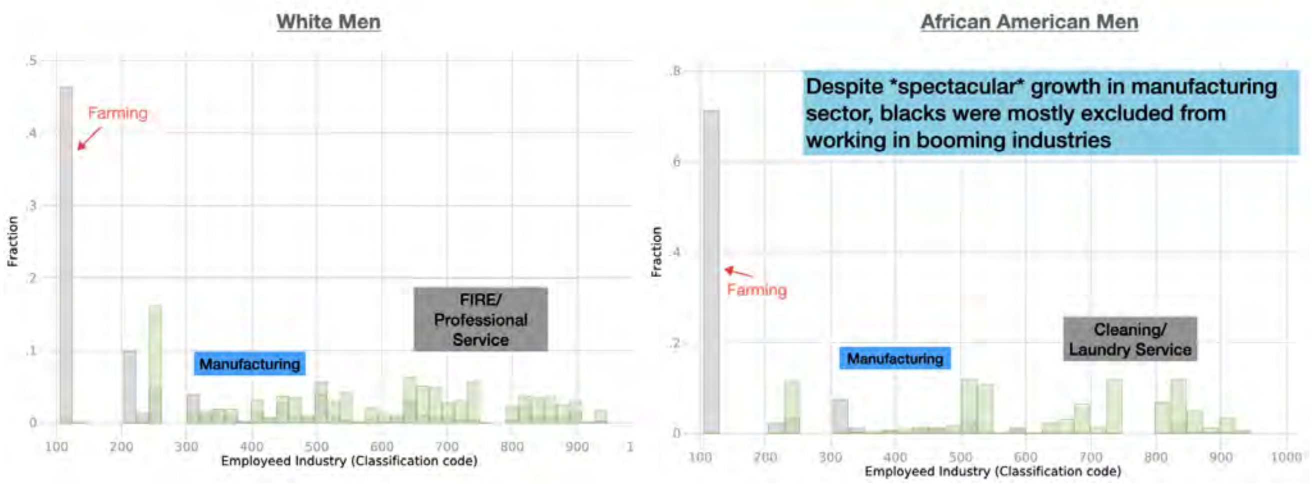
The figure on the right plots the same distribution of earning percentile rank for African Americans. Same as before, grey bars in rural areas imply that theirearnings were concentrated in low-earning occupations in rural areas. Once they migrated to an urban locations like New York, their earning distribution transitions from extremely low-jobs to predominantly less-than-median earning jobs. Therefore, still the urban wage premium was significant for African American migrants from the rural South to New York. However, as they rarely transitioned to above-median earning occupations, this implies that there was a "job ceiling" for African Americans.

Figure 17, derivedfrom panel data, reveals that both white and African American men were working primarily working in agriculture prior to their migration. However, once they arrived in New York locations, white men typically transitioned to high-growth industries including manufacturing, finance, insurance and real estate. However, such a industry transition did not materialize for African American migrants from the rural South to New York. They were systematically excluded from manufacturing industries and service industries that were considered "white collar." Instead, they were primarily engaged in personal services such as cleaning, laundry, running errands, and some types of transportation.

My individual panel data reveals that one's urban wage premium is largely explained by the industry that one works (the full regression results in Appendix Table 7). For example, when

I follow the same individuals' earning changes between two adjacent census years, people of all races who migrated from rural to urban locations enjoyed an increase of his earnings of 35-40%. A worker's upward mobility was explained by their migration decision to a large cities, and the diversity and type of "industries" (in rural areas, most of them are working in "Faming" industry) are what makes cities "the land of opportunities." In other words, the "job ceiling" for African Americans—where there are a prescribed set of trades or occupations, restricted exclusively to a group with ascribed social and economic status, that offer almost no upgrade—is rooted in the exclusion of African Americans from certain industries (as in Figure 16, 17).

Figure 17: Distribution of Employed Industry: Rural to Urban Migrants: by Race



Despite *spectacular* growth in manufacturing sector, blacks were mostly excluded from working in booming industries

Note: The figure on the left plots the distribution of the employed industry of white migrants from rural areas to New York. From the individual panel data, I show the distribution of one's primary employed industry in rural areas (pre-migration) in grey, and I depict the same for post-migration in New York in green. The figure on the left depicts the industry transition of white migrants from rural to New York, whereas the figure on the right depicts the same for African American migrants from the rural South to New York.

6.2 Could they have done better?

Among rural to urban migrants, I examine whether cities other than New York may have been a better "land of opportunities" for new migrants. Therefore, I calculate the mean earnings change of rural to urban migrants who migrated to urban cities other than New York. For example, Figure 18 plots the mean earnings across various cities for white and African American men who are both rural to urban migrants; Boston and Philadelphia (Upper Left and Upper Right in Figure 18) as migrants' urban destinations depict somewhat similar pattern to New York; urban wage premia for migrants to Boston and Philadelphia is stable at around 40% for both white and African American men; however, as in New York, African American men may face a similar "job ceiling" in these Norther cities as we do see that their mean earnings are significantly lower than those of white rural migrants to the same city.

Figure 18: Income Changes of Rural to Urban Migrants by Race

Figure 18 also reveals that Chicago (Bottom Left in Figure 18) may have been associated with ‘‘higher’’ level of urban wage premia. The urban wage premium among migrants to Chicago was higher for both white and African American men than it was in any other major cities; for example, the average urban wage premium was approximately 60% for white rural to Chicago migrants; whereas the urban wage premium was approximately 80% for African American rural to Chicago migrants. In understanding this bigger and substantial urban wage premium for African American migrants, one factor that is salient was that African Americans were employed in manufacturing industries in Chicago whereas we do not see such patterns in any other Northern (and Southern) cities up until 1930.

Atlanta (Bottom Right in Figure 18) depicts the most poignant divide across racial groups in terms of upward mobility for white and African American rural to urban migrants. For white rural to Atlanta migrants, their urban wage premium was more than 100% as they transitioned from mostly low-earning to high-earning occupations; however, for African American migrants, such opportunities were not given, and their urban wage premium was about 15% (systematically lower than all other Norther cities and Chicago).

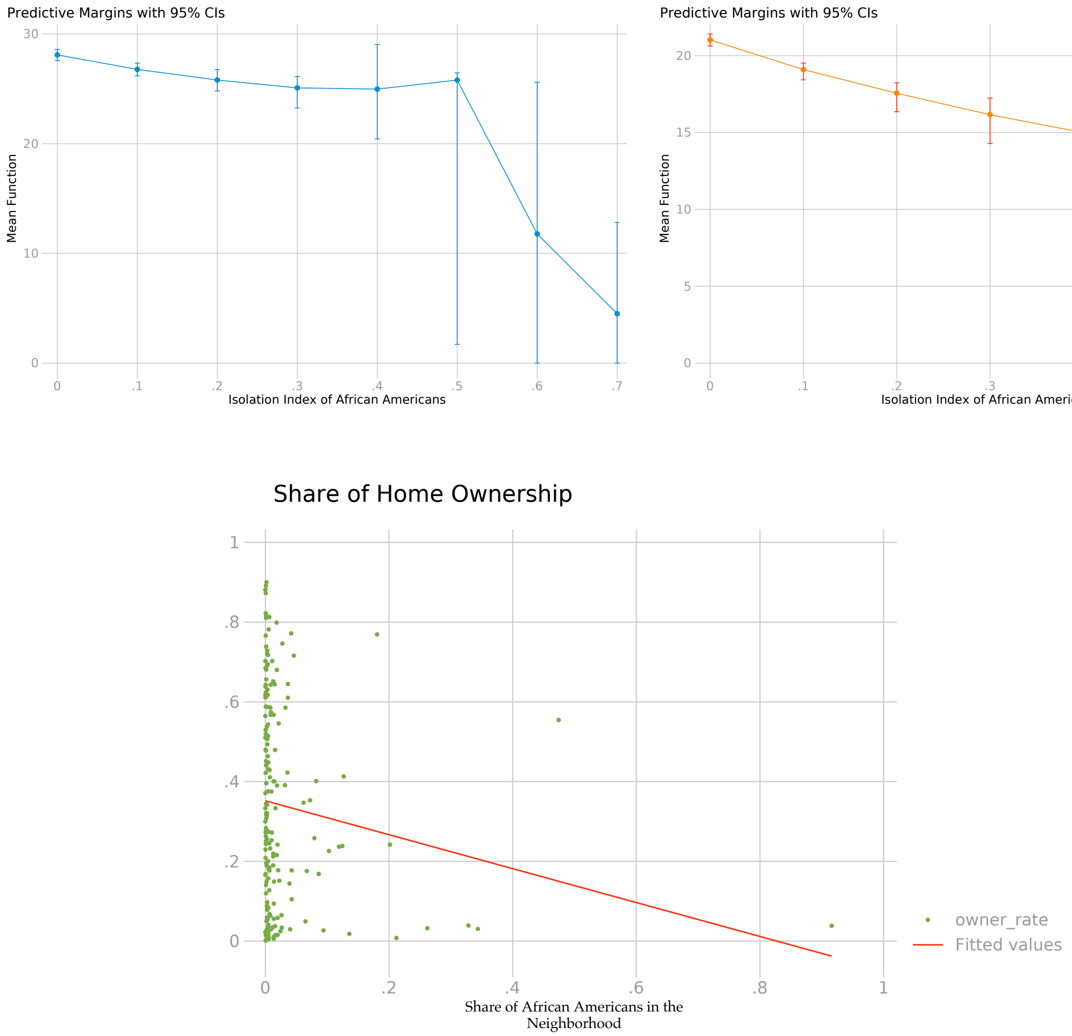
The fact that the urban wage premium differed substantially depending on rural migrants’ destination could imply that there were different types of opportunities and jobs that were available to rural migrants depending on their race. For example, Atlanta, GA was a medium-sized city with a large presence in textile and cotton industries. However, a huge concentration of a small number of industries (i.e. low industry diversity measure) coupled with the Jim Crow laws that enforced racial segregation in the Southern United States may be factors that made African Americans face a higher and stricter "job ceiling" than most other urban cities with higher industry diversity.

6.3 Systematic Barriers for Wealth Formation

In this section, I show the potential implications of minorities living in these highly segregated neighborhoods. Figure 19 demonstrates that the mean earnings of people living in segregated neighborhoods were significantly lower than those living in integrated neighborhoods. Considering Topa (2001)'s finding that contact processes and social interactions could explain labor market outcomes, people living in almost "100% segregated" neighborhoods would *only* be exposed to information about job openings that are low-paying, and this could create a larger *social multiplier*. When the social multiplier is large, we expect to see larger variation in aggregate endogenous variables such as crime rate, concentrated poverty, and housing market depression despite fundamentals (Scheinkman (2021)Scheinkman (2021); Glaeser and Scheinkman (2000)). In other words, the implications and consequences of minorities' extreme isolation could create larger social outcomes such as concentrated poverty, unemployment, and lack of upward mobility through "social (non-market) interactions."

Finally, in terms of wealth formation beyond labor earnings, Figure 19 shows that the share of home ownership decreases as the minority share of the neighborhood increases. Using the 1930 complete count population census, I calculate the minority share of homeownership by neighborhood by looking at every household's homeownership (all households of all races). In "minority" neighborhoods where the minority share is above 0.3, homeownership is extremely low (less than 5%) when compared to the city average of 40%. This implies that the initial wealth formation channel of home ownership was practically unattainable for most African Americans.

Figure 19: Barriers for Wealth Formation



Note: This plot is from a non-parametric regression with a Epanechnikov kernel after partitioning out zoning-regulation related fixed effects. The figure on the Left (blue) uses every man aged between 16 and 60 who worked and lived within the city. The figure on the Right (orange) uses every woman aged between 16 and 60 who worked and lived within the city respectively.

In the bottom, the plot depicts the share of home ownership from the 1930 census. Each dot represents a neighborhood in New York City and I calculate the share of home ownership by neighborhood by looking at every household's homeownership (covering all races in the neighborhood).

7 Conclusion

Through extensive efforts in digitization,record linking, and spatial analyses I establish that both transportation infrastructure and zoning were tightly connected to the decentralization of residential populations and the separation of production locations and residential locations; I also establish that neighborhoods became segregated by residents' socioeconomic status and race after the arrival or subways; Finally, I show that residential zoning heightened the intensity of residential segregation in the neighborhoods where African Americans were extremely isolated (regardless of their income).

Then, through the lens of neighborhood effects mediated by social interactions, I measure the impact of "isolation (non-exposure to the majority group)" on labor market outcomes for African Americans. Minorities living in "tipping" neighborhoods (and that being a stable equilibrium, see Card, Mas and Rothstein (2008)) could mean that their job opportunities may be systematically limited largely due to their social interactions. Economic indicators in segregated minority neighborhoods —absence of upward mobility, concentrated poverty, barriers of wealth formation, decreasing housing prices–tell us "separate but equal" had never existed to begin with. "`Separate and (Extremely) Unequal`" was the reality that people of different classes and races faced at a time of New York City's unprecedented urban transition and economic growth.

The findings that neighborhoods were highly segregated and residential zoning was tightly linked to African Americans' extreme "isolation" in the city imply that the racially-discriminatory policies were underway *way* before "redlining". Second, there could be a bigger welfare disparity between racial groups in terms of neighborhood amenities (e.g. noise, clean air) and long-term health consequences such as long-term exposure to air pollution or hazardous materials (Heblich, Trew and Zylberberg (2021); Logan and Parman (2018)). Third, two facts that 1. minorities' initial wealth accumulation were systematically blocked well before the redlining, and 2. the federal government's HOLC grade (called "redlining") assignment restricted access to credit for disproportionately minority neighborhoods call for conscious efforts in building more equitable and just public policies.

References

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Perez, "Automated Linking of Historical Data," *Journal of Economic Literature*, September 2021, *59* (3), 865--918.

Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey, "How Well Do Automated Linking Methods Perform? Lessons from U.S. Historical Data," Working Paper 24019, National Bureau of Economic Research November 2017.

Bairoch, Paul, *Cities and Economic Development* University of Chicago Press Economics Books, University of Chicago Press, 1991.

Barr, Jason M., *Building the Skyline: The Birth and Growth of Manhattan's Skyscrapers*, Oxford University Press, 2016.

Baum-Snow, Nathaniel, "Did Highways Cause Suburbanization?," Technical Report 2 05 2007.

Boustan, Leah, *Competition in the Promised Land: Black Migrants in Northern Cities and Labor Markets*, National Bureau of Economic Research, Inc, 2016.

_ **, Devin Bunten, and Owen Hearey**, *Urbanization in American economic history, 1800-2000*, Vol. 2, Oxford University Press,

Boustan, Leah Platt, "Was Postwar Suburbanization "White Flight"? Evidence from the Black Migration," *The Quarterly Journal of Economics*, 2010, *125* (1), 417--443.

Brook, Richard R. W. and Carol M. Rose, *Saving The Neighborhood: Racially Restrictive Covenants, Law, and Social Norms*, Harvard University Press, 2013.

Card, David, Alexandre Mas, and Jesse Rothstein, "Tipping and the Dynamics of Segregation," *The Quarterly Journal of Economics*, 2008, *123* (1), 177--218.
City Fixes Limit On Tall Buildings

City Fixes Limit On Tall Buildings*, The New York Times, 1916.*

Costa, Dora L. and Robert W. Fogel*, "Historical Urban Ecological Data, 1830-1930 [dataset]," 2015.*

Cronon, William*, Nature's metropolis : Chicago and the Great West, W.W.Norton, 1991.*

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor*, "The Rise and Decline of the American Ghetto," Journal of Political Economy, June 1999, 107 (3), 455--506.*

Duranton, Gilles and Diego Puga*, "The Growth of Cities," in Philippe Aghion and Steven Durlauf, eds., Handbook of Economic Growth, Vol. 2 of* Handbook of Economic Growth*, Elsevier, 2014, chapter 5, pp. 781--853.*

_ **and Matthew A. Turner***, "Urban growth and transportation," Technical Report 4 2012.*

Eli, Shari, Joshua K. Hausman, and Paul Rhode*, "California and the Car: Early Adoption of the Automobile in the Golden State.," Working Papers, University of Michigan 2021.*

Feigenbaum, James J.*, "Intergenerational Mobility during the Great Depression," 2015.*

Fishback, Price V., Jessica LaVoice, Allison Shertzer, and Randall Walsh, *"The HOLC Maps: How Race and Poverty Influenced Real Estate Professionalsâ Evaluation of Lending Risk in the 1930s," NBER Working Papers 28146, National Bureau of Economic Research, Inc November 2020.*

Fujita, Masahisa, Paul Krugman, and Anthony Venables, *The Spatial Economy: Cities, Regions, and International Trade, 1 ed., Vol. 1, The MIT Press, 2001.*

Glaeser, Edward L and José Scheinkman, *"Non-Market Interactions," Working Paper 8053, National Bureau of Economic Research December 2000.*

Glaeser, Edward L. and Matthew Kahn, *"Sprawl and urban growth," in J. V. Henderson and J. F. Thisse, eds., Handbook of Regional and Urban Economics, 1 ed., Vol. 4, Elsevier, 2004, chapter 56, pp. 2481--2527.*

_ , Bruce Sacerdote, and José A. Scheinkman, *"Crime and Social Interactions," The Quarterly Journal of Economics, 1996, 111 (2), 507--548.*

Goeken, Ron, Lap Huynh, TA Lynch, and Rebecca Vick, *"New methods of census record linking," Historical methods, 2011, 44 (1), 7--14.*

Grebler, Leo, *Experience in Urban Real Estate Investment, New York, Columbia University Press, 1955.*

Gyourko, Joseph and Raven Molloy, *"Chapter 19 - Regulation and Housing Supply," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., Handbook of Regional and Urban Economics, Vol. 5 of* Handbook of Regional and Urban Economics, *Elsevier, 2015, pp. 1289--1337.*

Haig, Robert Murray, *"Toward an Understanding of the Metropolis: II. The Assignment of Activities to Areas in Urban Regions," The Quarterly Journal of Economics, 1926, 40 (3), 402--434.*

_ , *"Major economic factors in metropolitan growth and arrangement; a study of trends and tendencies in the economic activities within the region of New York and its environs," Technical Report 1927.*

Heblich, Stephan, Alex Trew, and Yanos Zylberberg, *"East Side Story: Historical Pollution and Persistent Neighborhood Sorting," Technical Report May 2021.*

_ , Stephen Redding, and Daniel M Sturm, *"The Making of the Modern Metropolis: Evidence from London," The Quarterly Journal of Economics, 2020, 135 (4), 2059--2133.*

Jackson, Kenneth T., *Crabgrass Frontier: The Suburbanization of the United States, Oxford University Press, USA, 1985.*

Kahn, Matthew E., Ryan Vaughn, and Jonathan Zasloff, *''The housing market effects of discrete land use regulations: Evidence from the California coastal boundary zone,'' Journal of Housing Economics, December 2010, 19 (4), 269--279.*

Logan, Trevon and John Parman, *''The National Rise in Residential Segregation,'' NBER Working Papers 20934, National Bureau of Economic Research, Inc February 2015.*

Logan, Trevon D. and John M. Parman, *''Segregation and mortality over time and space,'' Social Science & Medicine, 2018, 199 (C), 77--86.*

Makielski, S J, *The politics of zoning : the New York experience, New York : Columbia University Press, 1966.*

Massey, Douglas S. and Nancy A. Denton, *American Apartheid: Segregation and the Making of the Underclass, Harvard University Press, 1998.*

Nicholas, Tom and Anna D. Scherbina, *''Real Estate Prices During the Roaring Twenties and the Great Depression,'' Real Estate Economics, March 2011, 18-09.*

Nix, Emily and Nancy Qian, *''The Fluidity of Race: âPassingâ in the United States, 1880-1940,'' Working Paper 20828, National Bureau of Economic Research January 2015.*

of New York Board of Estimate New York, New York : City and Committee on the city plan Apportionment, *''Final Report: 1916,'' Technical Report 1916.*

of New York, Regional Plan and Its Environs, *''Food, clothing and textile industries, wholesale markets and retail shopping and financial districts, present trends and probable developments,'' Technical Report 1927.*

_ and _ , *''Chemical, metal, wood, tobacco and printing industries; present trends and probable future developments,'' Technical Report 1928.*

O'Flaherty, Brendan, *City Economics, Harvard University Press, 2005.*

Price, Joseph, Kasey Buckles, Jacob Van Leeuwen, and Isaac Riley, *''Combining Family History and Machine Learning to Link Historical Records,'' Working Paper 26227, National Bureau of Economic Research September 2019.*

Redding, Stephen and Matthew Turner, *''Transportation Costs and the Spatial Organization of Economic Activity,'' in ''in,'' Vol. 5, Elsevier, 2015, chapter Chapter 20, pp. 1339--1398.*

Rothstein, Richard, *The Color of Law: A Forgotten History of How Our Government Segregated America*, Liveright, 2017.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek, *"IPUMS USA: Version 9.0 [dataset],"* 2019.

Scheinkman, José A., *"Social Interactions," Working Papers, Columbia University* 2021.

Schelling, Thomas C., *"Dynamic models of segregation," The Journal of Mathematical Sociology*, 1971, *1* (2), 143--186.

Shertzer, Allison, Randall P. Walsh, and John R. Logan, *"Segregation and Neighborhood Change in Northern Cities: New Historical GIS Data from 1900 to 1930," Technical Report* 2016.

Topa, Giorgio, *"Social Interactions, Local Spillovers and Unemployment," The Review of Economic Studies*, 04 2001, *68* (2), 261--295.

Tsivanidis, Nick, *"The Aggregate And Distributional Effects Of Urban Transit Infrastructure: Evidence From BogotÃ¡'s TransMilenio,"* 2019.

Ward, David and Olivier Zunz, *Landscape of Modernity: Essays on New York City, 1900-1940*, Russel Sage Foundation, 1992.

W.Dunlap, David, *"Zoning Arrived 100 Years Ago. It Changed New York City Forever.," The New York Times*, 2016.

Williamson, Jeffrey G. and Peter H. Lindert, *Unequal Gains: American Growth and Inequality since 1700*, Princeton: Princeton University Press, 2016.

Appendix

This Appendix reports additional information about the definitions and data sources used in the paper. Details about their creation and processing supplement the discussion in Section 3. Section A discusses data linking procedures and relevant details from 1870-1940. Section B discusses the construction of my geographic information systems (GIS) shapefiles from various sources including the digitization of Enumeration District maps from microfilms, the overground and underground railway networks over time, and zoning regulations in 1916.

A Record Linking

In this section, I describe the record linking procedure and relevant details. In constructing a panel of individuals, I use "Machine Learning," where the machine can self-link individuals after learning the pattern of ''true'' and ''false'' matches from training datasets. This method is implemented to link individuals across census years while maximizing the match rate and representativeness of linked datasets. I link complete-count US Federal Decennial Demographic Census records from 1850 to 1940 with newly transcribed socioeconomic variables such as occupation and employed industry.

A.1 Machine Learning Approach of Record Matching

The "machine learning" approach for record linking borrows insights from computer science and statistics, and I implement this method of classification and text comparison to link individual records. The rationale behind my choice of machine learning is to learn from big data. In essence, record linking without unique identifier is to predict whether certain linked records are "true" links of the same individual or not based on a set of features such as first name, last name, age, and place of birth. Similar efforts have been pioneered by Goeken, Huynh, Lynch and Vick (2011) which creates the IPUMS linked samples. Feigenbaum (2015) links individual records from the 1915 Iowa State Census to their adult-selves in the 1940 US Federal Demographic Census records. Relative to the mentioned works, my record linking is far more extensive in the scope of matching as it involves complete-count US Federal Decennial Demographic Census records of all years from 1850 to 1940. I teach a machine to learn to predict based on a set of features. I create a training datasetwhich contains both ''true'' and ''false'' matches and their characteristics (e.g some observations with ''true'' as an outcome would have same/very similar characteristics in terms of age, first and last name, parents' and his/her birthplaces whereas observations with ''false'' as an outcome would have quite different characteristics in terms of the above mentioned characteristics). In this case, the outcome is whether the matched records are a "true" or "false" match, given the observed characteristics. By taking this training data, I build a prediction model, or learner, which will enable us to predict the outcome for new, unseen records. A well-designed learner armed with a solid training dataset should accurately predict outcomes for new, unseen records.

I implement a supervised learning problem in the sense that the presence of outcome variable ("true" or "false" links) guides the learning process—in other words, the end-goal is to use the inputs to predict the output values. To summarize this process, I extract subsets of possible matches for each record and create training data in order to tune a matching algorithm so that the matching algorithm matches individual records by minimizing both false positives and false negatives while reflecting inherent noises in historical records. I have explored various models for model selection.

By comparing and analyzing matched records produced through various methods, I chose the random forest classification as it is *more conservative* in matching records—the number of matched records is lower than that of Support Vector Machine (hereafter, SVM)— and the number of unique matches are significantly higher than the standard SVM model. Although the choice of random forest classification may result in lower number match rate due to its conservative nature, I integrated household-level information in linking individual records to mitigate the concerns of low match rate.

A filtering process called "pruning" for non-unique matches

Although I largely follow the standard machine-learning record linking methodology suggested by Goeken, Huynh, Lynch and Vick (2011), I have extended the techniques of Goeken, Huynh, Lynch and Vick (2011) by inventing a two-step machine learning matching methodology. Especially, I make use of the parents and/or spouse information such as birthplaces and names to choose the "true" match among the set of candidate matches. This additional step of extracting household-level information and using it in selecting "true" matches among multiple candidates (instead of dropping non-unique matches, which have been the "standard" practices in the existing record matching literature) is novel. This procedure can not only save a number of matches that otherwise had to be dropped but also correct for selection bias (people with common characteristics such as common first and last names may be systematically under-represented in linked datasets)..

A.2 Record Linking in Practice: Innovations

The core of census matching is a classification problem. Given any pair of records from different census years, finding a true match is to find the mapping that classifies the pair as matched or unmatched based on the set of pre-determined features including name, gender, age, race and birthplace. However, since this set of features is far from unique, there are cases where one individual has several candidate matches (e.g. there are many ''John Smith'' with same age).

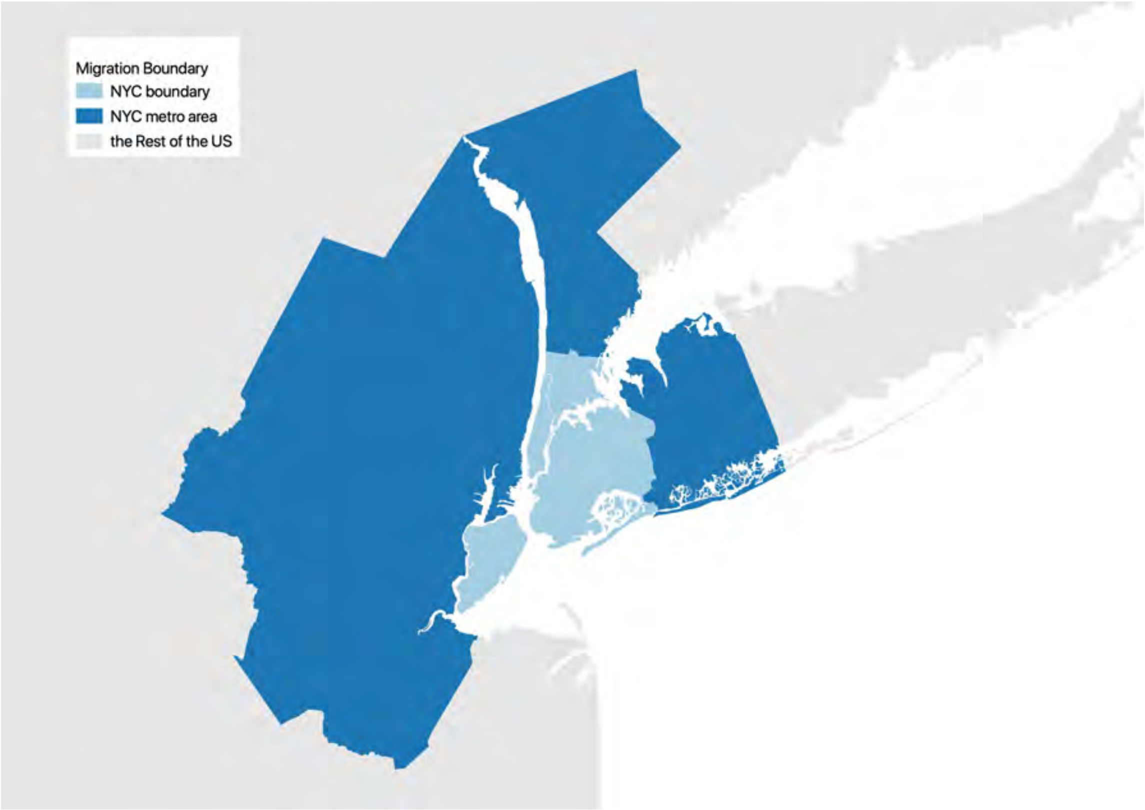
Most record linking approaches throw away non-unique matches. One of the contributions of my record linking approach is the use of household-level information to turn the non-unique types of matches (second to fourth type) to unique matches. Specifically, I use information such as the racial background, birthplaces, and birth year of an individuals father, mother, and spouse to identify the "true" match. . This not only increases the match rates but also alleviates the concern of systematic selection bias (e.g. people with common given given and last names may be systematically under-represented in the linked data).

B Geographic Data

B.1 Geographic Definition

Here are essential boundary definitions and taxonomy for this paper. A metro area, or metropolitan area, is a region consisting of a large urban core together with surrounding communities that have a high degree of economic and social integration with the urban core. I follow the IPUMS-definition, and delineation of the metro area of New York City.[35] Figure 20 shows the geographic boundaries of NYC, the NYC metro area, and the rest of the country. The IPUMS-delineation of the metro area of the city applies the 1950 Office of Management and Budget standards to historical statistics (Ruggles, Flood, Goeken, Grover, Meyer, Pacas and Sobek (2019)). This approach yields time-varying delineations of regions with a high degree of economic and social integration with the urban core which is ideal for my study (For example, Suffolk County, New York was not part of NYC metro area till 1920, however, as the economic integration between Suffolk county and NYC increased, Suffolk county became a part of NYC metro area since 1930). As in Figure 20, I define the five boroughs of New York City (in Light blue), the NYC metro area (in Dark blue), and the rest of the United States (in Light gray) by following the IPUMS delineations of the NYC metro areas.

Figure 20: Geographic Boundary of NYC Metro Area



B.2 Geographic Information Harmonization

Georeferencing and Creation of Shapefiles: Neighborhood Boundaries The primary geographic units used in the analyses are "Neighborhood Tabulation Areas" (hereafter, NTAs), which each feature at least 15,000 people in 2010 (there are 195 NTAs (neighborhoods) within the city). As datasets used in the analyses have different spatial units and/or the boundaries of the spatial unit constantly change, I create spatial crosswalks from historical spatial locations found in various data sources (e.g. "enumeration district" in US census records) to NTAs so that NTAs can be a

---

[35] Description and definition of a metropolitanare available here: `https://usa.ipums.org/usa-action/variables/METAREA#description_section`

time-invariant, consistent geographic unit of analyses. Therefore, all datasets used in the analyses are harmonized and geolinked to NTAs.

An Enumeration District is a historical version of a ‘‘census tract’’ where the historical US census enumerators recorded administrative divisions smaller than counties (and wards which were extensively used in existing literature). As the individual-level US Federal Demographic Census provides ED numbers, I can now aggregate the individual-level information to the neighborhood, or similar geographic levels, within the city. As Geographic Information System (hereafter, GIS) software enables researchers to know where these geographic units are in space, efforts of georeferencing ED images from microfilms and creating GIS-compatible shapefiles must be made to execute the analyses during the study period (i.e. 1870 to 1940).

This digitization effort has benefited from existing projects called the Urban Transition NHGIS (Logan et al, 2011) and Shertzer, Walsh and Logan (2016). I complement the existing sources by pushing the time horizon and geographic scope—1880 Enumeration District boundary files of Manhattan and Brooklyn were obtained from the Urban Transition Historical GIS project; I use Manhattan and Brooklyn ED boundary shapefiles from 1900 to 1930 from Shertzer, Walsh and Logan (2016). However, as Shertzer, Walsh and Logan (2016) mainly focus on studying the ten largest US cities, they did not digitize the relatively unpopulated areas of the Bronx, Queens, and Richmond. Therefore, I used microfilm scan images of New York City Enumeration District maps of the period1880-1940 and created historical GIS files for the remaining regions across time. For boroughs where microfilm scan images were not available in each period, such as Queens county in 1900 and Richmond County and Bronx county in 1910, I use the detailed street and building information of residential addresses from the individual-level census records to locate which ED corresponds to each neighborhood. Stephen P. Morse's website has resources for ED finding tools for 1900 to 1940 censuses `https://stevemorse.org/census/unified.html`, and I mainly reference this website to check the conversion between different census years and old street names and ED boundaries.

A major difficulty in making use of ED-level analysis while using the above-mentioned boundary files is that the ED boundaries change considerably across time, making it extremely challenging to form consistent neighborhoods.[36]I tackle this problem by taking the time-consistent neighborhood boundaries called Neighborhood Tabulation Areas (called ‘‘NTAs’’) created by the Department of City Planning in New York City.[37] To do this, I first find original ED maps — I found the original ED map images via FamilySearch, and some of those images were obtained from the National Archives Records Administrations (NARA) microfilm publication.[38] Then, I georeference these original maps (as seen in Figure 21) and constructed shapefiles for the ED boundary for every census year (see the black polygons and boundary lines in Figure 21). Then, I overlay ED shapefiles against the time-consistent neighborhood boundary shapefile (red line in Figure 21) to create spatial crosswalks from ED boundaries to NTA neighborhoods over the study period.

---

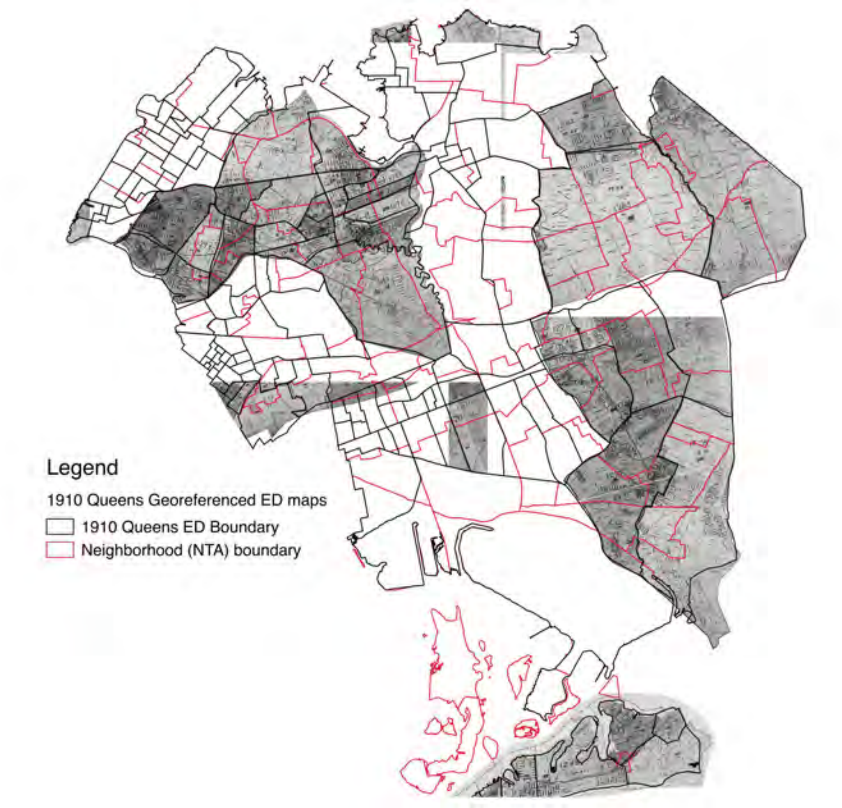
[36] Shertzer, Walsh and Logan (2016), for example, approach this problem by harmonizing ED data to temporally invariant geographically defined areas that they treat as “synthetic neighborhoods” for studying neighborhood change.

[37] Description and related GIS software-compatible files of Neighborhood Tabulation Areas are available here: `https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-nynta.page`

[38] I obtained some of these images titled ‘‘United States Enumeration District Maps for the Twelfth through the Sixteenth US Censuses, 1900-1940 images’’ from FamilySearch (`https://familysearch.org/pal:/MM9.3.1/TH-1961-35062-11409-75?cc=2329948:20June2014`; other images that were unavailable through FamilySearch were obtained through microfilm publications (Roll 42, New York, New York City boroughs; Niagara-Rockland 1900-1940, NARA microfilm publication A3378 (Wasington, D.C.: National Archives and Records Administration, 2003)).

For every ED and every NTA, I aggregate the variables by aggregating the complete-count US Demographic census. Examples of such are total population, age, family size, occupation-based earning and education measures, marital status, and race.

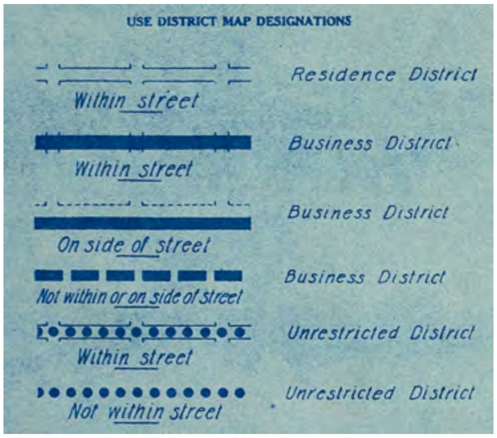
Figure 21: Geographic Boundary Harmonization Procedure: Original ED Maps & Georeferencing & Shapefile Creation and Spatial Crosswalk



1916 Zoning Resolution In addition to my GIS shapefiles of all Enumeration Districts found in the federal census from 1880 -1940, I also construct similar shapefiles for New York City's 1916 Zoning Resolution. Similar to the creation of ED boundary shapefiles, I georeference original maps of the 1916 Zoning Resolution in NYCand then construct area shapefiles for each land-use regulations — Use, Area, and Height. Figure 22 shows how these georeferenced maps were used to construct shapefiles for 1916 Zoning Resolution regarding land use; yellow lines represent modern streets, and I use overlaid maps to indicate which streets and areas were zoned for certain land uses and which areas are unrestricted; bold original lines in Figure 22 were designated for Business District uses only. The legend from the original zoning maps were coded into the modern street shapefiles and the original legend is available in Figure 23(b).

Figure 22: Creation of GIS shapefiles of 1916 Zoning Resolution

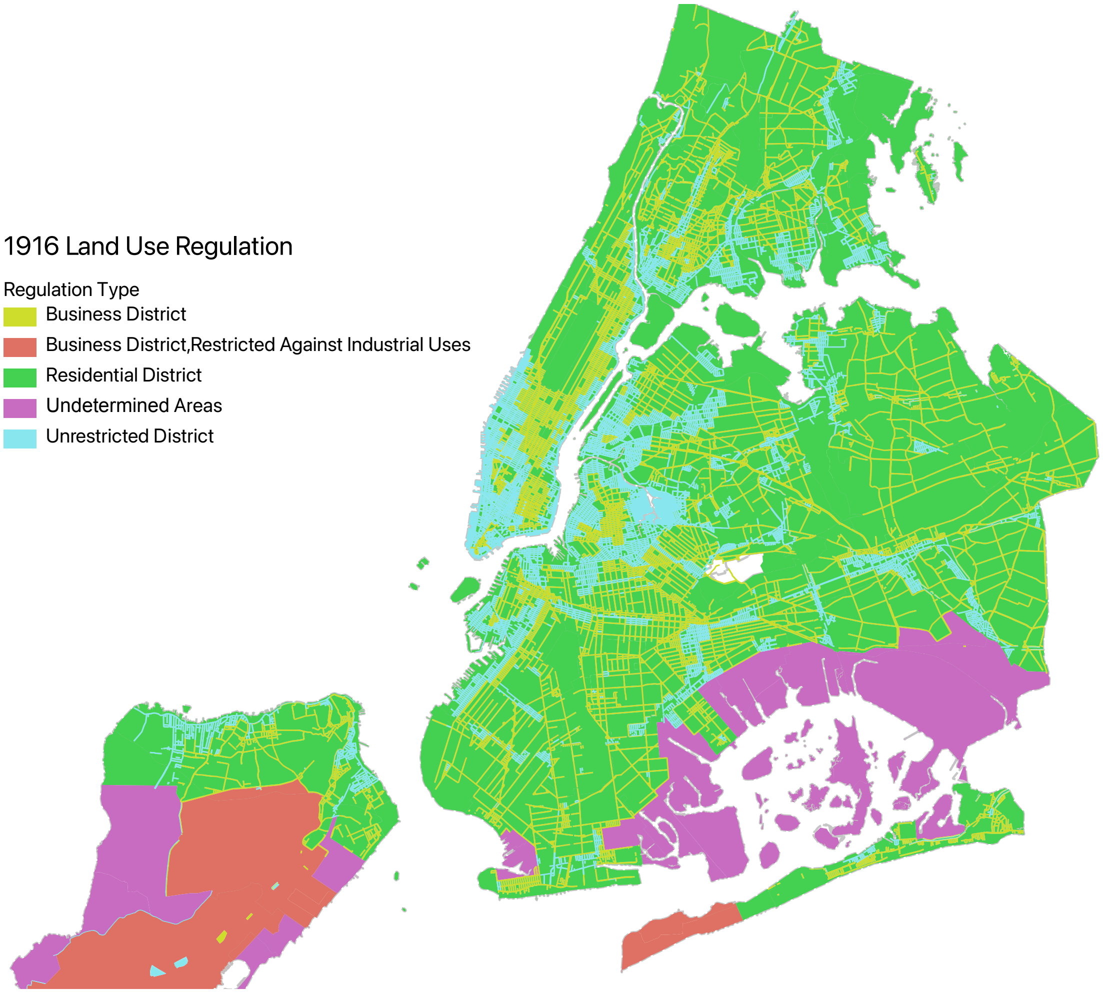
(a) Georeferenced 1916 Zoning Land-Use Restrictions map



(b) Legend of 1916 Zoning Resolution: Land Use

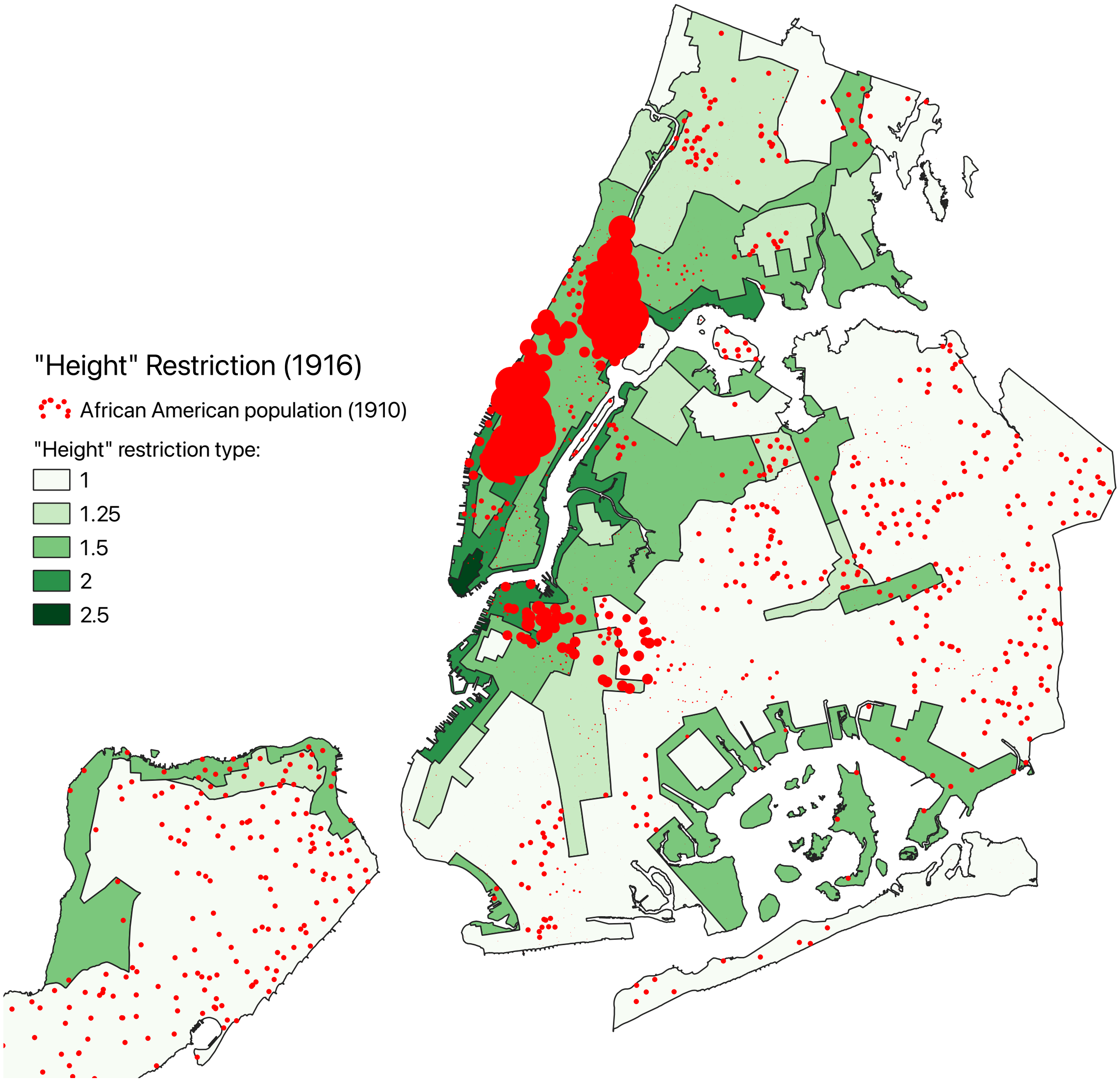
Note: The above figure shows the 1916 NYC Zoning maps related to the *Use regulations*. I construct GIS-compatible shapefiles by georeferencing the original maps from 1916 historical zoning maps that I obtain from the New York Public Library Map Division. Upon compiling all original images, I first georeference individual images of zoning map using GIS-software; then I overlay the georeferenced maps with the current streets file from New York State (`http://gis.ny.gov/gisdata/inventories/details.cfm?DSID=932`), and create what zoning land regulations were in place street by street. Different lines mean different types of land-use regulation and I document implemented use-related restrictions at the street-level.

Figure 23: Land-Use Restriction Zoning Map



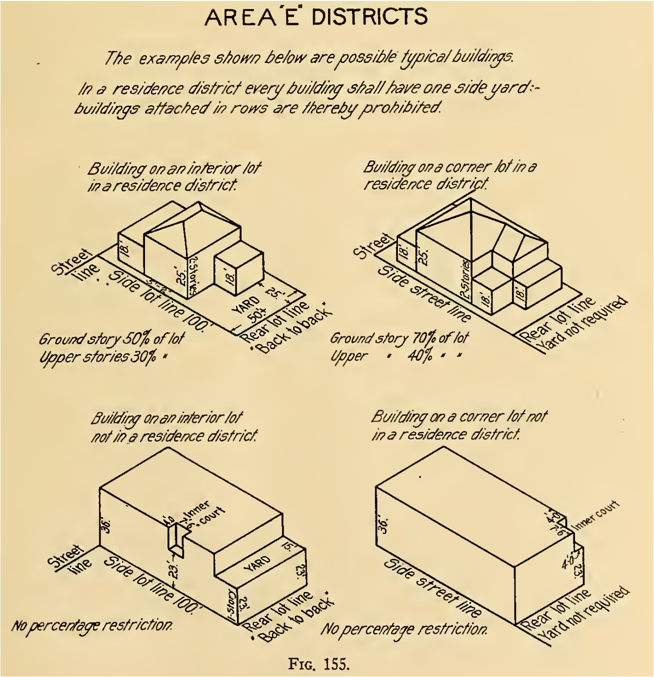
Note: Author's own creation based on digitized and georeferenced maps of 1916 Zoning Ordinances on Use-Restriction. Originally map images were scanned and provided the New York Public Library Map Division. I georeferenced the use-restriction maps down to street level accuracy by overlaying these use restriction shapefiles with the entire street shapefile of New York City (This GIS Data set for entire streets in New York State can be access in the following link `http://gis.ny.gov/gisdata/inventories/details.cfm?DSID=932`).

Figure 24: Height-Restriction Zoning Map (1916) & Connection to pre-zoning population trend (1910)

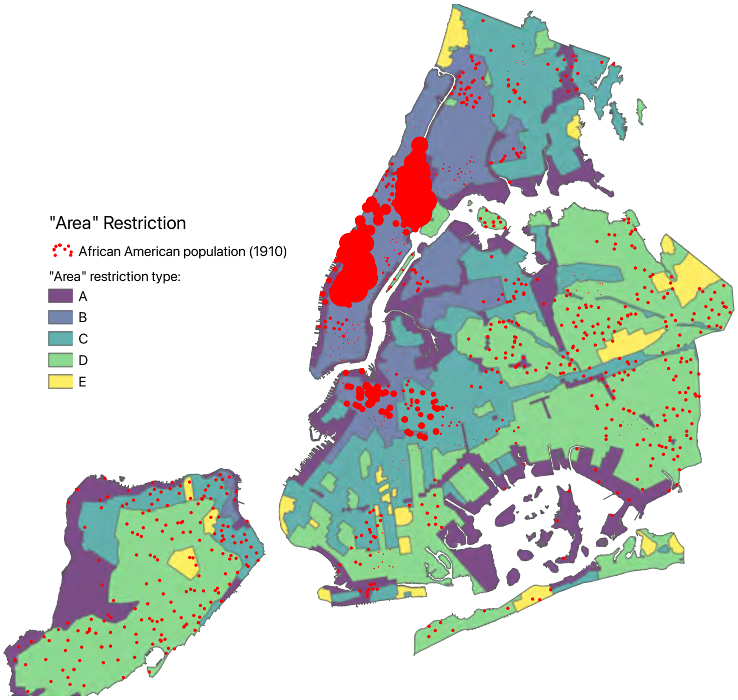


Note: Author's own creation based on digitized and georeferenced maps of 1916 Zoning Ordinances. Originally map images were scanned and provided the New York Public Library Map Division. Unlike Figures 9, 10, the connection between the spatial distribution of African Americans in 1910 before zoning and spatial assignment of ‘‘Height'' zoning restrictions (1916) seems less salient.

Figure 25: Illustrations of Area Restrictions & Connection to pre-zoning population trend



(a) Illustration of Area Restrictions



(b) pre-zoning population trend (1910) Height regulation (1916)

Note: Figure 26(a) was originally from New York and Apportionment (1916). Unlike Figures 9 and 10, the connection between the spatial distribution of African Americans in 1910 before zoning and spatial assignment of "Area" zoning restriction (1916) seems less salient.

C Supplementary Data, Details and Figures

C.1 Occupational data

OCCSCORE is a constructed 2-digit numeric variable that assigns occupational income scores to each occupation in all years of pre-1950 US censuses which represents the median total income (in hundred of 1950 dollars) of all persons with that particular occupation in 1950.[39] Then, OCC1950 is the base of creating OCCSCORE and OCC1950 is divided into 10 social classes and 269 occupational groups and has been the US standard for occupational coding due to its strength in comparability across years. However, it has potential shortcomings of not reflecting the relative wage changes, and relative wages that may differ across locations. Despite these potential shortcomings, this approach allows me to document neighborhood changes in terms of residents' skills over time (during my study period, the US Federal demographic census records asked neither one's income nor educational attainment with the exception 1870 and 1940).

C.2 Digitized Real Estate Transaction Records

I digitize the entire section of real estate sales transactions from the *Real Estate Record and Builders' Guide* from 1870 to 1940 in order to investigate how the provision of urban transportation infrastructure affected growth within cities and property values.[40] In addition to chronicling details of mortgages, permits, advertisements, and more, the *Real Estate Record and Builders' Guide* contains details on a significant percentage of property sales in the New York metropolitan area.

All records from all years contain basic information on the property sale in consideration including street and house number, distance and direction from the nearest street intersection, seller and buyer, and sale price. Records for all years are occasionally accompanied by size, property details (e.g. story count, building material), and information on related mortgages, titles, and foreclosures. Beginning in 1905, the majority of records include both a lot assessment and a property-and-lot assessment for tax purposes.

A representative record from 1912 is as follows:

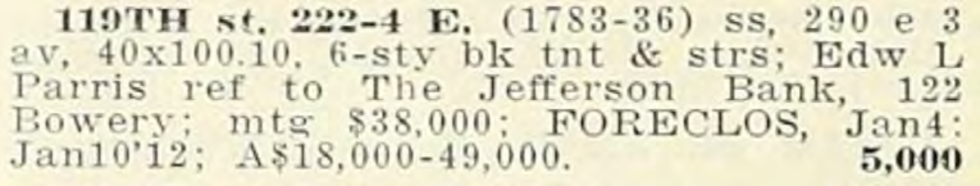
Record volumes were scanned and processed using ABBYY FineReader's optical character recognition engine. A predictive recursive descent parser was used to deterministically parse record text into usable transaction data. While record format and contents vary significantly by year, all records tightly follow a consistent structure dictating the sequential relationships of record components (e.g. mortgage data is terminated by a semicolon and followed by a deed filing date). This structure lends itself well to the capabilities of a parser that examines the start of an input string and selects the correct representative symbol from a set of contextually possible symbols, outputting a series of symbols representing the record. From the output of this parser, transaction

---

[39] Detailed description of "OCCSCORE" and "OCC1950" are available here: https://usa.ipums.org/usa-action/variables/OCCSCORE#codes_section, https://usa.ipums.org/usa-action/variables/OCC1950#description_section

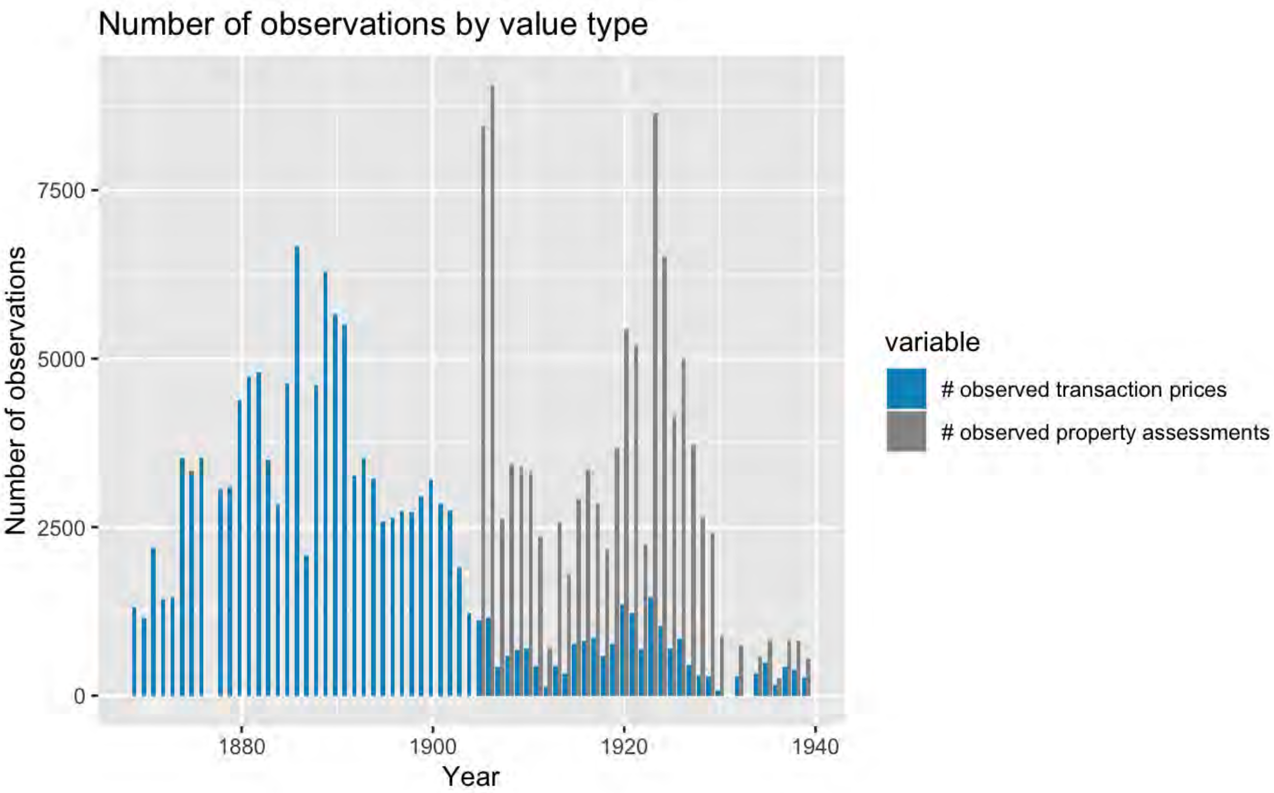
[40] *The Real Estate Record and Builders' Guide* were professional weekly publications of New York City real estate-related activities which included information on sales transactions, mortgages, leases, and liens along with prices of construction materials. Nicholas and Scherbina (2011) previously used these market-based transactions to document real estate prices during the 1920s and the Great Depression. However, they only collected 30 transactions per month for Manhattan between 1920 and 1939. Unlike them, I digitize the entire run of *the Builders' Guides* without sampling; this includes transactions for not only Manhattan but also other boroughs and suburbs in the New York metro area.

Figure 26: Real Estate Transaction Data Details



119TH st. 222-4 E. (1783-36) ss, 290 e 3 av, 40x100.10, 6-sty bk tnt & strs; Edw L Parris ref to The Jefferson Bank, 122 Bowery; mtg $38,000; FORECLOS, Jan4: Jan10'12; A$18,000-49,000. **5,000**

(a) Sample Real Estate Record Example



(b) No. Obs from House Sales Transaction Data

data can be extracted for analysis. For example, the above record from 1912 is decomposed into the following:

```
primary street : 119TH st
house number : 222-4 E, (1783-36)
street side : ss
secondary : 290 e 3 av
size : 40x100.10
property type : 6-sty bk ant & strs
seller : Edw L Parris ref
buyer : The Jefferson Bank, 122 Bowery
mortgage : mtg $38,000
foreclosure info: FORECLOS, Jan4
price: 5000
deed filing date : Jan10'12
tax assessment (lot, prop & lot): A$18,000-49,000
```

If the parsing output for a record was missing either its primary street or the intersecting secondary street, the record was dropped. Bundled properties were also dropped if detected, as they were relatively infrequent and their widely varying structure made parsing an intractable problem.

Each transaction record was then geocoded using the Google Maps Geocoding API. Geocoding queries were generated using the nearest intersection to avoid variance over time in house numbering. Because each record includes the property's bearing from its geocoded intersection, there is minimal information loss on the property's location. The intersection and bearing information can be used to merge the real estate records with other street-level data, like the new 1916 street-level building height regulations dataset.

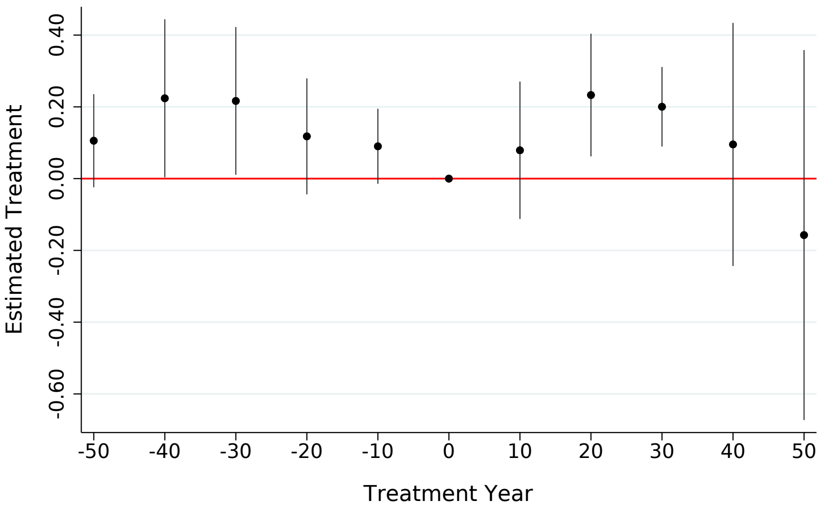
While nearly seventy percent of geocoding responses were matched as intersections (or train stations named after intersections), a portion of responses were street- or route-only matches to one of two streets in the geocoding query. Because a majority of analysis is done at the neighborhood-level, these street-only matches were preserved by filtering for streets that are fully encapsulated by a neighborhood (e.g. Bank St is fully encapsulated in the West Village neighborhood; thus a Bank St street-only geocoded record can be preserved). In addition, observations falling outside of the NYC Neighborhood Tabulation Areas were dropped..

Data cleaning primarily consisted of handling image-to-text character misclassifications and problems with image segmentation. Leading dollar signs were frequently misclassified as numbers and were handled based on each volume's notation conventions. Property-and-lot assessments were "imputed" (corrections) for a small subset of records where the lot assessment was initially found to be greater than the property-and-lot assessment value due to missing zeros digits. Finally, volume year was computed from the volume number, as the deed drawing and filing dates were subject to significant misclassification.

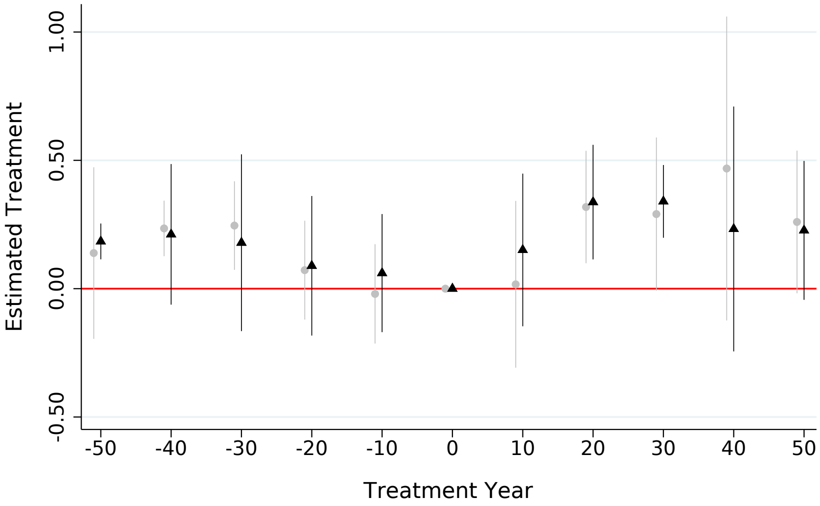
Price and assessment indexes were constructed with a pooled hedonic regression model. The hedonic regression method was chosen to avoid inconsistencies in house numbering and issues with property changes over time, both of which are known to affect the accuracy of repeated-sales indexes. Characteristics included in the regression include (but are not limited to) building materials, story count, foreclosure status, property usage, and reported square footage.

Figure 27: House Price Changes

All Neighborhoods in New York City



Non-Central Neighborhoods of New York City



Note: Estimated treatment effects from railway station arrival on the log hedonic house sales price index (by neighborhood); the sample includes 195 time-consistent neighborhoods in New York City. The digitized house sales data is at the transaction-level with a daily frequency for 70 years and approximately total 450,000 observations. All specifications include neighborhood fixed effects, year fixed effects, and neighborhood-specific time trends (the estimated coefficients and standard errors are reported in Appendix Table 5; 6).

The left figure shows estimated treatment effects (β_τ from equation 9) for all neighborhoods in the city; the right figure shows the heterogeneous treatment effects (γ_τ from equation 9) between neighborhoods in the city core and city periphery. The vertical lines in both left and right figure show the estimated 95% confidence intervals. Standard errors are clustered on boroughs. In the left and right figures, the first specification estimates equation 9(estimates are illustrated in grey circles) by using ‘‘Manhattan’’ as the definition of city-core, whereas the second specification estimates equation 9 (estimates are illustrated in black triangles) by using ‘‘5km from the Battery’’ as the definition of the city-core.

C.3 Visual Inspection of Residential Zoning's Impact on Segregating Minorities

Here, I visually show the spatial distribution of races across space over a one hundred year period. In all figures of Figure 28, one dot plots 500 people, and the color red denotes African American population whereas the color green denotes white population. Figure 29(a) exactly depicts New York where the majority of population was living near workplaces (within 5-kilometer distance from the Battery which is how I defined the "city center"). However, with the construction of elevated railway in parts of Brooklyn and Manhattan, the residential population began spreading out across the city. Starting in 1904, the city began building a more extensive network of subways and more people began decentralizing as parts of Bronx experienced substantial population gains.

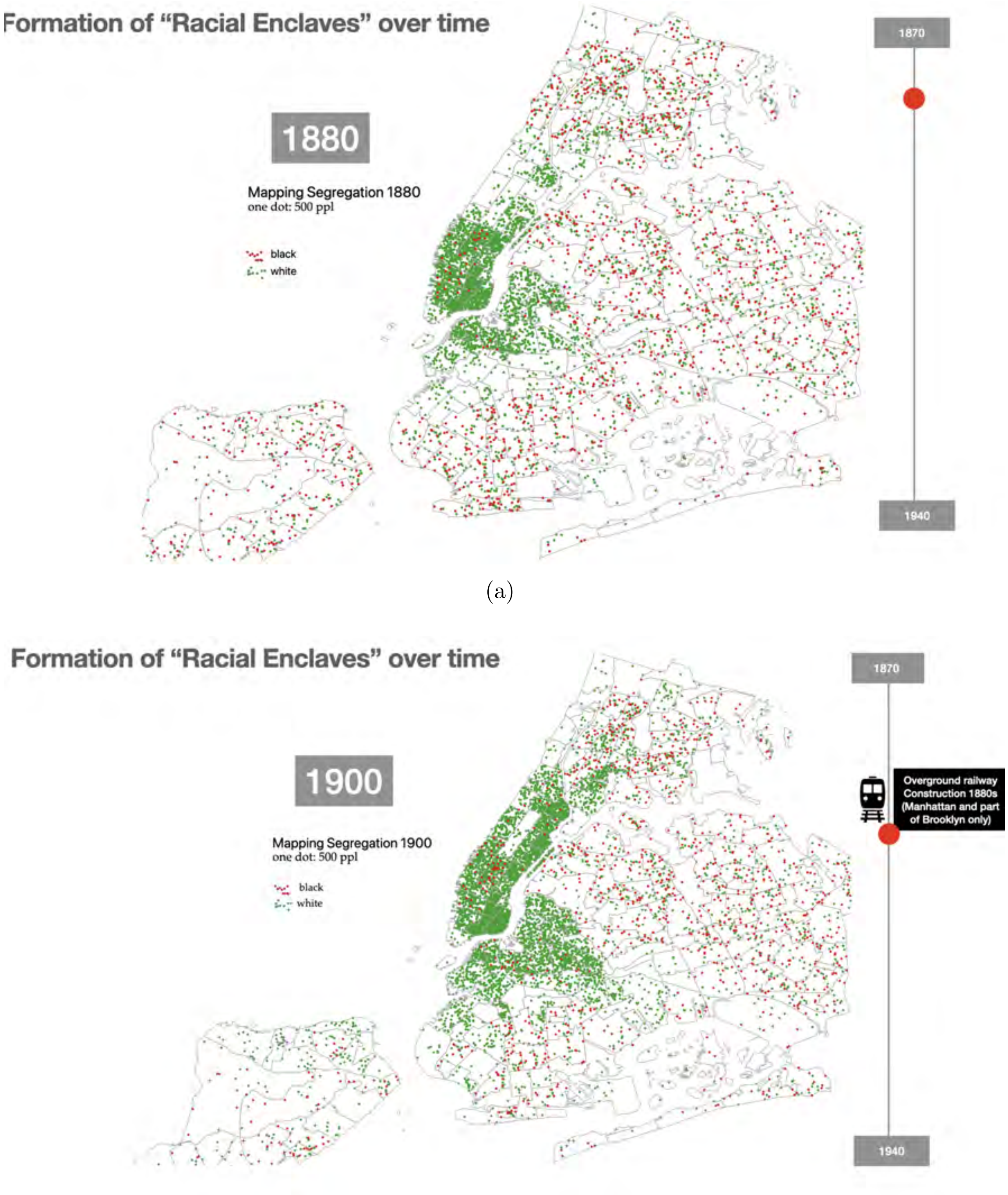
In terms of the spatial distribution of race integration, as captured by red and green dot together in proximity, African American and white population were not severely segregated up until 1910 (before the Great Migration). However, zoning was implemented in 1916 and by 1920 (figure 29(d)) African Americans were disproportionately more likely to live in a few neighborhoods in the city (i.e. Harlem, Manhattan and Weeksville/Crown Heights, Brooklyn). With the rise of Midtown starting around mid-1910s (with the opening of Penn Station), one of the African American neighborhoods called San Juan Hill, Manhattan (which is now a part of Lincoln Center) became a part of history.

By 1930 (Figure 29(e)), near the end of the first wave of the Great Migration, all African Americans lived in two extremely "isolated" neighborhoods where the isolation measure was close to 0.89 for Central Harlem.[41] Figure 29 shows the residential locations between the rich (the ones in top quartile in earning percentile rank) and poor (the ones in bottom quartile in earning percentile rank) African Americans. Figure 29 reveals that both rich and poor African Americans "shared" neighborhoods regardless of their income, "isolated" from the majority population.

---

[41] Isolation Index of 0 means a perfectly integrated/no segregation; Cutler, Glaeser and Vigdor (1999) defines Urban ghettos where minorities' Isolation Index is above 0.6.

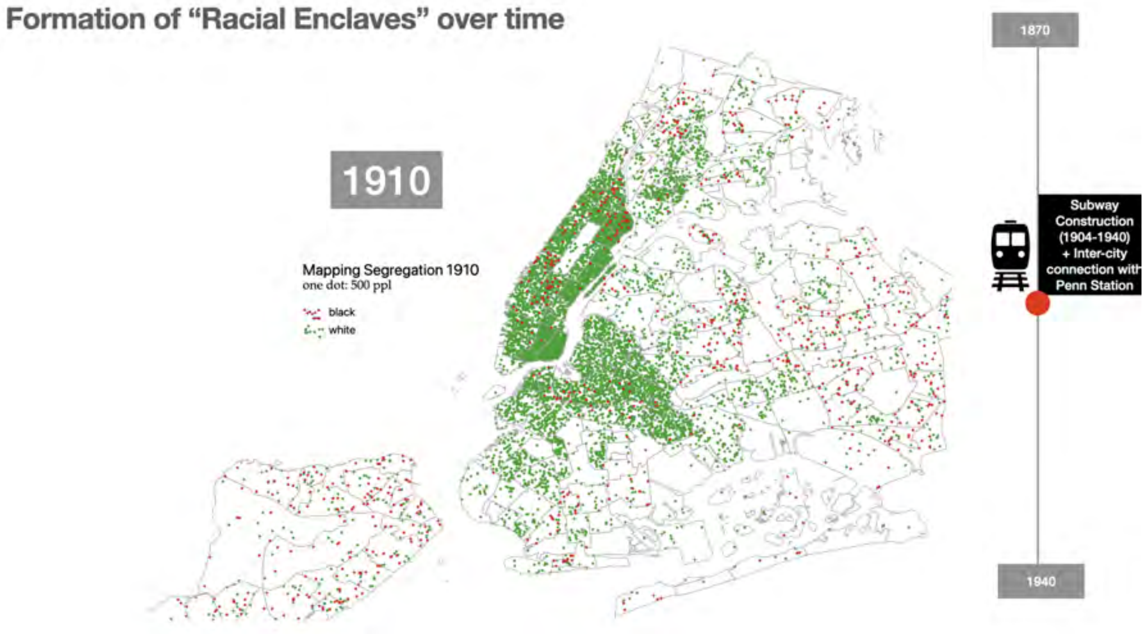
Figure 28: Mapping Residential Segregation by Race



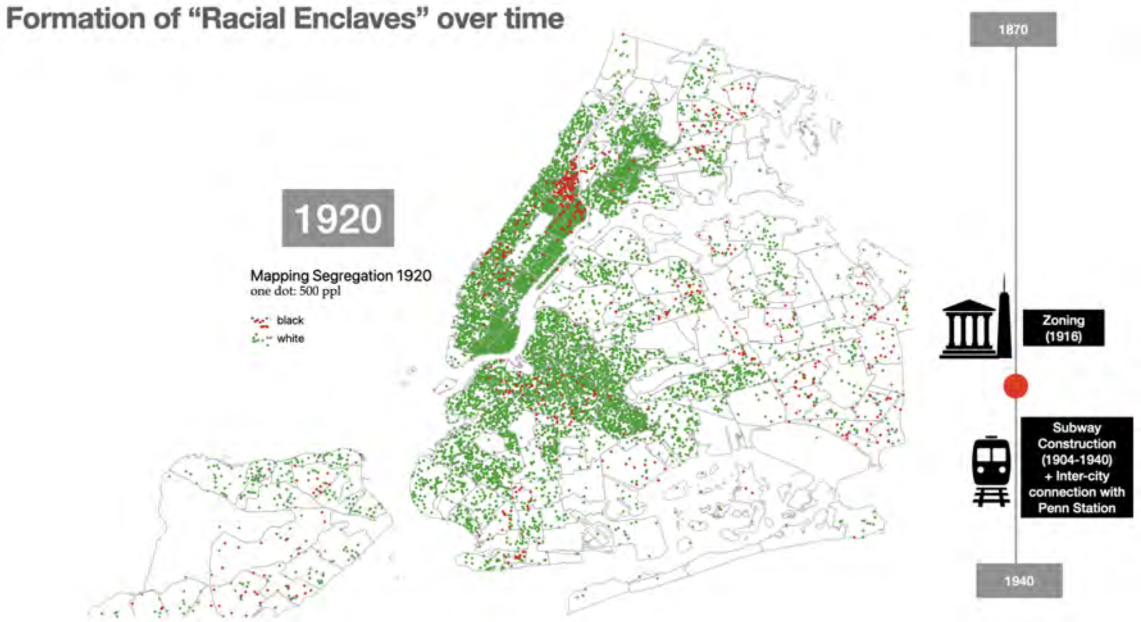
(a)

(b)

Figure 28: Mapping Residential Segregation by Race

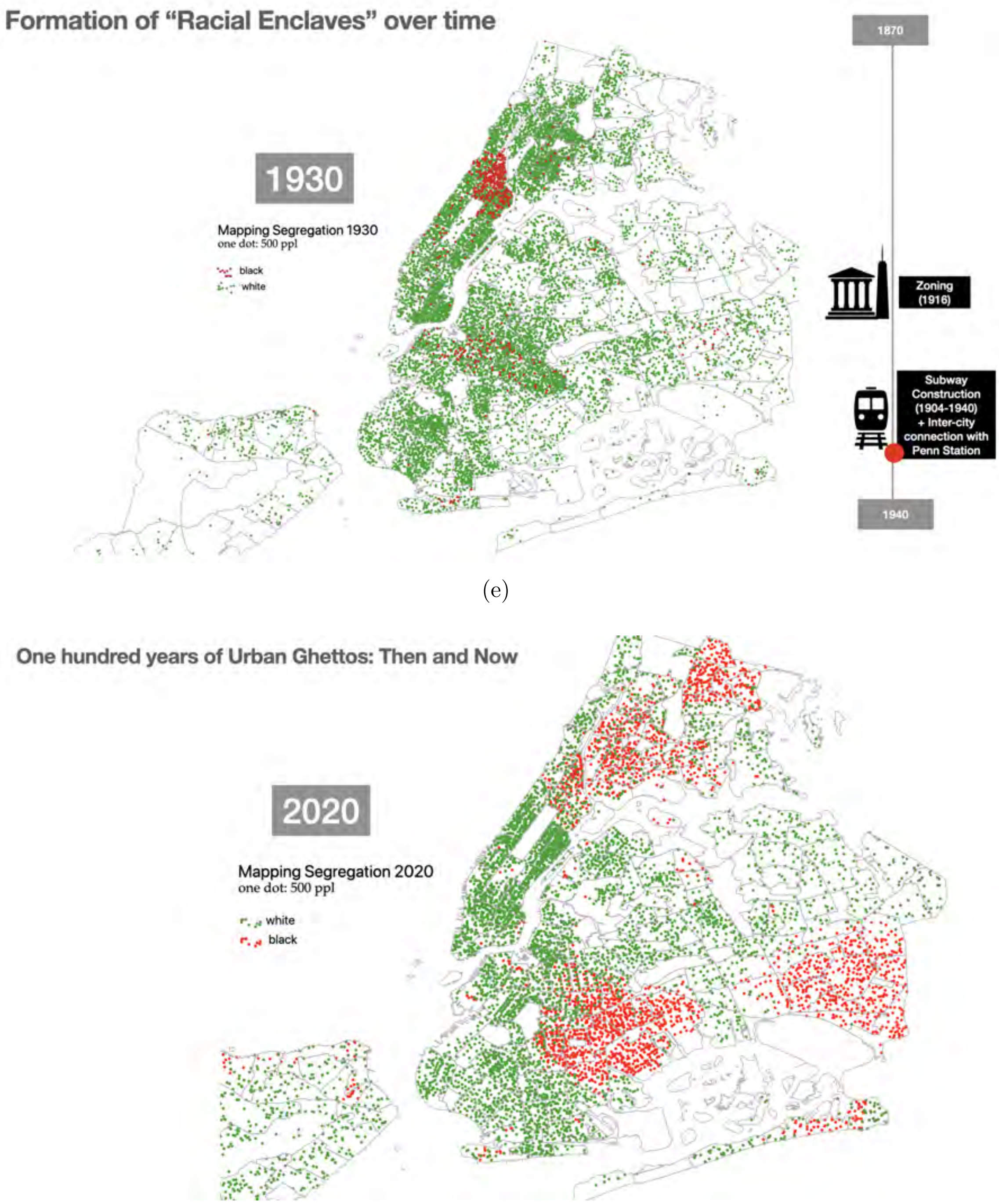


(c)



(d)

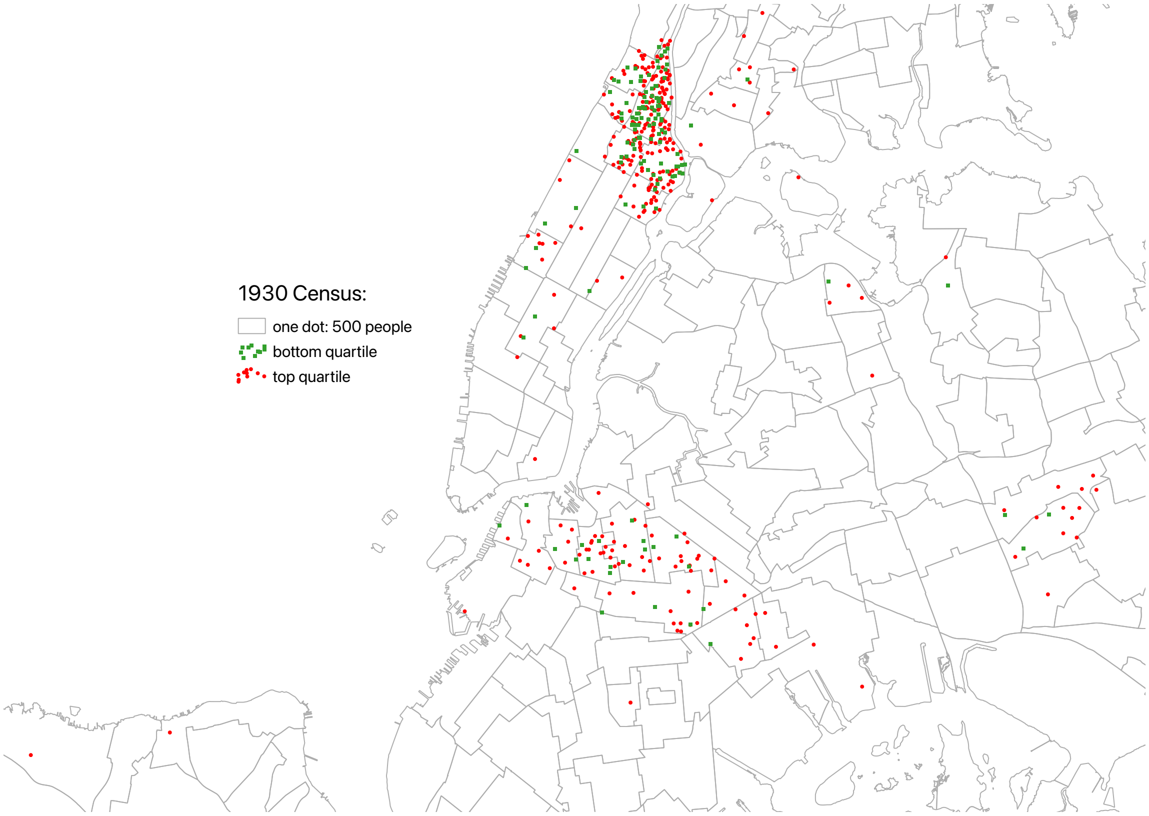
Figure 28: Mapping Residential Segregation by Race



(e)

(f)

Figure 29: Residential location of African Americans by income



D Regression Results

D.1 Population

Table 1: Treatment Effects for the Railway Network Arrival on Neighborhood Population in New York, 1870-1940

	(1)	(2)	(3)
	$\log(Population)_{it}$	$\log(Population)_{it}$	$\log(Population)_{it}$
$\beta_{\tau=40}$	-0.445	0.253	0.277
	(0.548)	(0.577)	(0.522)
$\beta_{\tau=30}$	0.851*	0.698	1.142**
	(0.438)	(0.450)	(0.452)
$\beta_{\tau=20}$	0.659*	0.533	0.917**
	(0.388)	(0.400)	(0.399)
$\beta_{\tau=10}$	1.257***	0.665*	1.330***
	(0.382)	(0.366)	(0.399)
$\beta_{\tau=0}$	–	–	–
$\beta_{\tau=-10}$	0.615	0.368	0.606
	(0.477)	(0.503)	(0.521)
$\beta_{\tau=-20}$	0.357	0.190	0.258
	(0.565)	(0.602)	(0.612)
$\beta_{\tau=-30}$	-0.317	-0.458	-0.559
	(0.511)	(0.527)	(0.518)
$\beta_{\tau=-40}$	1.093	0.853	0.947
	(0.664)	(0.689)	(0.698)
$\gamma_{\tau=40}$	–	-0.850	-4.169***
		(1.129)	(1.356)
$\gamma_{\tau=30}$	–	0.808	-2.534**
		(1.062)	(1.177)
$\gamma_{\tau=20}$	–	1.014	-1.972**
		(0.860)	(0.828)
$\gamma_{\tau=10}$	–	2.889***	-0.221
		(1.031)	(1.126)
$\gamma_{\tau=0}$	–	–	–
$\gamma_{\tau=-10}$	–	2.587***	1.708*
		(0.872)	(0.876)
$\gamma_{\tau=-20}$	–	2.181	3.035***
		(1.390)	(0.959)
$\gamma_{\tau=-30}$	–	1.717	4.435***
		(1.731)	(1.498)
$\gamma_{\tau=-40}$	–	1.347	2.615*
		(2.470)	(1.531)
Central City Defintion		Manhattan only	Within 5 km from Wall Street
Neighborhood FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	1365	1365	1365
R-squared	0.829	0.839	0.834

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

D.2 Segregation

D.2.1 Segregation (Dissimilarity Index) by Transit

Table 2: Treatment Effects for the Railway Network Arrival on Residential Segregation in New York

	(1)	(2)	(3)
	$\log(Dissimilar)_{it}$	$\log(Dissimilar)_{it}$	$\log(Dissimilar)_{it}$
$\beta_{\tau=40}$	0.019***	0.024***	0.017**
	(0.007)	(0.008)	(0.008)
$\beta_{\tau=30}$	0.017***	0.014**	0.015***
	(0.005)	(0.005)	(0.006)
$\beta_{\tau=20}$	0.011***	0.007*	0.010**
	(0.004)	(0.004)	(0.004)
$\beta_{\tau=10}$	0.009**	0.009**	0.007*
	(0.004)	(0.004)	(0.004)
$\beta_{\tau=0}$		–	–
$\beta_{\tau=-10}$	0.004	0.003	0.002
	(0.004)	(0.004)	(0.004)
$\beta_{\tau=-20}$	-0.005	-0.010	-0.008
	(0.006)	(0.006)	(0.008)
$\beta_{\tau=-30}$	-0.009	-0.010	-0.013
	(0.010)	(0.012)	(0.014)
$\beta_{\tau=-40}$	-0.003	-0.005	0.001
	(0.011)	(0.011)	(0.013)
$\gamma_{\tau=40}$	–	-0.005	0.015
		(0.014)	(0.014)
$\gamma_{\tau=30}$	–	0.014	0.016**
		(0.011)	(0.008)
$\gamma_{\tau=20}$	–	0.031***	0.005
		(0.008)	(0.006)
$\gamma_{\tau=10}$	–	0.001	0.008
		(0.009)	(0.005)
$\gamma_{\tau=0}$	–	–	–
$\gamma_{\tau=-10}$	–	0.007	0.007
		(0.011)	(0.007)
$\gamma_{\tau=-20}$	–	0.025**	0.010
		(0.013)	(0.009)
$\gamma_{\tau=-30}$	–	0.011	0.010
		(0.014)	(0.013)
$\gamma_{\tau=-40}$	–	0.004	-0.014
		(0.013)	(0.012)
Central Location Defintion		Manhattan only	Within 5 km from Wall Street
Neighborhood FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	920	920	920
R-squared	0.774	0.780	0.776

D.2.2 Segregation (Dissimilarity Index) by Transit & Zoning

Table 3: Treatment Effects for the Railway Network Arrival & "Zoning" on Residential Segregation in New York

	(1)	(2)	(3)
	$\log(Dissimilar)_{it}$	$\log(Dissimilar)_{it}$	$\log(Dissimilar)_{it}$
$\beta_{\tau=40}$	0.007***	0.007***	0.007***
	(0.002)	(0.001)	(0.001)
$\beta_{\tau=30}$	0.011**	0.010*	0.011**
	(0.003)	(0.004)	(0.003)
$\beta_{\tau=20}$	0.007	0.006	0.007
	(0.003)	(0.003)	(0.004)
$\beta_{\tau=10}$	0.003	0.003	0.004
	(0.004)	(0.004)	(0.003)
$\beta_{\tau=0}$	–	–	–
$\beta_{\tau=-10}$	0.001	-0.000	0.002
	(0.002)	(0.003)	(0.003)
$\beta_{\tau=-20}$	-0.003	-0.002	-0.002
	(0.007)	(0.007)	(0.006)
$\beta_{\tau=-30}$	-0.008	-0.009	-0.008
	(0.011)	(0.009)	(0.008)
$\beta_{\tau=-40}$	-0.007	-0.007	-0.005
	(0.005)	(0.006)	(0.006)
[Zoning: Land Use-Regulatoin]			
Business District	0.014**	0.013*	0.011*
	(0.004)	(0.005)	(0.005)
Business District Restricted Against Industry	0.029***	0.029**	0.024**
	(0.004)	(0.007)	(0.008)
Residential District	0.026***	0.026**	0.024**
	(0.004)	(0.007)	(0.007)
Undetermined Area	0.011	0.013	0.012
	(0.008)	(0.011)	(0.012)
Unrestricted District (Baseline)	–	–	–
[Zoning: Building Height]			
1 (Baseline)	–	–	–
1.25		0.013	0.015*
		(0.009)	(0.006)
1.5		-0.000	0.002
		(0.004)	(0.003)
2		0.006	0.009
		(0.009)	(0.005)
[Zoning: Land Area]			
A(Baseline)	–	–	–
B			0.011
			(0.008)
C			0.008*
			(0.003)
D			0.012**
			(0.003)
E			0.036
			(0.003)
Neighborhood FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	920	920	920
R-squared	0.584	0.594	0.603

D.2.3 Segregation (Isolation Index of African Americans) by Transit & Zoning

Table 4: Treatment Effects for the Railway Network Arrival & "Zoning" on Residential Segregation in New York

	(1) $\log(Isolation)_{it}$	(2) $\log(Isolation)_{it}$	(3) $\log(Isolation)_{it}$
$\beta_{\tau=40}$	-0.007	-0.007	-0.008
	(0.005)	(0.006)	(0.007)
$\beta_{\tau=30}$	0.004	0.003	0.003
	(0.014)	(0.013)	(0.012)
$\beta_{\tau=20}$	-0.008	-0.008	-0.007
	(0.005)	(0.005)	(0.004)
$\beta_{\tau=10}$	-0.007	-0.007	-0.006
	(0.004)	(0.004)	(0.003)
$\beta_{\tau=0}$	–	–	–
$\beta_{\tau=-10}$	0.003	0.003	0.004
	(0.007)	(0.007)	(0.007)
$\beta_{\tau=-20}$	-0.007	-0.008	-0.009
	(0.007)	(0.009)	(0.010)
$\beta_{\tau=-30}$	-0.001	-0.002	-0.002
	(0.008)	(0.012)	(0.012)
$\beta_{\tau=-40}$	-0.009	-0.011	-0.012
	(0.012)	(0.015)	(0.014)
[Zoning: Land Use-Regulatoin]			
Business District	0.017	0.019	0.018
	(0.009)	(0.011)	(0.010)
Business District Restricted Against Industry	0.017**	0.020*	0.017
	(0.004)	(0.007)	(0.008)
Residential District	0.012**	0.014**	0.013**
	(0.003)	(0.004)	(0.004)
Undetermined Area	0.006	0.008	0.007
	(0.006)	(0.010)	(0.010)
Unrestricted District (Baseline)	–	–	–
[Zoning: Building Height]			
1 (Baseline)	–	–	–
1.25		-0.008	-0.004
		(0.013)	(0.011)
1.5		0.005	0.010
		(0.013)	(0.011)
2		0.004	0.009
		(0.014)	(0.012)
[Zoning: Land Area]			
A(Baseline)	–	–	–
B			0.006*
			(0.003)
C			0.009**
			(0.003)
D			0.011
			(0.008)
E			0.022**
			(0.007)
Neighborhood FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	652	652	652
R-squared	0.054	0.061	0.064

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

D.3 House Price Index

As discussed in Section C.2, the digitized house sales data had about 450,000 transactions which were reported at a daily frequency (and published weekly) for 70 years. After geocoding each property's address, I map the property address into a time-consistent neighborhood boundary. By harmonizing each property's record by neighborhood and year, I can construct hedonic house sales price indexes by neighborhood and by year.

Table 5: Treatment Effects for the Railway Network Arrival & "Zoning" on Housing Prices in New York

	(1)	(2)	(3)	(4)	(5)	(6)
	$\log(HP)_{it}$	$\log(HP)_{it}$	$\log(HP)_{it}$	$\log(HP)_{it}$	$\log(HP)_{it}$	$\log(HP)_{it}$
$\beta_{\tau=50}$	-0.135	-0.130	0.275*	0.407**	-0.135	-0.134
	(0.193)	(0.168)	(0.123)	(0.106)	(0.148)	(0.108)
$\beta_{\tau=40}$	0.086	0.102	0.606**	0.663**	-0.026	-0.015
	(0.122)	(0.150)	(0.202)	(0.184)	(0.155)	(0.203)
$\beta_{\tau=30}$	0.201***	0.208***	0.331**	0.377**	0.158***	0.163**
	(0.034)	(0.039)	(0.104)	(0.083)	(0.025)	(0.055)
$\beta_{\tau=20}$	0.240**	0.238*	0.307*	0.343*	0.231**	0.233
	(0.056)	(0.094)	(0.112)	(0.135)	(0.083)	(0.120)
$\beta_{\tau=10}$	0.076	0.074	-0.006	0.011	0.078	0.076
	(0.069)	(0.035)	(0.083)	(0.074)	(0.071)	(0.039)
$\beta_{\tau=0}$	–	–	–			
$\beta_{\tau=-10}$	0.083	0.087**	-0.057	-0.037	0.031	0.033
	(0.040)	(0.021)	(0.046)	(0.043)	(0.038)	(0.034)
$\beta_{\tau=-20}$	0.116	0.116**	0.014	0.041	0.091	0.094
	(0.058)	(0.034)	(0.067)	(0.063)	(0.066)	(0.052)
$\beta_{\tau=-30}$	0.227**	0.221***	0.126	0.147***	0.172	0.187*
	(0.070)	(0.039)	(0.065)	(0.030)	(0.098)	(0.074)
$\beta_{\tau=-40}$	0.223**	0.212**	0.089	0.104	0.211	0.205
	(0.075)	(0.063)	(0.079)	(0.062)	(0.104)	(0.098)
$\beta_{\tau=-50}$	0.118*	0.082	-0.025	-0.007	0.199***	0.149*
	(0.044)	(0.048)	(0.200)	(0.171)	(0.034)	(0.058)
[**Zoning: Land Use-Regulatoin**]						
Business District	0.300**	0.305**	0.318**	0.336**	0.363*	0.357**
	(0.108)	(0.101)	(0.086)	(0.074)	(0.141)	(0.127)
Residential District	0.206	0.212	0.259	0.277	0.296	0.285
	(0.197)	(0.177)	(0.192)	(0.171)	(0.229)	(0.208)
Undetermined Area	0.398	0.423*	0.381	0.409*	0.490	0.487*
	(0.237)	(0.191)	(0.206)	(0.148)	(0.259)	(0.209)
Unrestricted District (Baseline)	–	–	–			
[**Zoning: Building Height**]						
1 (Baseline)	–	–	–			
1.25	0.432***	0.227	0.499***	0.292**	0.398***	0.181
	(0.083)	(0.128)	(0.046)	(0.091)	(0.080)	(0.130)
1.5	0.269**	0.069	0.327***	0.126	0.263**	0.037
	(0.059)	(0.111)	(0.038)	(0.063)	(0.061)	(0.118)
2	0.891***	0.690***	0.963***	0.766***	0.926***	0.693***
	(0.041)	(0.069)	(0.078)	(0.047)	(0.056)	(0.095)
[**Zoning: Land Area**]						
A(Baseline)	–	–	–			
B	0.546	0.536	0.641	0.677	0.547	0.519
	(0.378)	(0.397)	(0.429)	(0.446)	(0.405)	(0.431)
C	0.754	0.523	0.842	0.649	0.739	0.490
	(0.492)	(0.505)	(0.549)	(0.559)	(0.522)	(0.545)
D	0.850*	0.619	0.957*	0.777	0.860	0.600
	(0.385)	(0.409)	(0.404)	(0.410)	(0.420)	(0.465)
E	1.324**	1.084*	1.475**	1.339**	1.258*	0.961
	(0.420)	(0.442)	(0.434)	(0.415)	(0.482)	(0.531)

THIS TABLE TO BE CONTINUED

Table 6: Treatment Effects for the Railway Network Arrival & "Zoning" on House Price Index in New York (Ctnd)

	(1) $\log(HP)_{it}$	(2) $\log(HP)_{it}$	(3) $\log(HP)_{it}$	(4) $\log(HP)_{it}$	(5) $\log(HP)_{it}$	(6) $\log(HP)_{it}$
$\gamma_{\tau=50}$			0.000	0.000	0.031*	0.034
			(.)	(.)	(0.013)	(0.063)
$\gamma_{\tau=40}$			-0.246	-0.168	0.354**	0.351*
			(0.178)	(0.218)	(0.089)	(0.127)
$\gamma_{\tau=30}$			0.199*	0.283*	0.205***	0.206***
			(0.087)	(0.123)	(0.043)	(0.014)
$\gamma_{\tau=20}$			0.229	0.314	0.069	0.063
			(0.189)	(0.251)	(0.116)	(0.116)
$\gamma_{\tau=10}$			0.473***	0.587***	0.041	0.053
			(0.030)	(0.088)	(0.103)	(0.104)
$\gamma_{\tau=0}$	–	–	–	–	–	–
$\gamma_{\tau=-10}$			0.599**	0.718*	0.439	0.443
			(0.202)	(0.273)	(0.294)	(0.295)
$\gamma_{\tau=-20}$			0.552***	0.645***	0.317**	0.310**
			(0.082)	(0.103)	(0.082)	(0.075)
$\gamma_{\tau=-30}$			0.474***	0.575***	0.247*	0.194*
			(0.030)	(0.106)	(0.106)	(0.088)
$\gamma_{\tau=-40}$			0.747***	0.845***	0.099	0.102
			(0.061)	(0.104)	(0.100)	(0.085)
$\gamma_{\tau=-50}$			0.854***	0.811***	-0.076	-0.020
			(0.055)	(0.103)	(0.055)	(0.033)
log(% African Americans)		-0.008		-0.009		0.002
		(0.017)		(0.019)		(0.023)
Central Location Defintion			Manhattan only	Manhattan only	< 5 km from Wall Street	< 5 km from Wall Street
Borough FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3403.000	3327.000	3403.000	3327.000	3403.000	3327.000
R-squared	0.402	0.412	0.411	0.420	0.410	0.419

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Estimated treatment effects from the railway arrival station on log hedonic house sales price index (by neighborhood); the sample includes time-consistent 195 neighborhoods in New York City (made up with 5 boroughs) in digitized real estate data from 1870 to 1940; The original digitized house sales data had about 450,000 house sales transactions which was reported at a daily frequency (and published weekly) for 70 years).

D.4 Urban Wage Premium from Panel Data

$$\log \text{Earnings}_{iht} = \alpha_i + \beta X'_{it} + d_t + Household_h + \epsilon_{iht} \tag{10}$$

where i denote an individual i from a household h at time t; $earnings_{iht}$ denote individual i from household h's earning at time t, α_i is individual i's time-invariant characteristics, β is a time-invariant vector of coefficients of a characteristic's marginal contribution to individual i's earnings, $Household_h$ is the household fixed effects that the individual (i) belongs to and all members in the household h may share that could explain one's individual i's earnings (e.g. diligence, personality traints, health characteristics that may be common among household members), and d_t is a time fixed effect. The unobserved heterogeneityϵ_{iht} is assumed to be normally distributed.

Table 7: Urban Premium for Rural to NY migrants

		White Men					Black Men			
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
age	0.017***	0.016***	0.020***	0.017***	0.031***	0.033***	0.026***	0.026***	0.027***	0.043***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.013)
age^2	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
[Migration Decision]										
NYC leaver to NY metro	-0.020***	-0.020***	-0.018***	-0.005**	0.005	0.010	0.009	0.009	0.011	-0.047
	(0.003)	(0.003)	(0.003)	(0.002)	(0.005)	(0.020)	(0.019)	(0.019)	(0.018)	(0.049)
NYC leaver to non-NY metro	-0.061***	-0.060***	-0.044***	-0.016***	0.005	-0.053***	-0.053***	-0.052***	-0.029***	-0.014
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.009)	(0.009)	(0.009)	(0.008)	(0.043)
NYC entrant from NY Metro	-0.052***	-0.006**	0.011***	0.003	0.008	-0.042	-0.005	-0.005	-0.006	0.028
	(0.003)	(0.003)	(0.003)	(0.003)	(0.007)	(0.027)	(0.027)	(0.027)	(0.022)	(0.054)
NYC entrant from non-NY Metro	-0.160***	-0.070***	-0.046***	-0.013***	-0.014***	-0.244***	-0.034***	-0.034***	0.000	-0.021
	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.009)	(0.009)	(0.010)	(0.009)	(0.036)
NYC stayer + Neighborhood mover	-0.018***	-0.018***	-0.008***	0.002	0.011***	0.001	-0.000	-0.000	0.006	0.021
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.009)	(0.009)	(0.009)	(0.008)	(0.037)
Neighborhood (+ NYC) stayer (baseline)	–	–	–	–	–	–	–	–	–	–
[Urban]										
Rural (baseline)	–	–	–	–	–	–	–	–	–	–
Urban		0.358***	0.373***	0.022***	0.058***		0.409***	0.409***	-0.000	0.057**
		(0.003)	(0.003)	(0.002)	(0.007)		(0.009)	(0.009)	(0.009)	(0.028)
[Nativity]										
Native-Born: Both parents native-born	–	–	–	–	–	–	–	–	–	–
Native-Born: Father foreign, mother native			0.028***	0.009***	0.010			0.035	0.039	0.087
			(0.003)	(0.003)	(0.007)			(0.091)	(0.106)	(0.069)
Native-Born: Mother foreign, father native			-0.005	0.005	0.006			0.023	0.080*	0.271***
			(0.005)	(0.005)	(0.009)			(0.056)	(0.046)	(0.048)
Native-Born: Both parents foreign			-0.026***	-0.016***	-0.003			0.215**	0.187***	0.358
			(0.002)	(0.001)	(0.003)			(0.098)	(0.063)	(0.254)
Foreign-Born			-0.102***	-0.067***	-0.013***			-0.001	-0.016**	-0.031
			(0.001)	(0.001)	(0.003)			(0.007)	(0.007)	(0.046)
log(Father's Earnings)					0.094***					-0.008
					(0.004)					(0.040)
IND1950 fixed effects	No	No	No	Yes	Yes	No	No	No	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	447856	447856	447856	352616	41336	17439	17439	17439	14788	1327
R-Sqaured	0.039	0.091	0.107	0.516	0.636	0.117	0.220	0.221	0.581	0.748

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$