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

Sun Kyoung Lee[†]
Yale University

Jan 31, 2022

Abstract

I examine the influence of the United States' first comprehensive land-use regulations ("zoning") and the growth of transportation infrastructure on segregation and intra-city migration in New York City from 1870 to 1940. Combining a new panel dataset derived from historical US federal population censuses with newly digitized real-estate sales transaction records, I find that between-neighborhood socioeconomic segregation increased dramatically after the building of transit infrastructure, that zoning was largely an endogenous response to the socioeconomic segregation facilitated by transit infrastructure and that the combination of zoning and building new subways in a rapidly expanding city produced a pattern where inner-city neighborhoods zoned for factories and multi-family dwellings 'flipped' to African American majorities. I estimate the value of a "tipping point" beyond which mixed-race neighborhoods are no longer sustainable and show that housing prices in tipped neighborhoods decrease by 40% relative to non-tipped neighborhoods. Finally, while white and African American migrants from the rural South to New York City benefit from urban wage premia of about 40%, this premium is reduced for African American migrants living in segregated minority neighborhoods.

1 Introduction: New York's Rise as the Megacity

What is the role of urban planning when cities grow? Processes of economic development in most countries are accompanied by urbanization —a spatial transformation process whereby the population moves from rural to urban locations (typically cities). This process is typically accompanied by urban planning policies designed to reduce congestion. Two canonical examples are subway construction and zoning restrictions: transit infrastructure reduces commuting time, while zoning limits the geographical scope of activities with large negative externalities (e.g. manufacturing that generates pollutants harmful to health).[1]

In this paper, I demonstrate that these two policies contributed to racial segregation by neighborhoods and a worsening of economic outcomes for African American migrants, relative to white counterparts, in America's biggest city during an era of unprecedented economic and population growth (1870-1940).

To do so, I use extensive new digitized data from original historical records. Using the neighborhood-level panel dataset (1870-1940) and a difference-in-differences methodology, I first establish that after the arrival of the subway, neighborhoods became segregated by income. Building subways facilitated socioeconomic segregation between neighborhoods by inducing a geographic sorting of households by income. As subways increased the geographic scope of the city, central locations (where most manufacturing and office jobs were concentrated) saw both a population decline and a substantial decline of residents' socioeconomic status.

Next, using cross-sectional analysis I show that the nation's first comprehensive zoning regulation in 1916 formed an endogenous response to neighborhood socioeconomic segregation as of 1910. Arguably, this segregation had been facilitated by subway building, assigning neighborhoods, for example, for single-family housing or for factories and multi-family dwellings. Zoning assignment was predominantly assigned according to residents' socioeconomic status and demographic composition, and further intensified segregation by inducing household sorting by income and race. Zoning assignment resulted, in very short order, in the formation of racial enclaves.

Shortly before zoning was implemented, in 1910, Central Harlem had 7% of African American residents. By 1930, this had increased to 96%. A natural question is whether the combination of building subways and zoning contributed to inner-city neighborhoods zoned for factories and multi-family dwellings flipping to African American majorities while white residents used new subways to move to neighborhoods zoned as residential districts in the city periphery. To answer this, I digitize transactional-level house sales records with actual sales price information and construct a neighborhood-level house sales price index spanning 1870-1940. Using housing sales

---

[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)).

price time-series, I test whether a structural break exists in the relationships between the housing sales price and time of transit access by neighborhood (Bai and Perron (1998)). In neighborhoods undergoing a tipping process, property prices decline by 40% relative to non-tipping neighborhoods, and property prices remain depressed.

Thus, a combination of urban planning policies and the neighborhood tipping process produced an extreme residential racial segregation where African Americans were "isolated" in few minority-neighborhoods. Finally, I investigate the impact of residential segregation on labor market outcomes using panel data of individuals that I construct using machine-learning record linking techniques. I find adverse effects for African Americans living in segregated minority neighborhoods. In my context, residential "isolation" of minorities against the majority population could mean that minorities' job opportunities are 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 40% earnings increases for the white and African American population). 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; these forces include exclusion from high-growth industries and extreme, residential segregation.[3]

As described above, an important innovation is the collection and harmonization of large, mostly newly-digitized data from original sources. To summarize, 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-based record linking methods to construct individual panel data; here, I link population census data without unique individual identifiers (e.g. social security number) using time-invariant information from the census.[5] Next, to

---

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

[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

measure the impact of urban planning policies, I collect zoning and subway planning documents and convert maps from the Archives into Geographic Information System (henceforth, GIS)-compatible shapefiles for performing spatial analyses on time-consistent neighborhoods. Finally, to observe 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 from 1870 to 1940. 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]

This paper also extends the literature on "tipping" and the dynamics of segregation (as in Card, Mas and Rothstein (2008)). I establish the connection between demographic composition change and housing sales prices. In particular, I first identify the neighborhoods that underwent a "tipping" process by documenting population changes and demographic shifts between African American and white populations. I document that when neighborhoods "tip", the following patterns are observed: 1. a significant increase in numbers of African Americans, 2. a disproportionate decrease of white original residents leading to net population loss, 3. the share of African Americans reaches a significant level (above 0.3). I use the high-frequency neighborhood-level panel data of housing sales prices to test if and when a structural break occurs. I show that Central Harlem North was the first neighborhood that underwent the "tipping" process both captured by demographic changes (in 1910, the same neighborhood had a 7% share of African Americans; and this number jumps to 50% by 1920, and 96% by 1930, reaching an almost 100% segregation), and housing sales price data with its structural break in 1917. Additionally, I document that geographical contiguity matters in understanding why neighborhoods tip. As Central Harlem North went through a tipping process from 1910, in the following decade, nearby neighborhoods began exhibiting similar demographic composition shifts -and structural breaks in terms of house sales prices in these neighborhoods are identified shortly after 1917. Finally, I also show how housing prices change in "tipping" neighborhoods—housing prices decrease by 40% compared to non-tipping neighborhoods throughout the dynamic "tipping" process, and the housing market continues to be depressed over time, perhaps due to hysteresis.

This may be due to a dramatic population loss —I document that the inflow of one African American in tipping neighborhoods is associated with an outflow of 2.5 whites, leading to a net population loss. Therefore, a dramatic population loss in "tipping" neighborhoods could explain 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.

decrease in 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.

This paper is also related to previous research on residential segregation. Boustan (2010) 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.[7] Cutler, Glaeser and Vigdor (1999) document changes in racial segregation across U.S. cities and find that the period 1890 to 1940 saw the birth of ghetto in the United States, accompanied perhaps by African American migration from the rural South to the North. However, Logan and Parman (2015) derive 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 the aforementioned works, my paper finds that the 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]

Relatedly, a growing strand of research studies the impact of the US government's discriminatory policies and practices. One strand of research 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) suggest that racial bias in the construction of the HOLC

---

[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 migrants 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").

maps can explain at *most* 4 to 20 percent of the observed concentration of African American households in the lowest-rated ("redlined") zones. Fishback, LaVoice, Shertzer and Walsh (2020) also suggest that the majority of African American households were located in such zones *well* before the federal government's involvement in mortgage markets. My findings that 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.

My paper also contributes to the literature of social (i.e. non-market) interactions, which are not regulated by the price mechanism, among individuals. Scheinkman (2008) 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 (2008)).[10] Relative to the existing literature on social interactions, my paper studies the dynamics of segregation using high-frequency housing sales data. It captures how social interactions could lead to sudden demographic composition changes and housing market crashes.

The literature on land-use regulation. Gyourko and Molloy (2015) review the existing literature on local government regulation of urban land and concludes that most studies have found substantial effects on the housing market. Only a handful of papers on zoning study effects *beyond* the housing market.[11] My paper shows that zoning was an endogenous response to the segregation pattern that subways created by enabling the geographical sorting of households by income. The local segregation dynamics that I analyze in the Harlem area also show that zoning assignment changed "who lived where" by influencing the location choices of households of different races and incomes, intensifying segregation even further.

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

---

[10]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 an 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).

[11]Kahn, Vaughn and Zasloff (2010) is a notable exemption that makes a connection between residential regulation and sorting. They, for instance, look at 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.

growth (Duranton and Puga (2014); Redding and Turner (2015)).[12] Relative to most papers that study the role of transit infrastructure investments, my paper focuses on ''unanticipated'' consequences —segregating neighborhoods by income and race. Related to this, 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 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) (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.[13] For example, I digitize a wealth of spatial data from 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 new set of questions based on the interaction of datasets. For instance, residential segregation was studied very thoroughly on its own (as in Logan and Parman (2015)). However, as I combine racial segregation data and public policies with a spatial dimension (i.e. neighborhood-variations of zoning), I can establish the impact of residential zoning on residential racial 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 the long-run ''unanticipated'' impacts of public

---

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

[13]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 locations 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 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 underrepresented in linked datasets, better linking may reduce selection bias coming from the common characteristics.

policies *well* before redlining.

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

2 Data and Background: New York City, 1870-1940

In this section, I discuss the main sources of data and relevant background that is essential in understanding New York City's urban planning and segregation. 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 high-frequency novel transaction-level real estate data that I digitize from the archives. I augment other sources of data including major public policies (i.e. subways and zoning); augmentation also includes the construction of panel data that follow the same individuals over time. Further details are discussed in the Appendix.

2.1 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]

---

[14]The first underground line of the subway opened in 1904, almost 40 years after the opening of the first elevated

Figure 1: Evolution of Spatial Links by the Intra- and Inter-city Transit Network
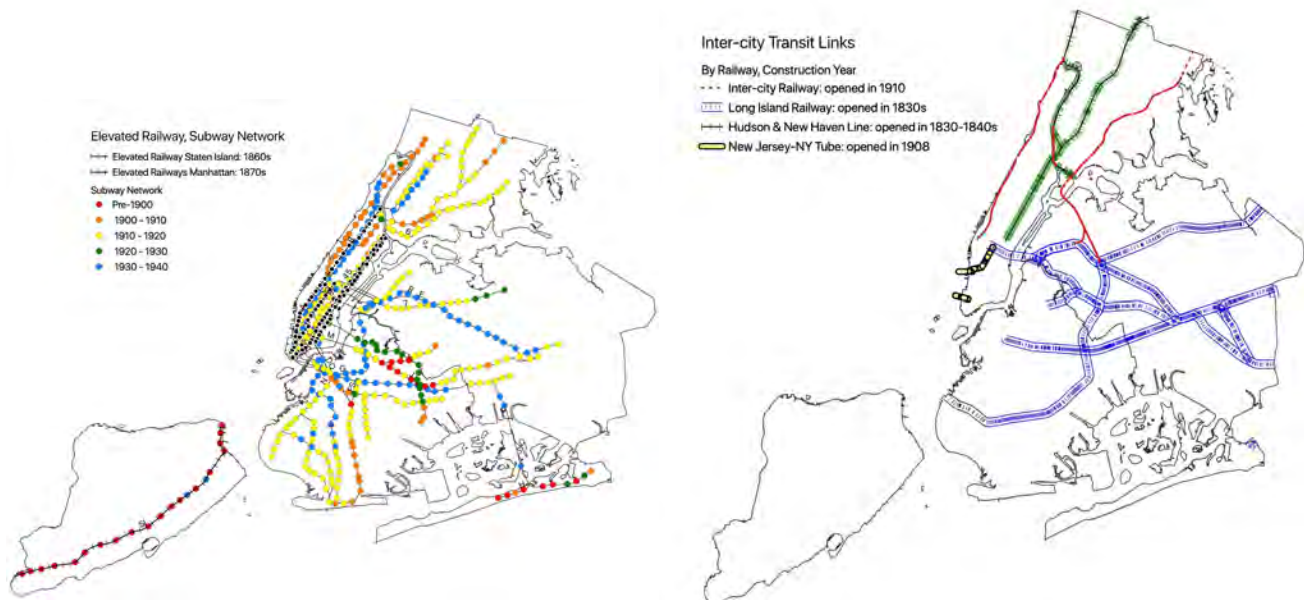
Note: The above figures show the evolution of within and between city spatial links in terms of the elevated railways and subways (for within-city transit network, Left figure) and railroad network (for between-city transit network, Right figure) over the study period. For within-city transit network (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/).

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 1).[15] As seen in Figure 1, inter-city transit 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.

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

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))[17]

---

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 (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

[17]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*

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 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)).

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 enable me to document zoning-regulations that were in place down to the street and building level. Details of these procedures and sources are available in the Appendix Paragraph B.2 and Figure 23, 24, and 26(b).

2.3   Population Data & Trend

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.[18] 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 in various data sources to time-consistent boundaries.[19] See the Appendix Section B.2 for details regarding this procedure.

---

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)). 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.

[18]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.

[19]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

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 individuals 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) evaluate different automated methods for record linking, performing a series of comparisons across methods and against hand linking. I largely follow the standard machine-learning record linking methodology suggested by Goeken, Huynh, Lynch and Vick (2011). Yet, I have extended the techniques of Goeken, Huynh, Lynch and Vick (2011) by inventing a two-step machine learning matching methodology that make use of household-level information. See the Appendix Section A for details of the census record linking.

Population Trend that the New Neighborhood-level Panel Data Reveals The total population of New York City increased from 1.48 million to 7.5 million during my study period (1870-1940);[20] 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).

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. 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*. Figure 2 reveals 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 2 shows mean occupational income

---

[20]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.

trends for all employed men and women aged between 16-60 at varying geographic scopes.

Figure 2: Mean Occupational Income Trend: Men and Women
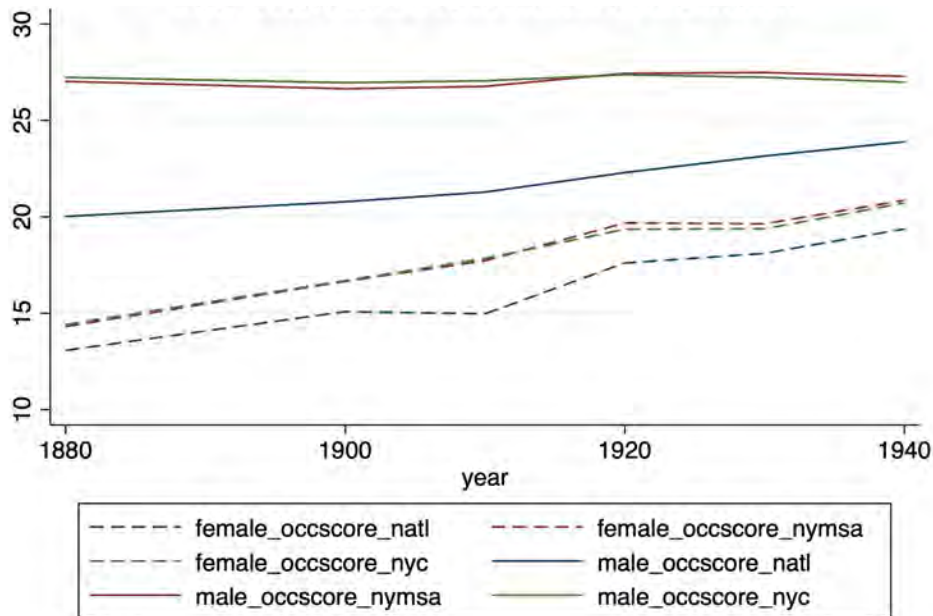
Note: Solid lines indicate men, whereas dotted lines indicate women; in terms of geography, the national average is in Blue, New York metro area average is in Red, and New York City average is in Green.

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

2.4 Urban Land Use

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) pioneer 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.[21]

---

[21]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

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 conveyed in bona fide sales.[22] 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 3 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.[23] Starting 1905, the records also began listing assessed values for each property. As Figure 3 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.[24]

---

regressions on bundled transacted properties with one price.

[22]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.''

[23]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.

[24]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

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 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.[25] See Section C.2 for more details on real estate records.

Figure 3: Real Estate Transaction Record Examples (1871 and 1922)

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".

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,

---

[25]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).

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.[26]

I digitize maps that feature the location of jobs by industry to accurately document the primary location of jobs in the city. Figure 4, 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.

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 5 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.[27]

---

[26]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.

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

Figure 4: Location of Women's Clothing Industry (1900 - 1922)
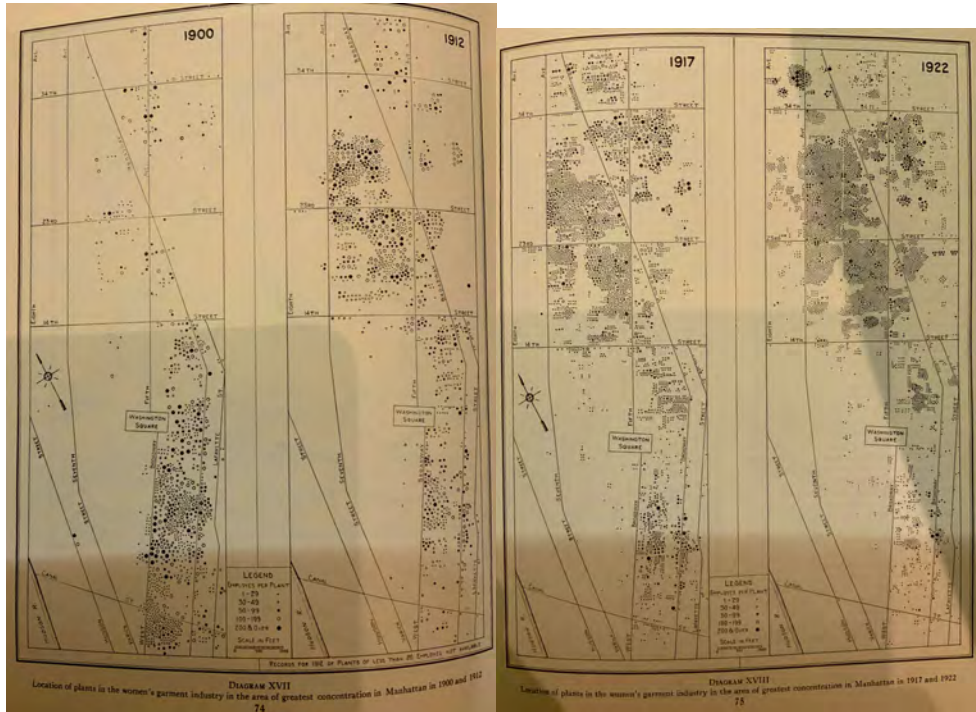
Note: The above figure shows the location of plants in women's garment industry in New York in 1900, 1912, 1917, and 1922. It reveals that the women's garment industry tends to concentrate in one central district. The industrial survey states that this tendency toward concentration is due in part to 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))

Figure 5: Home and Work Locations from City Directory
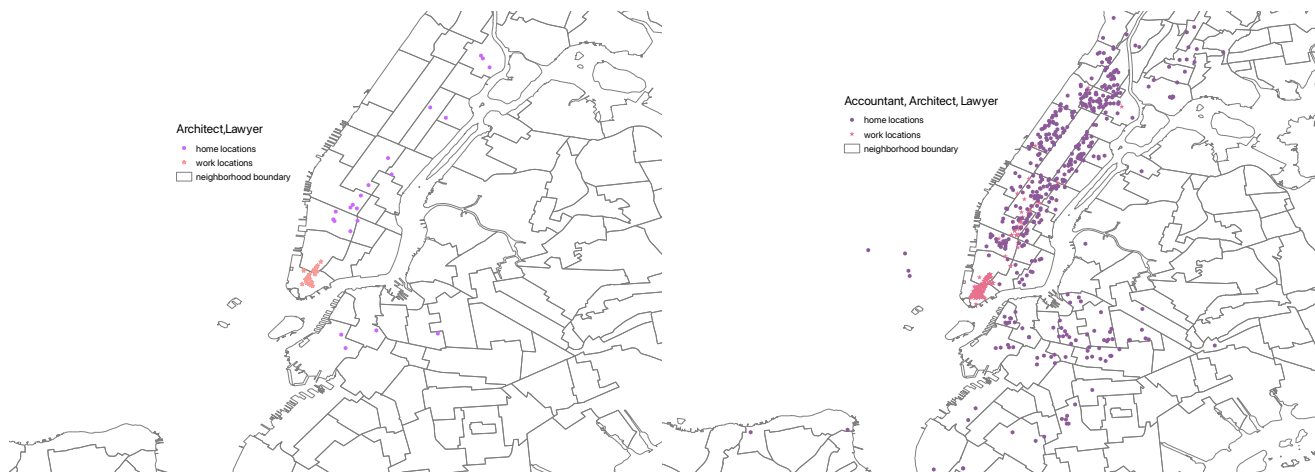
Note: The above figures show the commuting patterns by plotting home and work locations for the same category of jobs (i.e. accountants, architects and lawyers) in 1879 (Left) and 1906 (Right). In 1879, in the absence of rapid commuting transit infrastructure, home and work locations were not entirely 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.

3 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 assume 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. $m > m^*$), $\frac{\partial b^w(n^w, m)}{\partial m} < 0$.

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.

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 $b^m(m, m) = b^w(1 - m, m)$. 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.[28] Card, Mas and Rothstein (2008) show 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)

---

[28]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)*".

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.
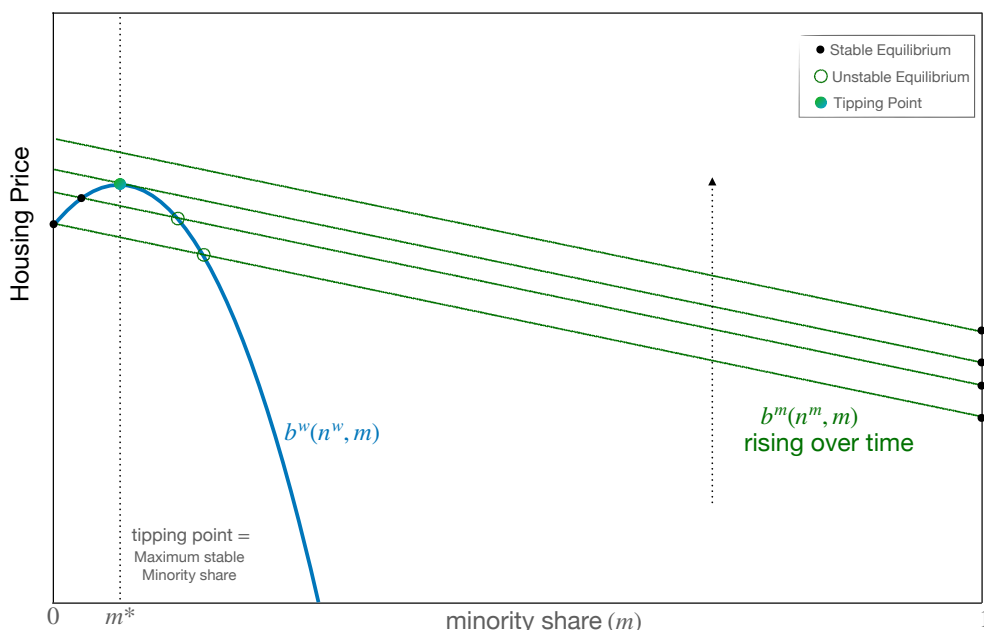
Finally, Card, Mas and Rothstein (2008) note that the value of the "tipping point" (m^*) depends on the white/majority population's tolerance for growing minority neighbors. If the majority's aversion toward minority is "strong" (captured by the sharp negative slope of $\dfrac{\partial b^w}{\partial m}$), then even at a relatively low level of minority share (m), the neighborhood could undergo the tipping process. Relatedly, Card, Mas and Rothstein (2008) discuss 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.

3.1 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) state 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 4 and 5) is a natural step.

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) note that the location of tipping point depends on the strength of white/majority distaste for minority neighbors (i.e. $\frac{\partial b^w}{\partial m}$)

4 Reduced-Form Evidence

4.1 Staggered Triple 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, residents' socioeconomic

status, and segregation as it relates to the arrival of transit infrastructure. The baseline specification is as follows:

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

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 residential population, residents' socioeconomic status, 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; 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)). 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.

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

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 outcome variable, depending on neighborhoods being located in central locations or not.

4.2 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 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 3. 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 occurred 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). In my main analysis, I define "city core" as locations within 5 kilometers of the Battery which is the Southern tip of Manhattan. 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 2.4 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 manufacturing 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).[29] This definition of the city core captures the Central Business District where commercial and business activities are highly concentrated.

As in Figure 7, the heterogeneous treatment effects (γ_τ from equation 3) between neighborhoods in the city core and city periphery reveal that central locations 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 population from

---

[29]Alternatively, I also define "city core" as all locations in Manhattan county, and report the estimated coefficients under this definition in Appendix Table 3.

the neighborhood-specific time trends shortly after the railway arrival which is an exact antithesis from the trends before the railway arrival.
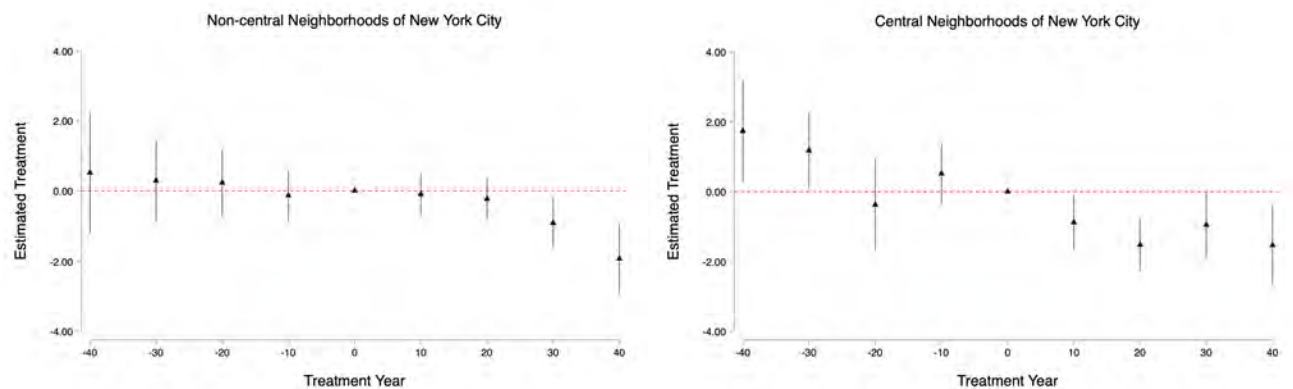
Figure 7: Population Changes
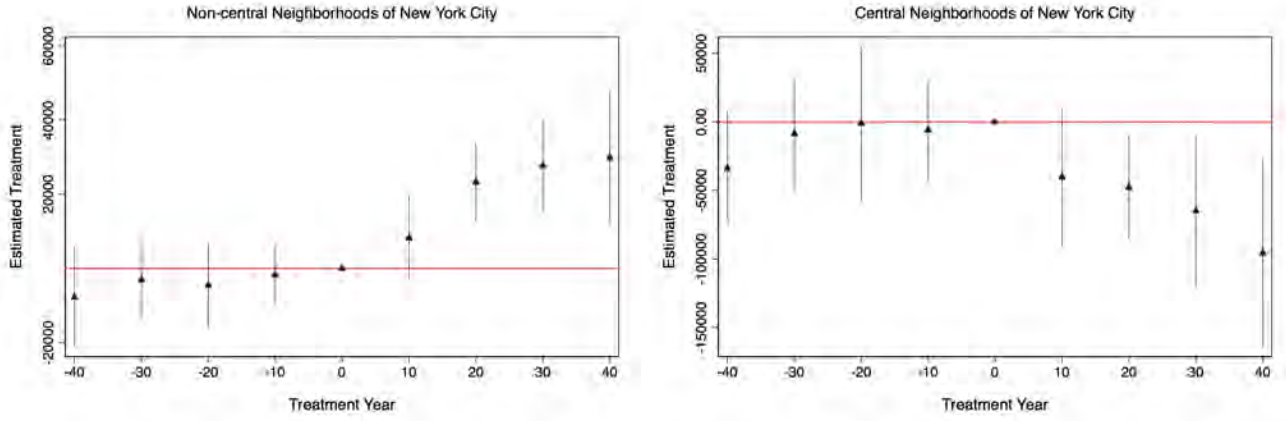
Note: Estimated treatment effects from the railway station arrival on log neighborhood population; the sample includes 195 time-consistent neighborhoods in New York City (made up of 5 boroughs) and comes from population censuses from 1870 to 1940 (every 10 years, except for 1890 where the original census was lost due to fire); all specifications include neighborhood fixed effects, year fixed effects (the estimated coefficients and standard errors are reported in Appendix Table 3).

The left figure shows estimated treatment effects (β_τ from equation 3) for neighborhoods outside the city core; the right figure shows the heterogeneous treatment effects (γ_τ from equation 3) between neighborhoods in the city core and city periphery. The vertical lines in both left and right figure show the estimated 95% confidence intervals. I define neighborhoods to be located in the city core if a neighborhood is located within 5km from the Battery.

4.3   Socioeconomic Status (SES) Changes

I follow the regression specification as in equation 3 which allows for heterogeneous treatment effects, with residents' socioeconomic status as an outcome variable. Figure 8 shows statistically significant and negative heterogeneous treatment effects (γ_τ from equation 3) of the railway arrival on residents' socioeconomics status. After the railway arrival, centrally located neighborhoods exhibit a substantial decline of residents' SES which amounts to 10-20% decresed of residents' SES. However, such patterns were not observed in non-centrally located neighborhoods.

Figure 8: Socioeconomic Status Changes

Note: Estimated treatment effects from the railway station arrival on residents' socioeconomic status; 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; all specifications include neighborhood fixed effects, year fixed effects (the estimated coefficients and standard errors are reported in Appendix Table 4). The left figure shows estimated treatment effects (β_τ from equation 3) for neighborhoods outside the city core; the right figure shows the heterogeneous treatment effects (γ_τ from equation 3) between neighborhoods in the city core and city periphery. The vertical lines in both left and right figure show the estimated 95% confidence intervals. I define neighborhoods to be located in the city core if a neighborhood is located within 5km from the Battery.

4.4   Segregation

Stratification of neighborhoods by income and race I show that there is a greater degree of residential "sorting" by income immediately after the railway arrival. I measure the mean earnings of residents 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) 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 two groups. The dissimilarity index for two groups in the city, Rich and Poor, for example, is measured as follows (Massey and Denton (1998)):

$$\text{Dissimilarity Index}_{it} = \frac{1}{2} \sum_{i=1}^{n} \left| \frac{r_i}{R} - \frac{p_i}{P} \right|$$

where n equals the number of historical census tracts within the neighborhood (i.e. aggregation of census tracts); r_i denotes the number of rich population in census tract i; R denotes the total rich population in the city; p_i denotes the number of poor population in census tract i that makes up the neighborhoood; P denotes the total poor 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 (that makes up the city) 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.3, I report staggered difference-in-differences specifications that establish the connection between the transit network and neighborhood stratification (by income and race). Figure 9 shows the estimated treatment effects (β_τ) from equation 2 and the 95% confidence intervals, with the full regression results in Appendix Table 5. 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)).

I follow the regression specification as in equation 3 which allows for heterogeneous treatment effects, with an outcome variable of dissimilarity index. Consistent with notations in equation 3, $\mathbb{I}_i^{center}$ 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 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).

Figure 10 on the right shows that segregation trends do not exhibit heterogeneous treatment effects (γ_τ) from the railway arrival (the full regression results in Appendix Table 5). While overall segregation increased after the railway arrival (as in Figure 9), this trend was largely driven by neighborhoods in the city periphery (as in Figure 10 on the left where it plots the estimated treatment effects (β_τ) of railway arrival on neighborhood segregation for city periphery neighborhoods).
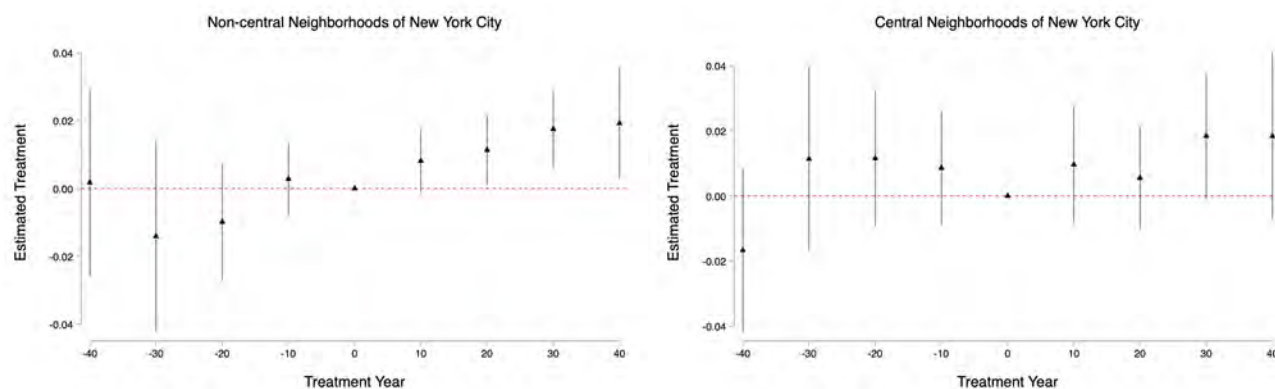
Figure 9: Estimated Treatment Effects (β_τ) of Railway Arrival on Neighborhood Segregation
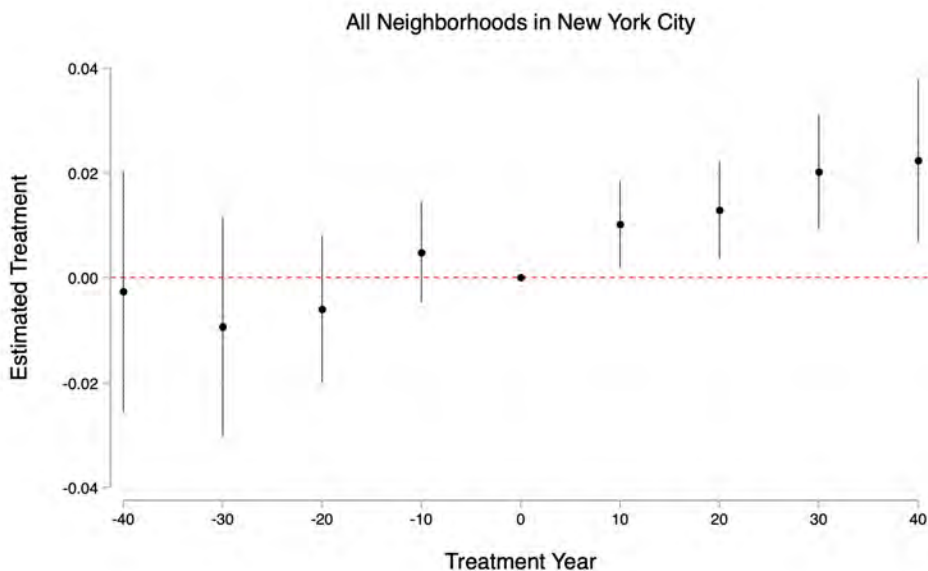
Note: Estimated treatment effects (β_τ from equation 2) from railway station arrival on neighborhood dissimilarity index; the sample includes 195 time-consistent neighborhoods in New York City (made up with 5 boroughs) and comes from population censuses from 1870 to 1940 (every 10 years, except for 1890 where the original census was lost due to fire); all specifications include neighborhood fixed effects, and year fixed effects; the estimated coefficients and standard errors are reported in Appendix Table 5.
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 τ capture the pre-treatment years. The figure shows estimated treatment effects (β_τ from equation 2) for neighborhoods. The vertical lines in the figure show the estimated 95% confidence intervals. Standard errors are clustered on boroughs.

Figure 10: Differential Segregation Trend

Note: Estimated treatment effects from the railway station arrival on residents' socioeconomic status; 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; all specifications include neighborhood fixed effects, year fixed effects (the estimated coefficients and standard errors are reported in Appendix Table 3).

The left figure shows estimated treatment effects (β_τ from equation 3) for neighborhoods outside the city core; the right figure shows the heterogeneous treatment effects (γ_τ from equation 3) between neighborhoods in the city core and city periphery. The vertical lines in both left and right figure show the estimated 95% confidence intervals. I define neighborhoods to be located in the city core if a neighborhood is located within 5km from the Battery.

Predicting Census-tract-level Zoning What predicts how (nation's first) zoning was assigned at the census tract level? I combine 1910 population census (latest population census before zoning) and type of zoning assignment at the historical census tract level and show how residents' demographic characteristics and socioeconomic status affected zoning assignment. Multinomial logistic regression predicts how zoning was assigned by residents' demographic composition (i.e. share of African Americans in the census tract) and socioeconomic status (based on residents' mean earnings, I grouped historical census tracts into tercile income groups; high, medium, and low). The zoning land-use assignment is the outcome variable which consists of three categories of land-use regulation (i.e. residential, business, and unrestricted). The output corresponds to the equations below:

$$\ln\left(\frac{P(zoning = residential)}{P(zoning = unrestricted)}\right) = b_{10} + b_{11}(share\,minority) + b_{12}(SES)$$

$$\ln\left(\frac{P(zoning = business)}{P(zoning = unrestricted)}\right) = b_{20} + b_{21}(share\,minority) + b_{22}(SES)$$

where b's are the regression coefficients.

Based on regression coefficients, Table 1 reports the relative risk ratio which is the ratio of the probability of choosing one outcome (i.e. business or residential) category over the probability of choosing the baseline ("unrestricted") category. It reveals that census tracts with higher share of minority population would be much less likely assigned for residential districts; high-income

neighborhoods are disproportionately more likely to be assigned for residential districts.

Figure 11: Predicting census-tract-level zoning assignment
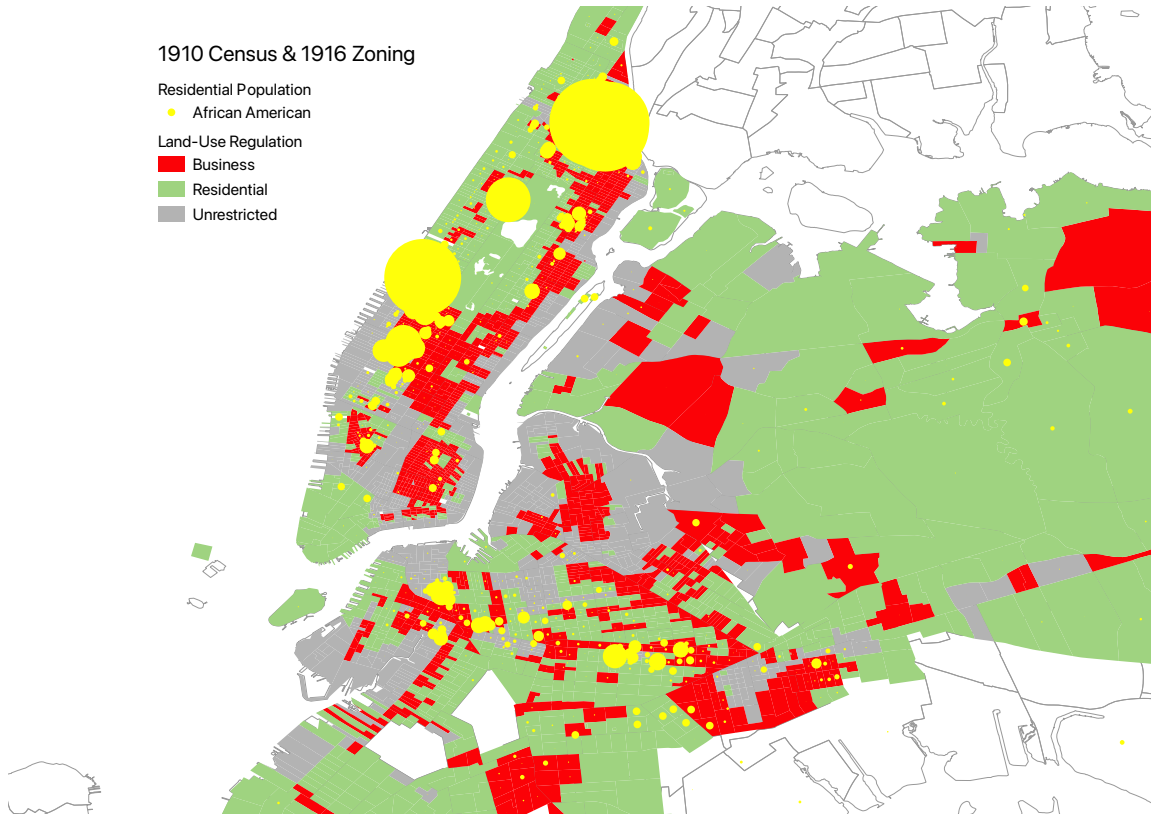
Note: The above figures show the spatial distribution of African Americans in 1910 (6 years before 1916 Zoning Ordinances); Bigger yellow bubbles indicate a higher number of African Americans. Areas where African-Americans resided in high numbers (presented by big yellow bubbles in the figure above) were disproportionately more likely to be zoned as Business districts (in red shade) where manufacturing activities were allowed. Neighborhoods where African Americans lived in very small numbers (presented by small or non-existing yellow bubbles in the figures above) and/or residents' income is high were disproportionately more likely to be zoned for residence districts where any type of manufacturing or industrial activities was strictly prohibited.

Table 1: Zoning Assignment & Residents' Chracteristics

	Business vs. Unrestricted		Residential vs. Unrestricted	
Variable	Relative Risk Ratio	Std Err	Relative Risk Ratio	Std Err
Demographic Characteristics				
Share of African American	1.383	0.863	0.250*	0.181
Socioeconomic Characteristics				
High Income	3.755***	0.561	13.710***	2.025

Land-use regulation being "Unrestricted" is the base outcome. Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.5 Neighborhood Tipping

I combine population census and high-frequency housing sales to document how demographic composition and housing sales price chang during the "tipping" process. While Section 3 illustrates the model of tipping as in Card, Mas and Rothstein (2008), this section explain the data pattern and dynamic patterns of racial segregation. Suppose there exist a threshold (let's call that m^*) where if the minority share in the neighborhood is abovee this threshold (m^*), we observe how previously predominantly white neighborhood neighborhood could "tip" where integrated mixed equilibrium disappears and it will move toward 100% minority equilibrium (i.e. $m = 1$) where minority population is perfectly "isolated".

Demographic Changes

By following the definition of neighborhood tipping, I identify 6 neighborhoods that are mostly located near Harlem. In identifying tipping neighborhoods, I focus on the following demographic changes: 1. population increase of African Americans, 2. disproportionate population decrease of whites leading to net population loss, 3. dramatic increase of share of African Americans. Figure 12 plots such demographic changes. It shows between 1910 and 1920, there was a significant inflow of African Americans and outflow of whites in Central Harlem North and East Harlem North. Consequently, we see the adjacent neighborhoods of these neighborhoods also began experiencing the sudden, dramatic demographic shifts in the following decade.

Figure 12: Population Inflow and Outflow of "Tipped" and other Neighborhoods
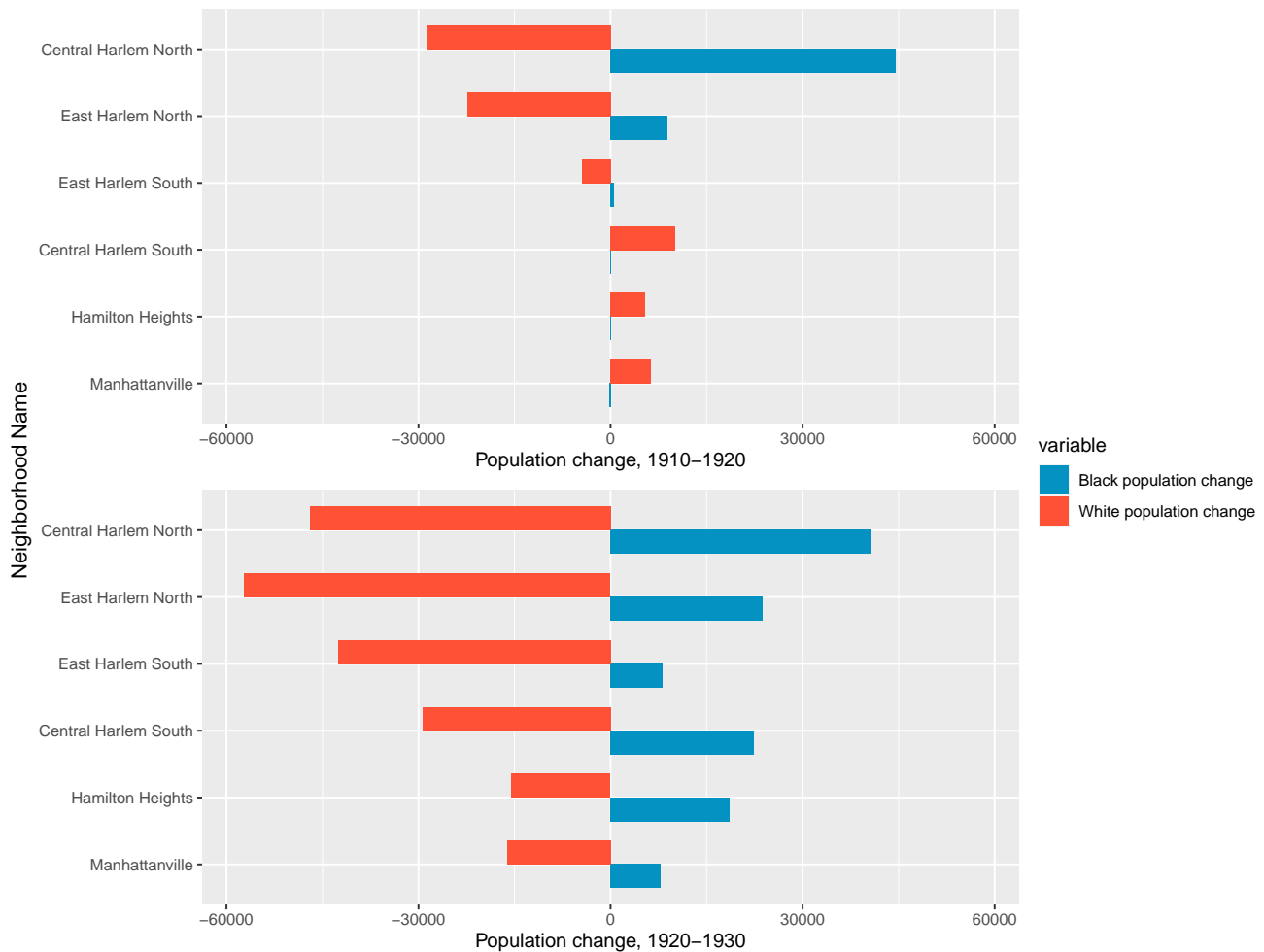
Note: I plot net neighborhood population flows by decade for neighborhoods that "tipped" (Central and East Harlem and neighborhoods near Harlem which are Hamilton Heights and Manhattanville) between 1910 and 1930. Net population analyses reveals that arrival of one African American is associated with 2.5 white departures for "tipping" neighborhoods. Out of 196 neighborhoods in the city, about 6 neighborhoods were going through this dramatic demographic transition , and the rest of neighborhoods were gaining population significantly by margin (ranging from 30- 300% population increases) between 1910 and 1930.

Housing Prices of "Tipped" Neighborhoods This population change was dramatic —for instance, Central Harlem North in 1910, share of African Americans was around 7%; 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 96% 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 near Harlem where the share of minorities was less than 5%, but in just one decade the share sharply increases. When neighborhoods underwent such demographic shirts, what happens to housing prices?

As discussed in Section 3, 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.

Then, in Figure 13, I plot housing price movements of "tipped" neighborhoods in Harlem (in Blue line) over the period 1870-1940. The dark grey dot line is the hedonic price index based on maximum assessed valuations of properties; whereas the light grey dash line is the hedonic priced index based on minimum assessed-valuations; the blue line is the hedonic sales price index based on actual transacted prices. Using high frequency time-series hedonic sales price index by neighborhood for 70 years, I test if and when break(s) occur especially during this dramatic demographic changes (Bai and Perron (1998)) in those identified neighborhoods.

Figure 13 reveal that sales price index of Central Harlem neighborhoods reveal a structural break in 1917 (a year after zoning ordinances were adopted), whereas adjacent neighborhoods in East Harlem reveal a structural break in 1920 in terms of housing sales prices. In identifying structural breaks, I also control for transit access changes and distance from the Battery; I also use alternative measures as outcome variables which are deviations of sales price index from assessed valuations (minimum and maximum respectively) and they yield robust timing of structural breaks respectively. As in Table 6, housing sales price decrease by 40% relative to non-tipping neighborhoods, and housing prices in these neighborhoods continue to be depressed due to hysteresis.

Figure 13: Housing Prices of "Tipped" Neighborhoods
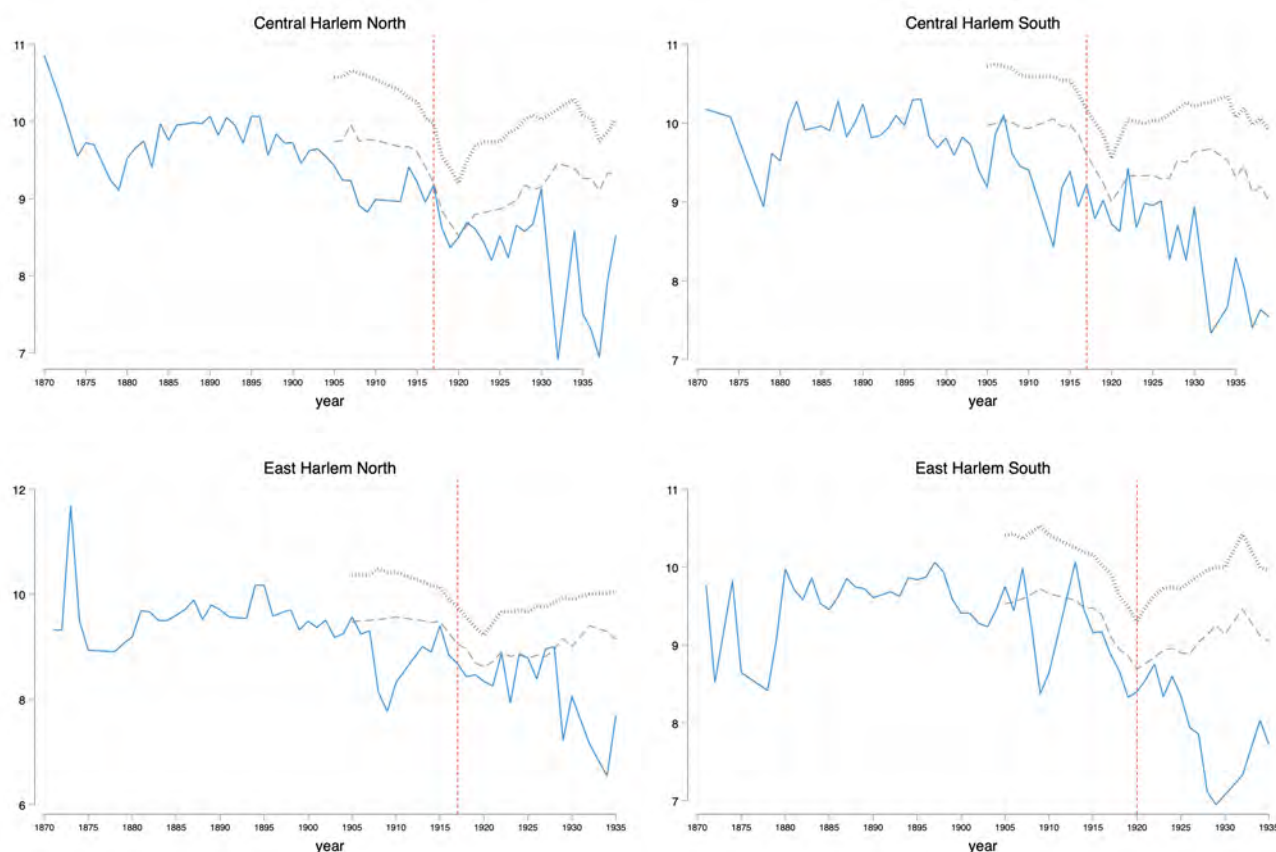
Note: I plot hedonic house sales price indexes by tipping neighborhoods over 70 years. In all subfigures here, the dark grey dot line plots the hedonic maximum assessed-valuations-based price index ; the light grey dash line plots the hedonic minimum assessed-valuations -based price index; blue line plots the hedonic sales price-based price index. Using neighborhood-level hedonic sales price index, I identify the existence and timing of structural break between 1910 and 1930 wh the dynamic process of residential racial segregationCompared to their pre-trends before 1910 and fundamental values "measured by the assessed values" (in grey lines), housing prices in "tipping" neighborhoods were declining.

5 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.[30] Boustan, Bunten and Hearey (2018) documents that the urban wage premium in the US was remarkably stable over the past two

---

[30]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, raced people."*

centuries, ranging between 15 to 40 percent.[31] 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.[32]

## 5.1 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 14 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 reveals why African American men's urban wage premium was lower than that of 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.

---

[31]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.

[32]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.

Figure 14: Income Changes: Pre- and Post- Urban Migration
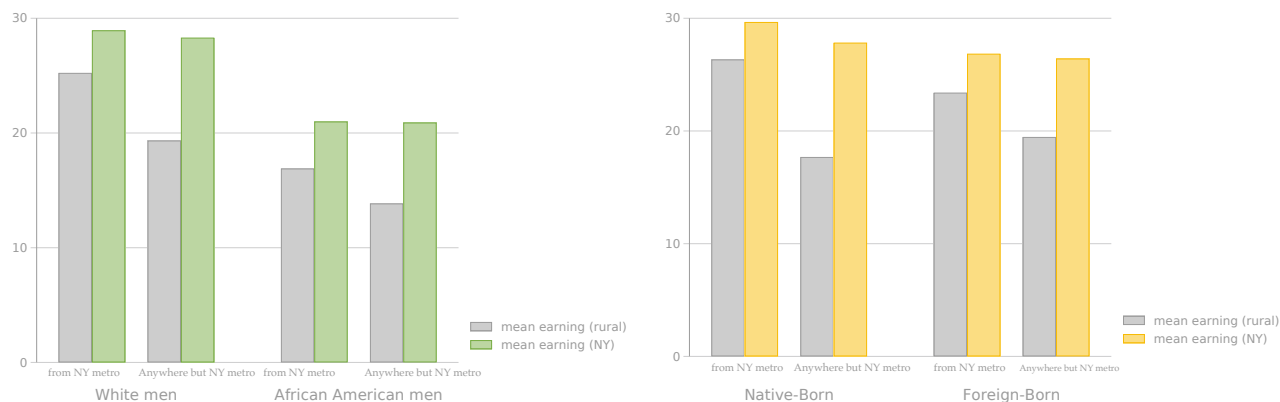
Note: The figure on the left plots the mean earnings of white and African American migrants from rural areas to New York. Their prior occupation-based earnings in rural areas are depicted in grey, whereas their mean earnings in New York are depicted in green. Once they migrated to urban locations like New York, their earnings increased by 35-40% for all races. Although returns to migration from rural areas to urban city were substantial for people of all races, the mean earnings of African American migrants in New York was significantly lower than those of white migrants.

The figure on the right plots the mean earnings of native and foreign-born migrants from rural areas to New York. Their prior occupation-based earnings in rural areas are depicted in grey, whereas their mean earnings in New York are depicted in yellow. Once they migrated to urban locations like New York, their earnings increased by significantly for both native and foreign-born individuals. Unlike African American men where we see the sizable earnings gap between racial groups, there was no gap of ''urban wage premia'' between native and foreign-born individuals.

Figure 16, derived from 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 primarily sort in personal service industry where there exsits relatively less innovation and growth.

Figure 15: Distribution of Earned Income Percentile Rank Changes of Rural to Urban Migrants: by Race
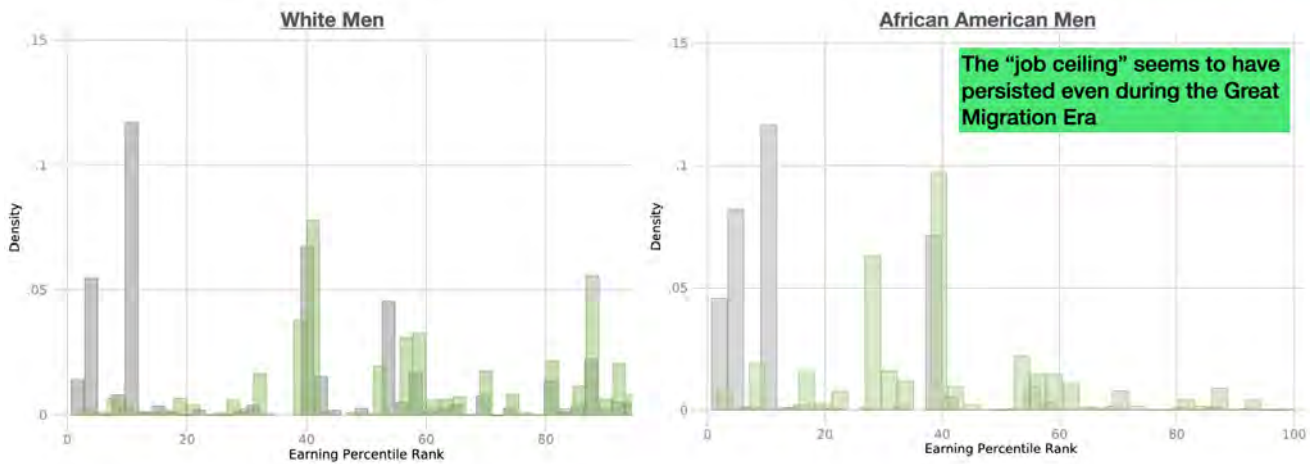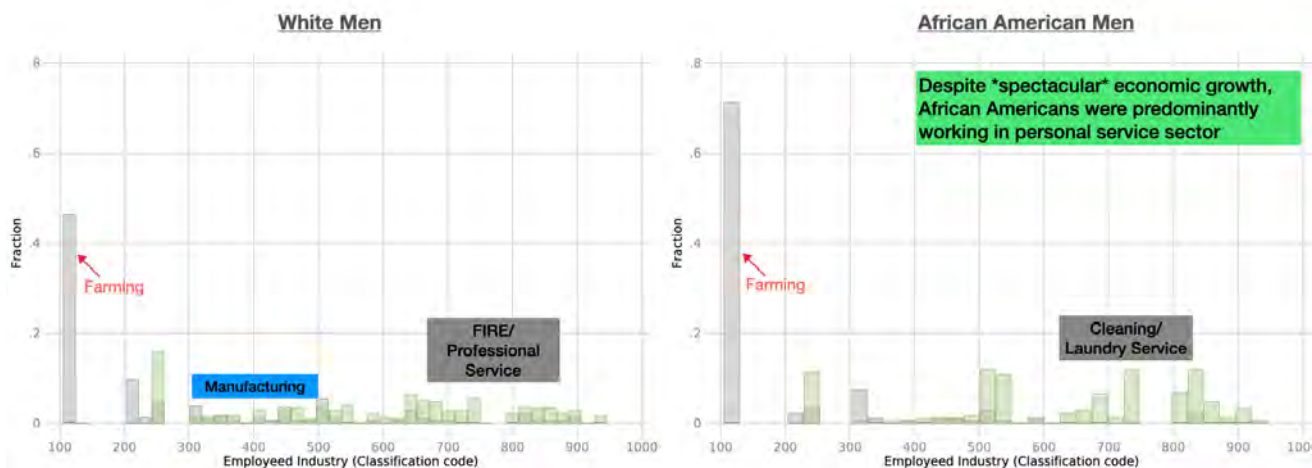

Note: The figure on the left plots the distribution of earning percentile rank of white migrants from rural areas to New York. Their prior occupation-based earnings distribution are depicted in grey, whereas theirearning distribution in New York is depicted in green. Higher concentration of grey bars 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 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 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.

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 my panel data of individuals most workers of all races in rural areas worked in farming industry) that disproportionately exist in urban locations. 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 15, 16).

Figure 16: Distribution of Employed Industry: Rural to Urban Migrants: by Race

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.

5.2 Long-term impacts of living in Segregated Minority Neighborhoods

I use panel data of individuals that include rural to urban (i.e. New York) migrants to measure their economic mobility due to rural to urban migration. I calculate the earnings changes of the same individuals between two adjacent censuses. Table 2 shows that earnings increase about 6.7 (approximately 40% earnings increases) for all men who migrated from rural to New York City (Column 1). However, this premium is reduced for African American migrants living in segregated minority neighborhoods (Column 3).

$$\text{Earnings}_{it} = \alpha_i + x'_{it}\beta + d_t + u_{it}$$

Table 2: Urban Premium for Rural to NY migrants

|	(1) Income	(2) Income	(3) Income
Urban (i.e. New York City)	6.680***	6.708***	6.708***
	(0.050)	(0.053)	(0.053)
Urban x African American		-0.336***	-0.170
		(0.129)	(0.144)
Urban x African American x Minority Nghd			-0.296***
			(0.113)
Observations	932718	932718	932718
R-Squared	0.046	0.046	0.046

Standard errors in parentheses

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

5.3 Systematic Barriers for Wealth Formation

Beyond the finding that it was the segregated minority neighborhood that adversely affected earnings of African Americans, I also show the differential outcome of wealth formation beyond labor earnings. Figure 17 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 17: Barriers for Wealth Formation

Share of Home Ownership
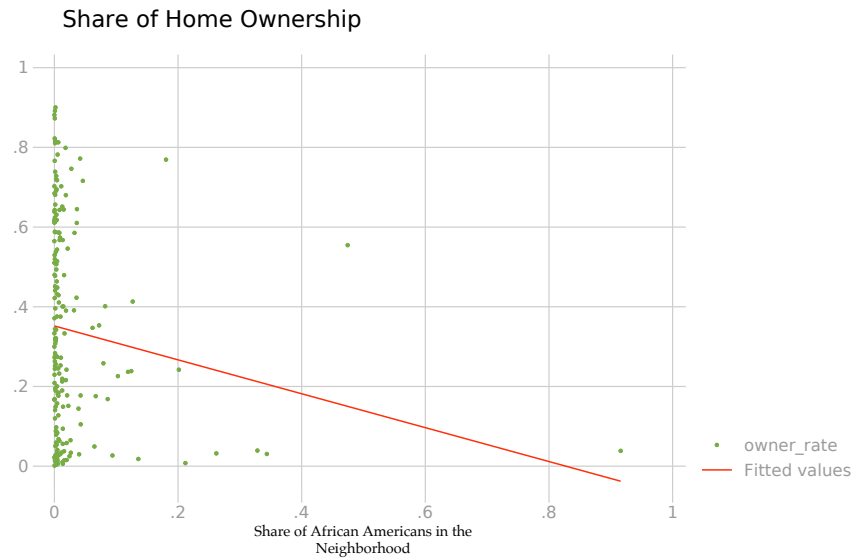
Note: The above plot depicts the share of home ownership using households of all races in each neighborhood from the 1930 census. Each dot represents a neighborhood in New York City and I calculate the share of home ownership by neighborhood by looking at every household's homeownership (covering all races in the neighborhood).

5.4 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.

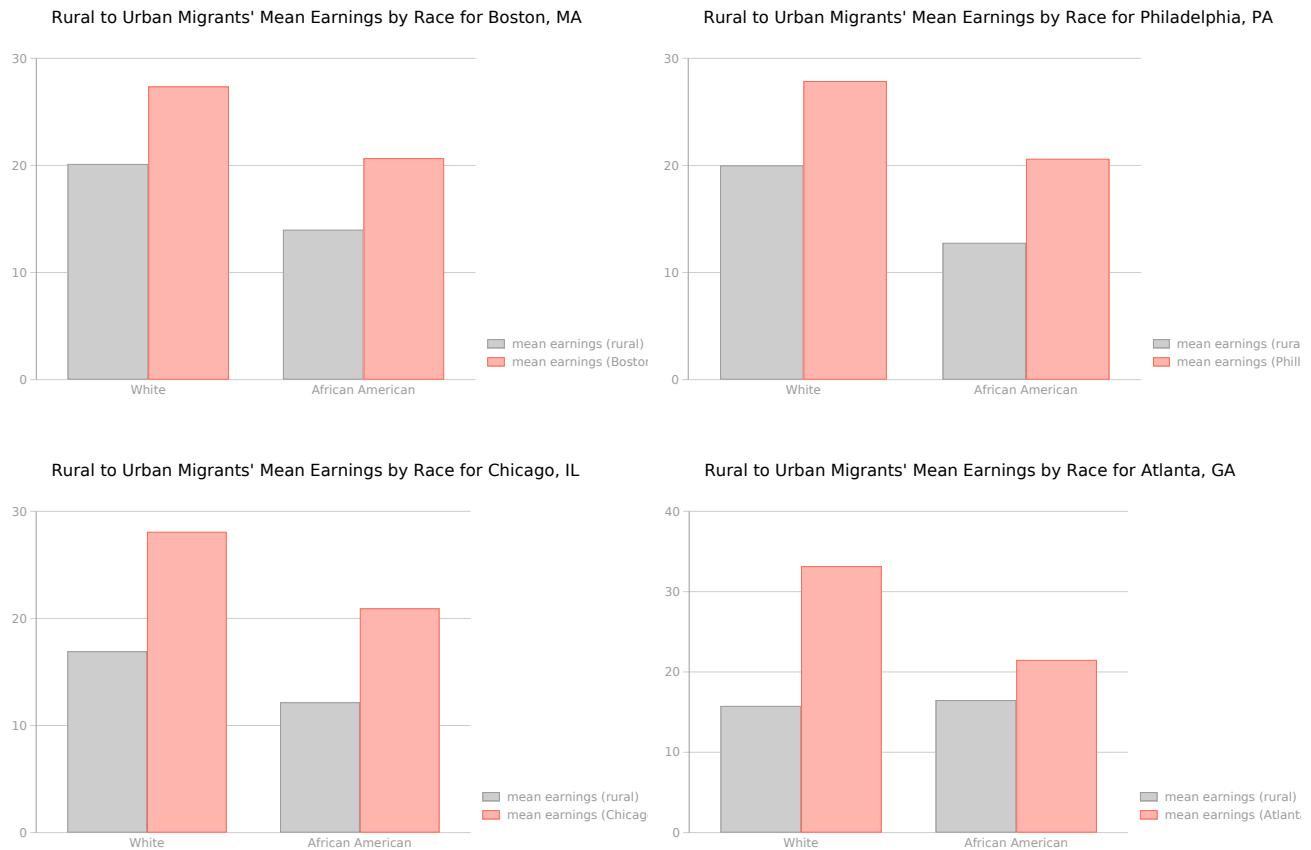Figure 18: Income Changes of Rural to Urban Migrants by Race

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 Northern 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   Conclusion

Through extensive efforts in digitization, record linking, and spatial analyses, I establish that both transportation infrastructure and zoning increased growth but had "unanticipated" consequences—increasing isolation of African Americans in industries, occupations, and neighborhoods. After the railway arrival, neighborhoods became increasingly segregated both by income and race (increased segregation between neighborhoods), and the nation's first comprehensive zoning policy (especially residential zoning policy) induced African Americans to become extremely "isolated" in a few minority-neighborhoods (increased segregation within neighborhoods).

I use neighborhood "tipping" models and the dynamics of segregation (as in Card, Mas and Rothstein (2008)) to establish a dynamic connection between housing prices and residential segregation over time. By following the same neighborhoods and documenting population flows in neighborhoods, I estimate the value of the "tipping point" (i.e. share of minority population in each neighborhood where mixed equilibrium is no longer sustainable), and I also show that housing price changes of "tipping" neighborhoods gradually decreased throughout the dynamic "tipping" process. I find the inflow of one African American family in tipping neighborhoods is associated with an outflow of 2.5 Whites, leading to net population loss in "tipping" neighborhoods that could explain the decrease of housing prices.

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 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–reveal that "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.

Relatedly, there might have been a bigger welfare disparity between racial groups beyond economic indicators (which were the main elements that I discuss in this paper). For example, there may have been a significant welfare disparity in terms of public service and utility provisions (e.g. sewer system, better policing activities) between neighborhoods. Also, there could have been long-term health consequences due to long-term exposure to air pollution or hazardous materials (See, for example, Heblich, Trew and Zylberberg (2021); Logan and Parman (2018)).

Moreover, my findings are still relevant to cities and countries that are implementing public policies such as the provision of a large-scale transit infrastructure investments and land-use regulations. My findings based on ''1870-1940 New York City'' could apply to cities and developing countries in the 21st century that are experiencing rapid economic and population growth, and therefore such public policy interventions are being implemented. For instance, after New York City's 1916 Zoning Ordinances, many other American cities implemented land-use regulations (that include exclusionary ''residential'' zoning); zoning remains to be one of the most widely and extensively used urban planning tools to this date in many parts of the world.

Finally, my findings also show that the effort of desegregating neighborhoods and communities could and *should* go far beyond ''redlining.'' For instance, my paper shows that the majority of African American households were located in few economically disadvantageous ''tipped'' neighborhoods *well* before the federal government's involvement in mortgage markets; this implies that racial segregation and racially discriminatory practices have been underway decades before the federal government's ''redlining'' in the 1930s. Recently, the Justice Department announced the launch of the department's new Combatting Redlining Initiative.[33] Although this new initiative is an extremely meaningful effort in addressing racial inequality, this initiative alone is unlikely to address the issues of segregation completely. Other efforts of desegregating neighborhoods such as Inclusionary Zoning, removal of interstate highways have been discussed and implemented to build a more inclusive society. Yet, as inspirational and material these efforts are, drawing lessons from the past could help us choose whether to reflect and revise some of the proposed solutions (such as taking down the highways).[34] As mentioned in the article, *''reconnecting neighborhoods is more complicated than breaking them apart''*; not carefully thought-out efforts may generate another set of ''unanticipated'' long-term consequences.

---

[33]According to the Justice department, ''redlining'' is an illegal practice in which lenders avoid providing services to individuals living in communities of color because of the race or national origin of the people who live in those communities.

[34]See, for example, New York Times article ''Can Removing Highways Fix America's Cities?'': https://www.nytimes.com/interactive/2021/05/27/climate/us-cities-highway-removal.html

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.

Bai, Jushan and Pierre Perron, ''Estimating and Testing Linear Models with Multiple Structural Changes,'' *Econometrica*, 1998, *66* (1), 47--78.

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.*

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," The New Palgrave Dictionary of Economics Online (Palgrave Macmillan, 2008) Steven N. Durlauf and Lawrence E. Blume (eds),, 2008.*

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 Bogotá'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 2. 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 19 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 19, 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 19: Geographic Boundary of NYC Metro Area
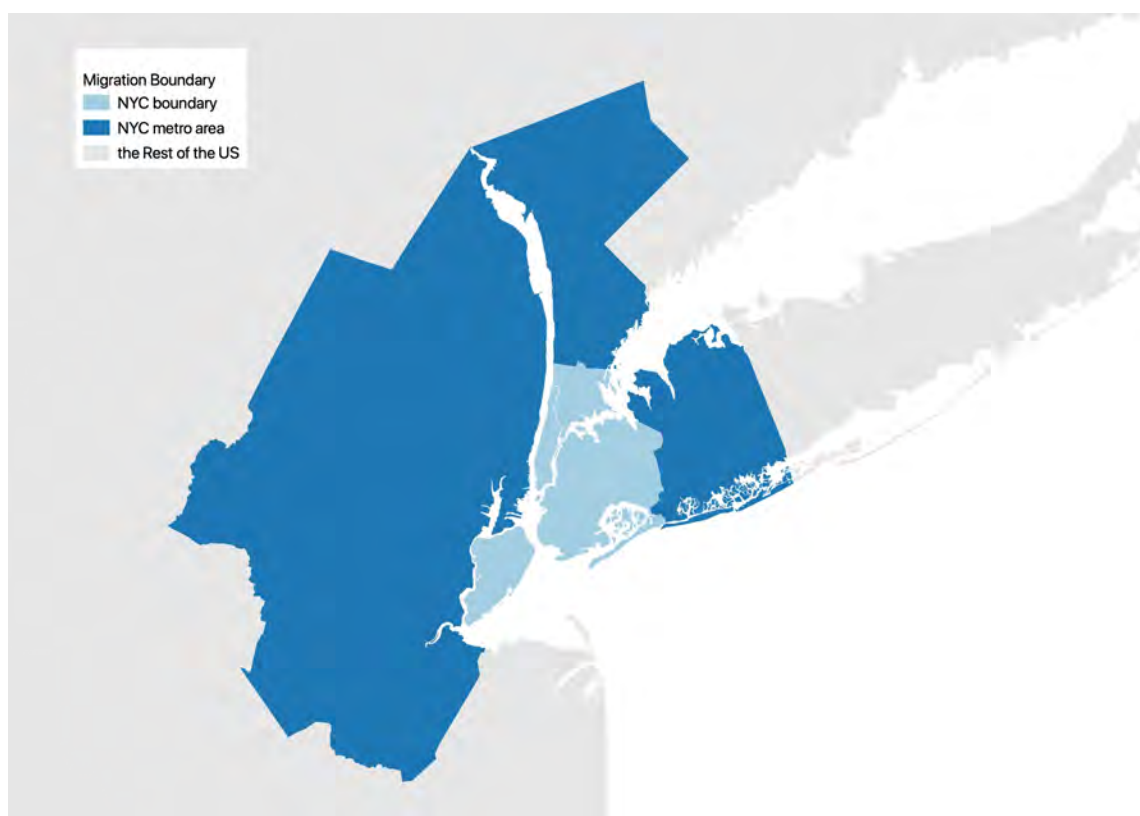
B.2 Geographic Information Harmonization

Georeferencing and Creation of Shapefiles: Neighborhood Boundaries   The primary geographic units used in the analyses are "Neighborhood Tabulation Areas" (hereafter, NTAs),which is an aggregation of historical census tracts called "enumeration districts (EDs)". There are 195 time-consistent neighborhoods in the city with a minimum of 15,000 to 100,000 residents during the study period. 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

---

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

found in various data sources (e.g. "enumeration district" in US census records) to NTAs so that NTAs can be a time-invariant, consistent geographic unit of analyses. Therefore, all datasets used in the analyses are harmonized and geolinked to NTAs. Figure 20 shows NTA boundary in blue, and historical census tracts that make up each NTA in light grey.

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

---

[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)).

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. 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 20: Neighborhood Boundary of NYC

Neighborhood Definition:
- Census enumeration district boundary
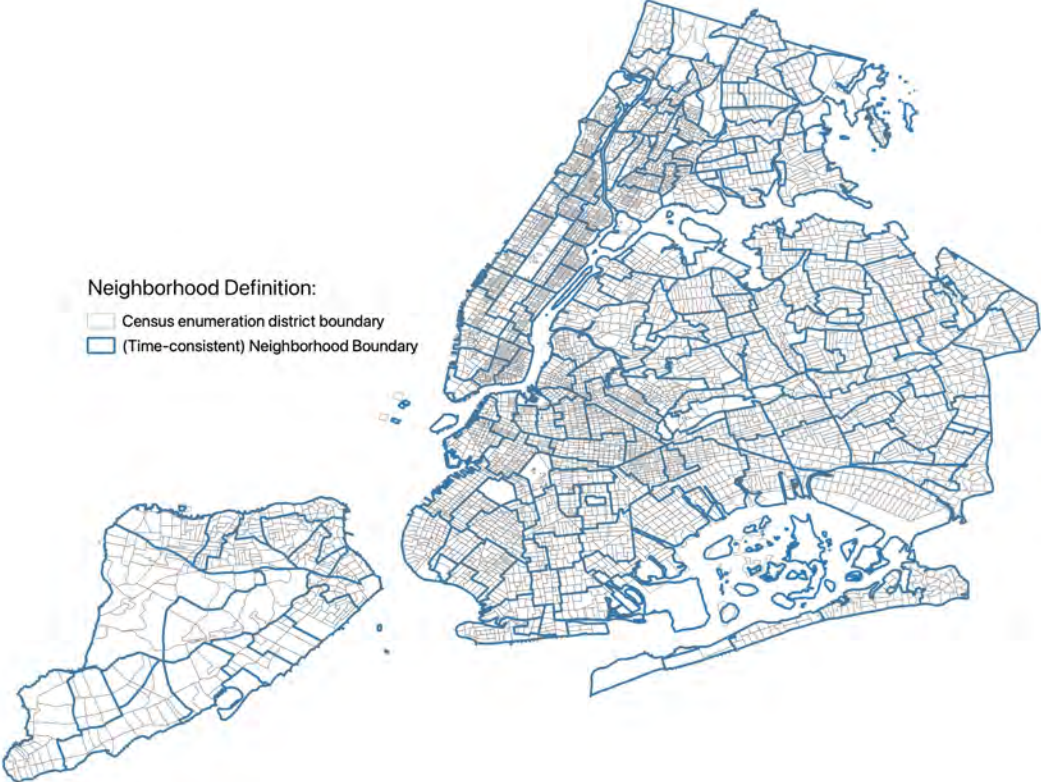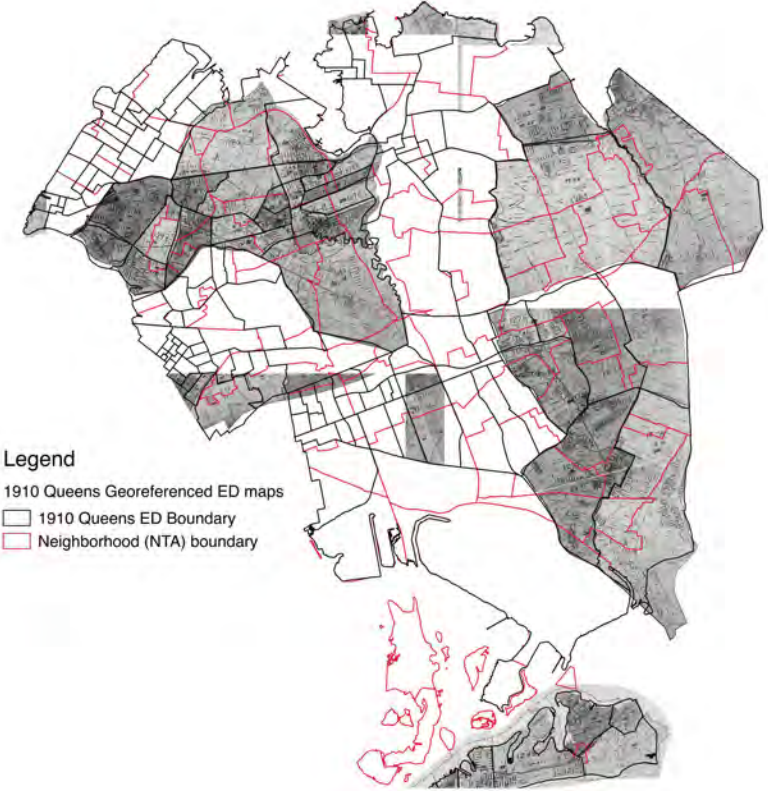- (Time-consistent) Neighborhood Boundary

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
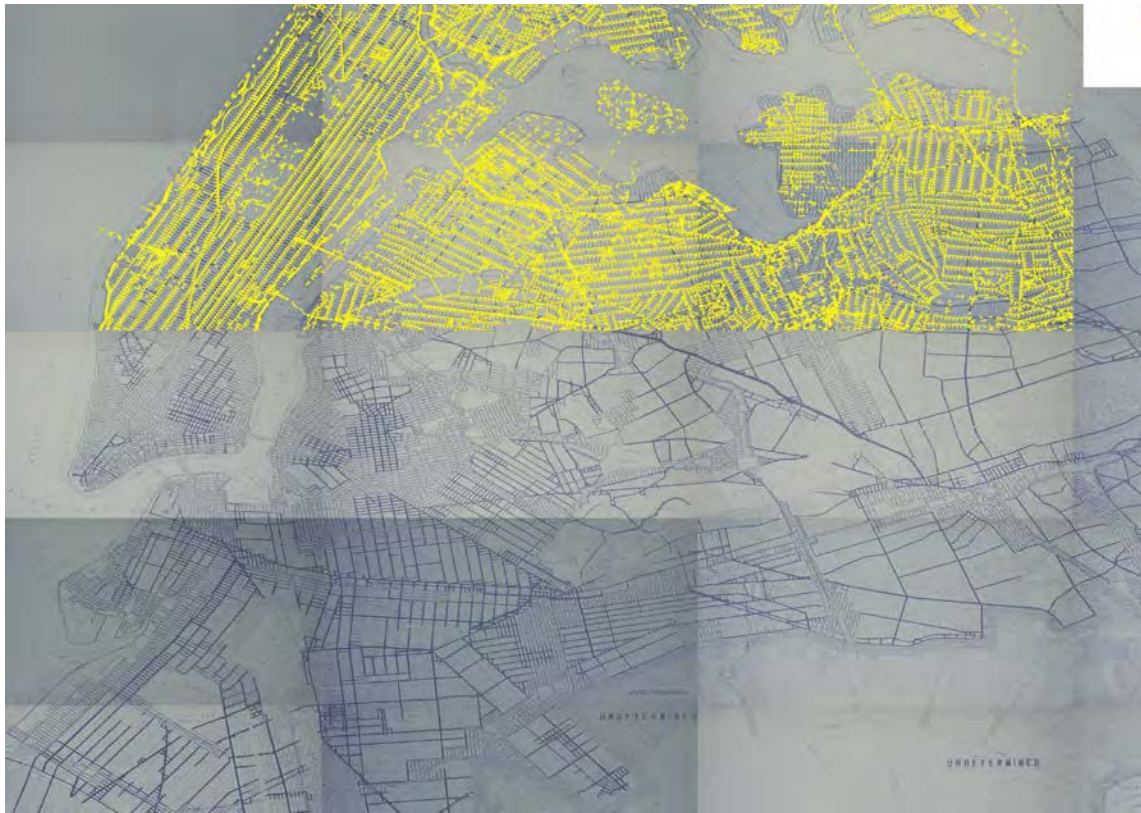
(a) Georeferenced 1916 Zoning Land-Use Restrictions map
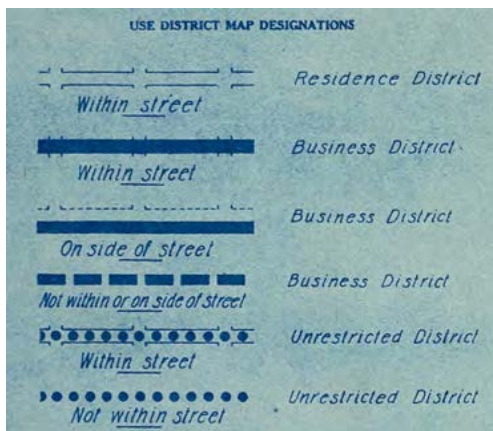
(b) Legend of 1916 Zoning Resolution:
Land Use

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
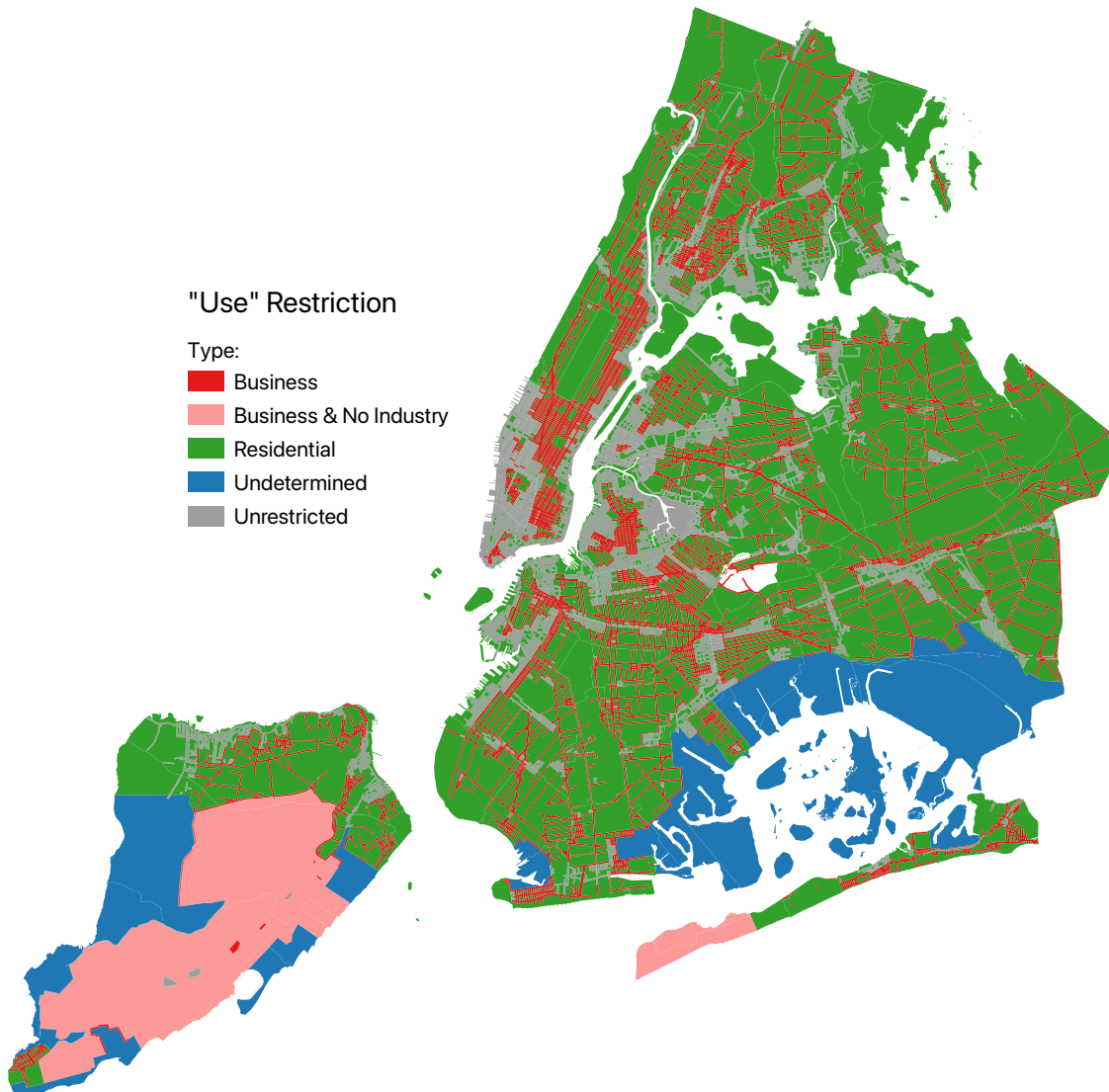
Note: Author's own creation based on digitized and georeferenced maps of 1916 Zoning Ordinances on Use-Restriction. Originally map images were scanned and provided the New York Public Library Map Division. I georeferenced the use-restriction maps down to street level accuracy by overlaying these use 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)
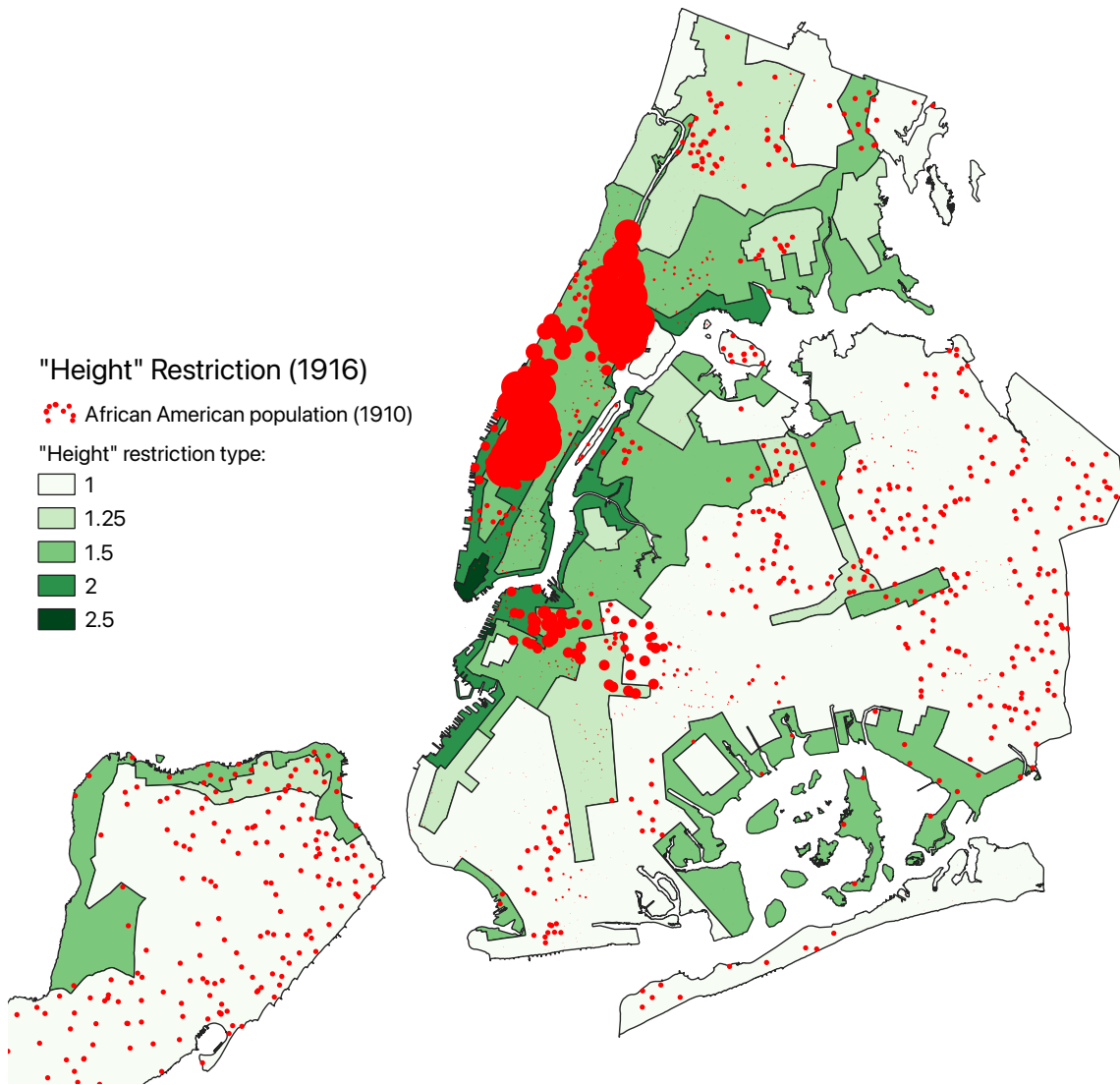
"Height" Restriction (1916)

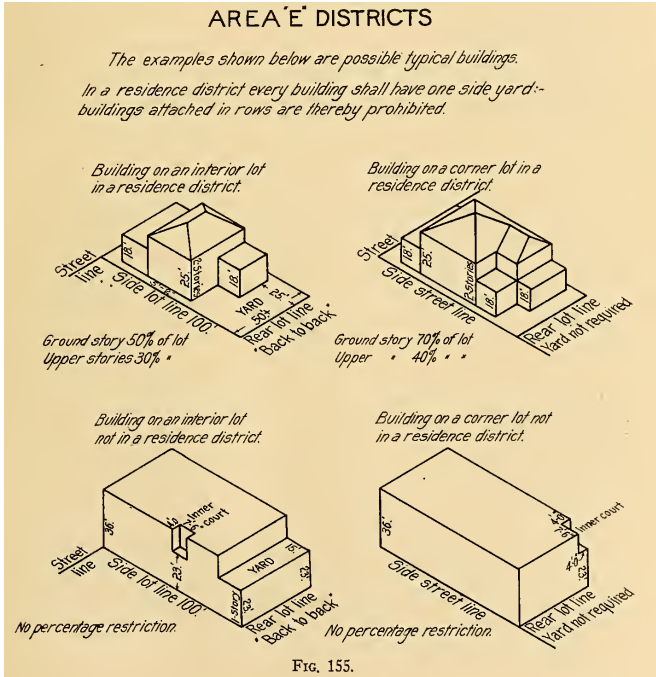African American population (1910)

"Height" restriction type:
- 1
- 1.25
- 1.5
- 2
- 2.5

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.

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

(a) Illustration of Area Restrictions
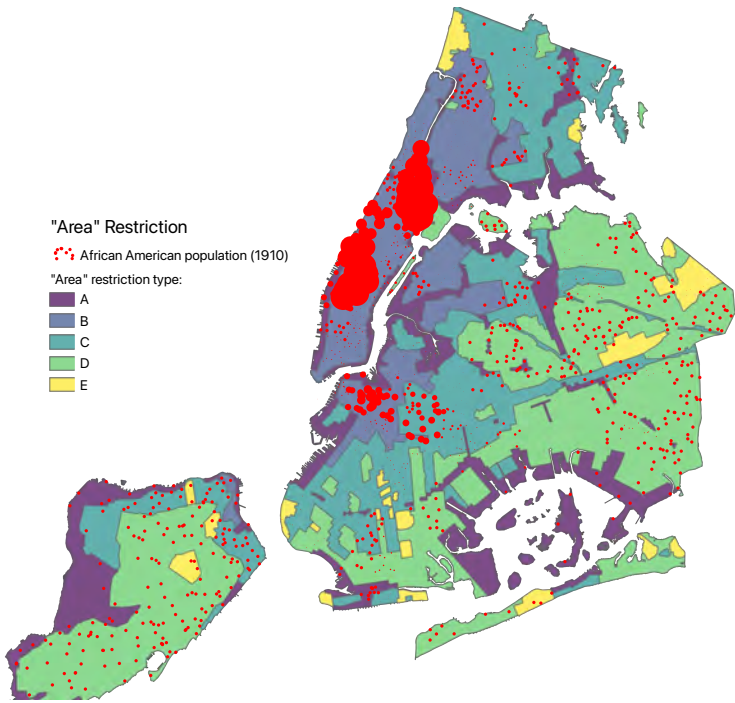
(b) pre-zoning population trend (1910) Height regulation (1916)

Note: Figure 26(a) was originally from New York and Apportionment (1916).

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
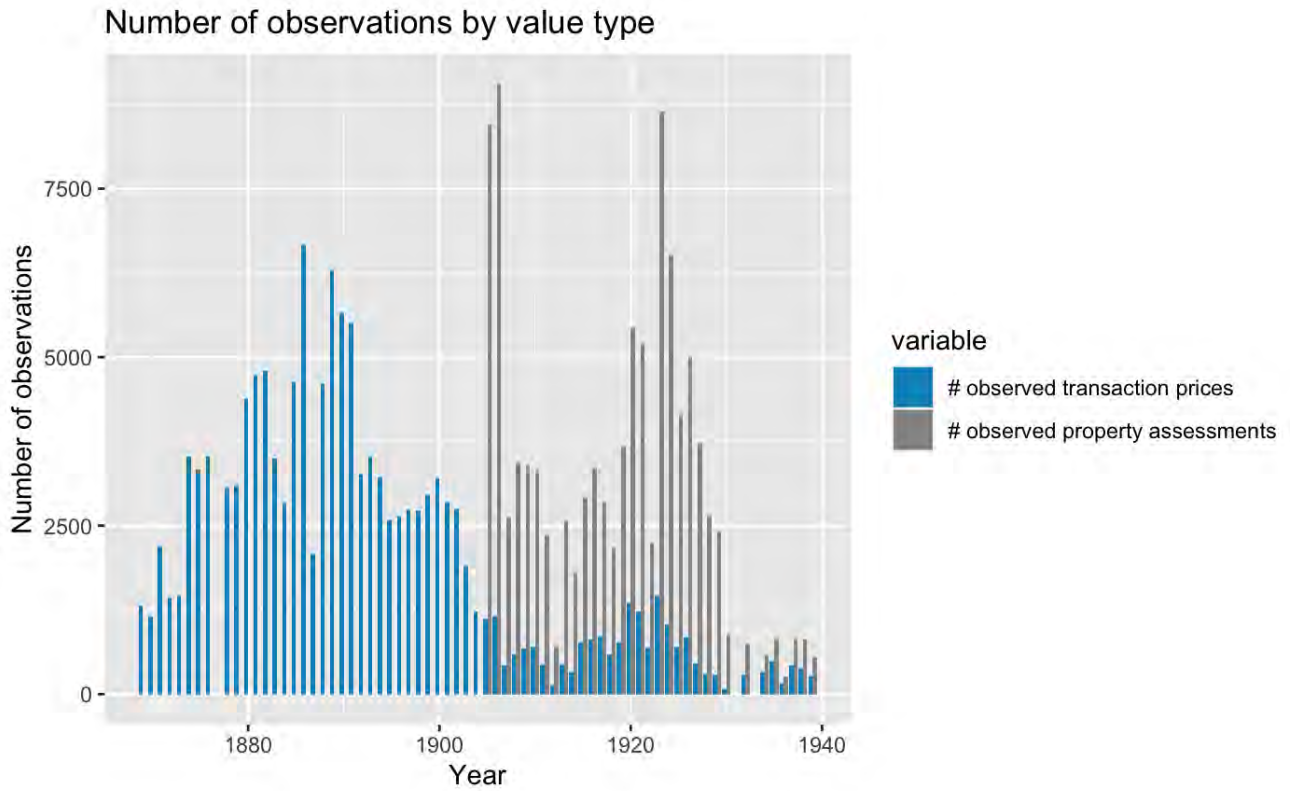
---

[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

Number of observations by value type

variable
- # observed transaction prices
- # observed property assessments

(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.

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 27, one dot plots 500 people, and the color red denotes African American population whereas the color green denotes white population. Figure 28(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 28(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 28(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 28 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 28 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 27: Mapping Residential Segregation by Race

(a)

(b)

Figure 27: Mapping Residential Segregation by Race

(c)

(d)

Figure 27: Mapping Residential Segregation by Race

(e)
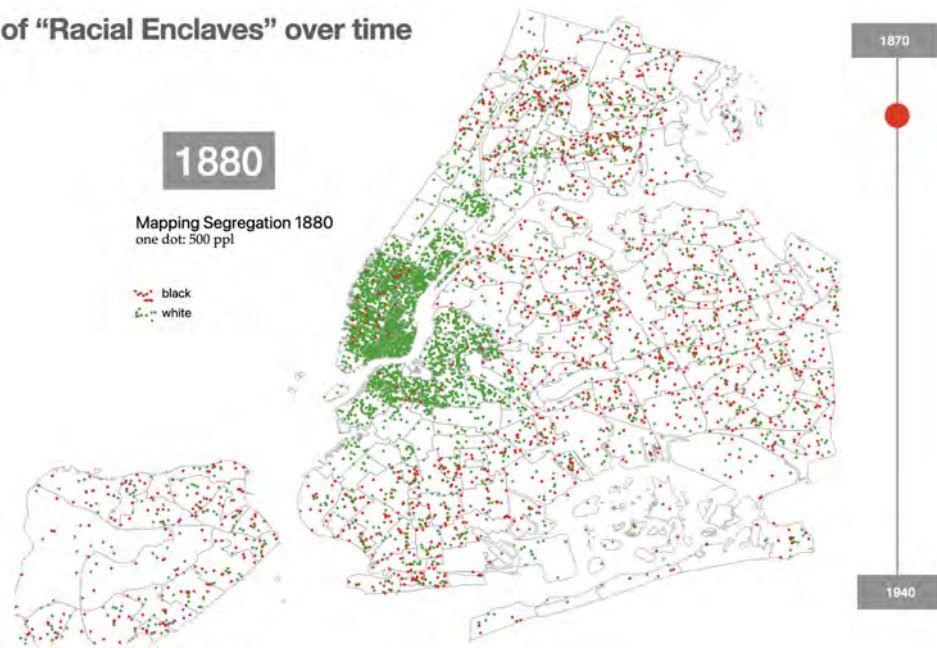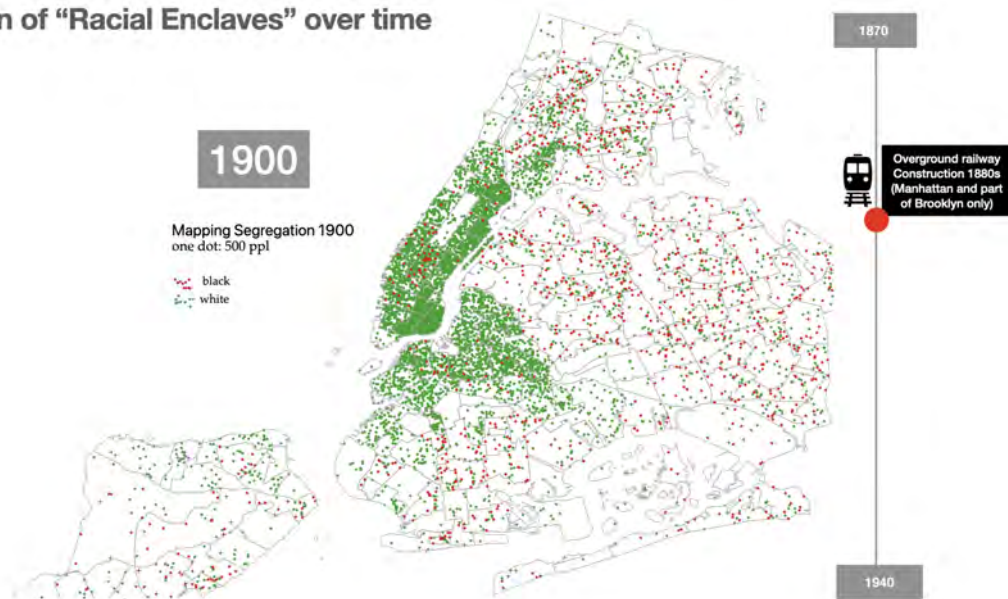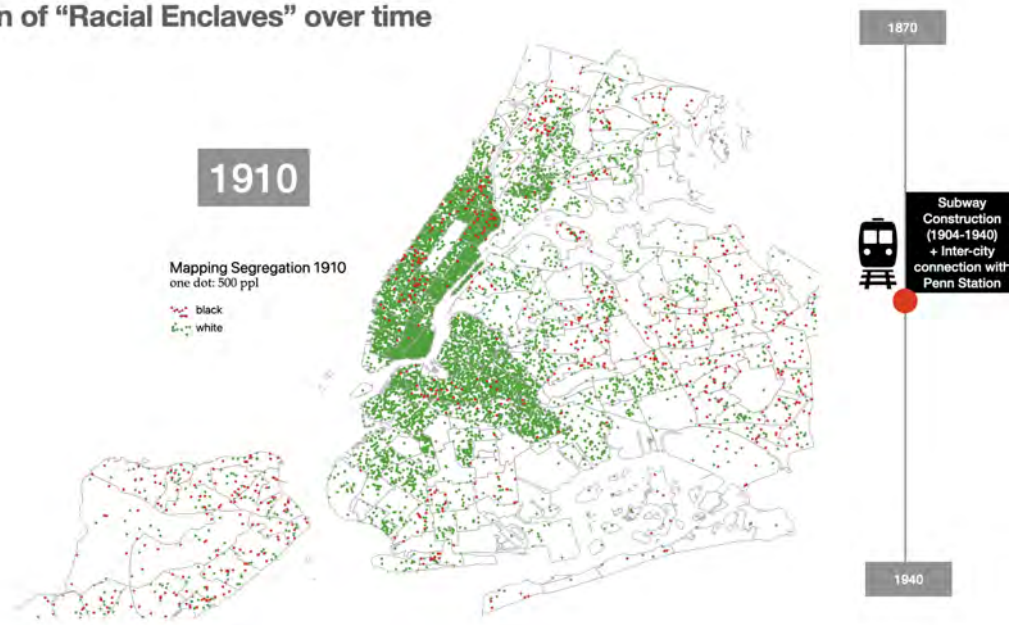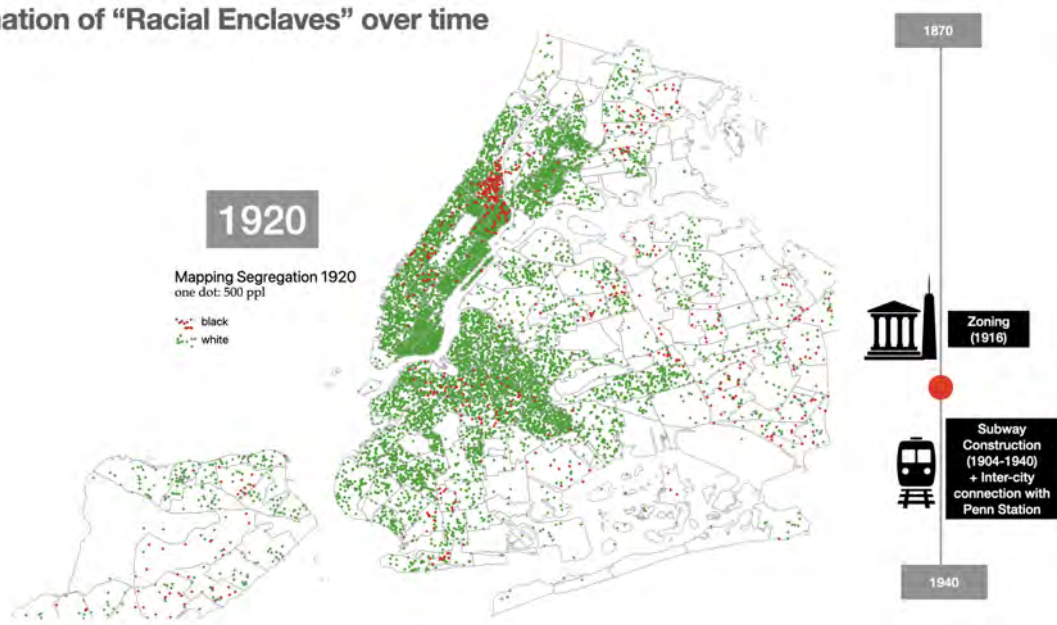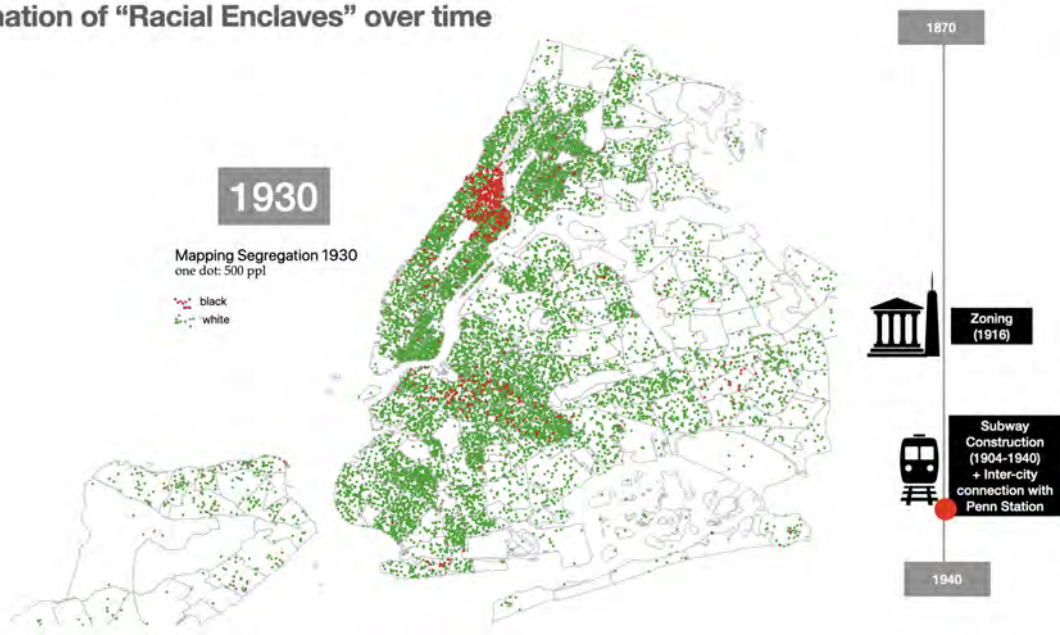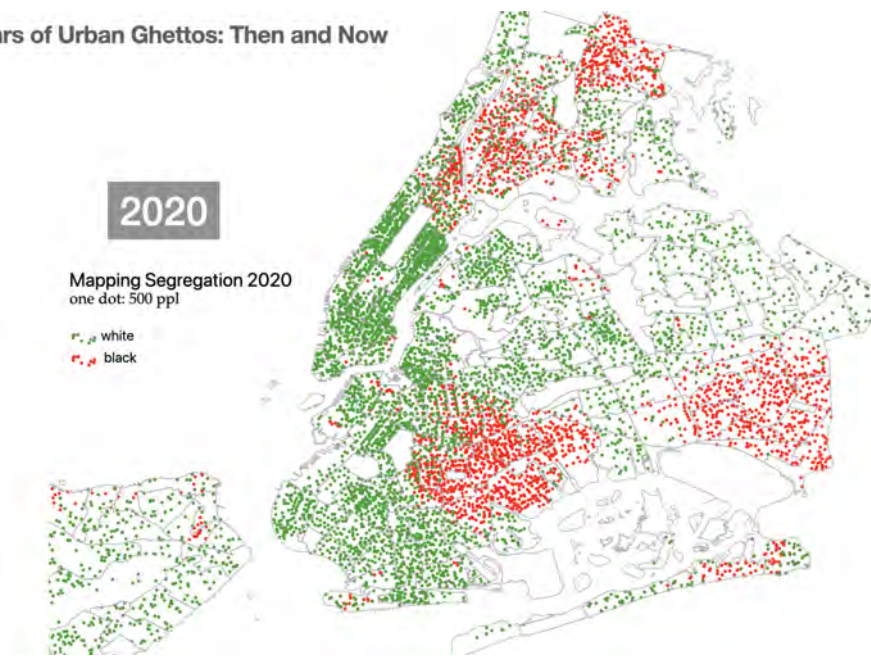
(f)

Figure 28: Residential location of African Americans by income

1930 Census:

one dot: 500 people

bottom quartile

top quartile
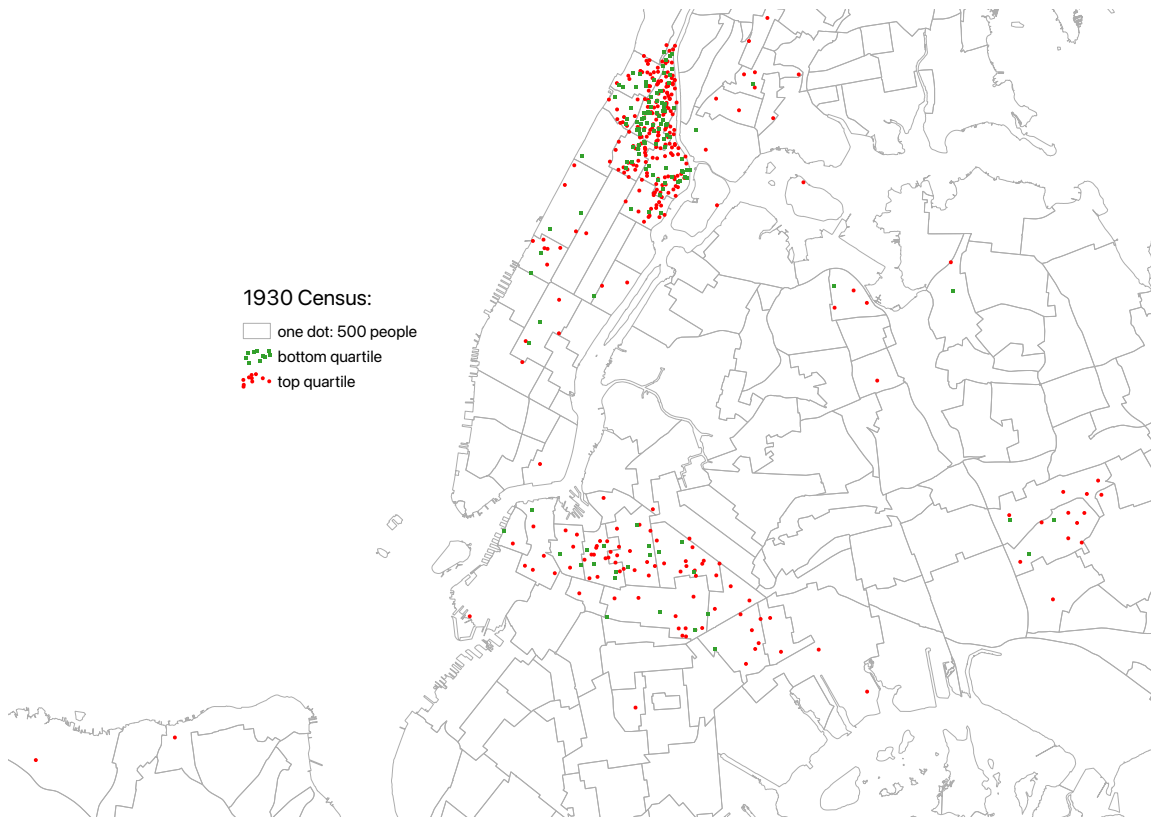
D   Regression Results

D.1   Population

Table 3: Treatment Effects for the Railway Network Arrival on Neighborhood Population in New York, 1870-1940

	(1)	(2)	(3)
	$Population_{it}$	$Population_{it}$	$Population_{it}$
$\beta_{\tau=40}$	12416.30	22254.19***	29868.40***
	(10302.05)	(3361.13)	(9174.58)
$\beta_{\tau=30}$	18279.24**	23914.07***	27701.28***
	(7285.22)	(3197.81)	(6397.46)
$\beta_{\tau=20}$	14954.13**	20086.17***	23308.61***
	(5988.00)	(2980.04)	(5421.98)
$\beta_{\tau=10}$	1300.99	7937.43***	8160.25
	(6593.16)	(2170.06)	(5744.99)
$\beta_{\tau=0}$	–	–	–
$\beta_{\tau=-10}$	-5476.77	-1794.09	-1696.94
	(4613.26)	(2569.95)	(4226.14)
$\beta_{\tau=-20}$	-8820.41	-4639.07	-4522.07
	(6886.54)	(3670.72)	(5906.75)
$\beta_{\tau=-30}$	-8307.65	-3683.16	-3074.23
	(6180.09)	(2693.46)	(5494.29)
$\beta_{\tau=-40}$	-16819.80**	-9138.56***	-7815.37
	(7893.72)	(3340.52)	(6886.33)
$\gamma_{\tau=40}$	–	-27626.06	-95585.04***
		(27036.29)	(35138.36)
$\gamma_{\tau=30}$	–	-20608.14	-65106.31**
		(23595.94)	(27862.50)
$\gamma_{\tau=20}$	–	-17565.73	-47744.98**
		(19315.12)	(19382.75)
$\gamma_{\tau=10}$	–	-27088.54	-40012.14
		(24624.96)	(25650.83)
$\gamma_{\tau=0}$	–	–	–
$\gamma_{\tau=-10}$	–	8750.62	-5889.96
		(23395.51)	(18783.84)
$\gamma_{\tau=-20}$	–	-14731.33	-930.03
		(25329.87)	(28779.57)
$\gamma_{\tau=-30}$	–	-14773.67	-8566.63
		(22713.72)	(21335.22)
$\gamma_{\tau=-40}$	–	-54741.67	-34048.24
		(39423.78)	(21773.71)
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	0.63	0.64	0.67

Standard errors in parentheses

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

D.2 Socioeconomic Status

Table 4: Treatment Effects for the Railway Network Arrival on Residents' Socioeconomic Status

	(1)	(2)	(3)
	SES_{it}	SES_{it}	SES_{it}
$\beta_{\tau=40}$	-2.204***	-1.147**	-1.925***
	(0.481)	(0.499)	(0.508)
$\beta_{\tau=30}$	-1.032***	-0.879**	-0.900**
	(0.341)	(0.343)	(0.361)
$\beta_{\tau=20}$	-0.417	-0.312	-0.230
	(0.277)	(0.272)	(0.296)
$\beta_{\tau=10}$	-0.261	-0.125	-0.096
	(0.266)	(0.278)	(0.302)
$\beta_{\tau=0}$	–	–	–
$\beta_{\tau=-10}$	-0.100	0.086	-0.132
	(0.330)	(0.343)	(0.375)
$\beta_{\tau=-20}$	0.063	0.714*	0.234
	(0.395)	(0.369)	(0.484)
$\beta_{\tau=-30}$	0.631	0.783	0.290
	(0.483)	(0.541)	(0.584)
$\beta_{\tau=-40}$	0.979	1.492*	0.520
	(0.724)	(0.795)	(0.875)
$\gamma_{\tau=40}$	–	-2.817***	-1.534***
		(0.767)	(0.576)
$\gamma_{\tau=30}$	–	-0.995	-0.950**
		(0.736)	(0.483)
$\gamma_{\tau=20}$	–	-0.344	-1.524***
		(0.835)	(0.385)
$\gamma_{\tau=10}$	–	-0.826	-0.882**
		(0.691)	(0.403)
$\gamma_{\tau=0}$	–	–	–
$\gamma_{\tau=-10}$	–	1.292*	0.517
		(0.682)	(0.442)
$\gamma_{\tau=-20}$	–	-1.895*	-0.374
		(1.084)	(0.668)
$\gamma_{\tau=-30}$	–	1.461**	1.178**
		(0.735)	(0.554)
$\gamma_{\tau=-40}$	–	1.069	1.739**
		(0.750)	(0.738)
Central Location Defintion		Manhattan only	Within 5 km from Wall Street
Neighborhood FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	921	921	921
R-squared	0.544	0.561	0.554

D.3 Segregation

Table 5: Treatment Effects for the Railway Network Arrival on Residential Segregation in New York

	(1)	(2)	(3)
	$Dissimilar_{it}$	$Dissimilar_{it}$	$Dissimilar_{it}$
$\beta_{\tau=40}$	0.022***	0.027***	0.019**
	(0.008)	(0.008)	(0.008)
$\beta_{\tau=30}$	0.020***	0.016***	0.018***
	(0.006)	(0.006)	(0.006)
$\beta_{\tau=20}$	0.013***	0.009*	0.011**
	(0.005)	(0.005)	(0.005)
$\beta_{\tau=10}$	0.010**	0.010**	0.008*
	(0.004)	(0.005)	(0.005)
$\beta_{\tau=0}$	–	–	–
$\beta_{\tau=-10}$	0.005	0.004	0.003
	(0.005)	(0.005)	(0.005)
$\beta_{\tau=-20}$	-0.006	-0.011*	-0.010
	(0.007)	(0.007)	(0.009)
$\beta_{\tau=-30}$	-0.009	-0.011	-0.014
	(0.011)	(0.012)	(0.014)
$\beta_{\tau=-40}$	-0.003	-0.005	0.002
	(0.012)	(0.012)	(0.014)
$\gamma_{\tau=40}$	–	-0.005	0.018
		(0.014)	(0.013)
$\gamma_{\tau=30}$	–	0.017	0.018*
		(0.012)	(0.010)
$\gamma_{\tau=20}$	–	0.037***	0.006
		(0.012)	(0.008)
$\gamma_{\tau=10}$	–	0.001	0.010
		(0.012)	(0.009)
$\gamma_{\tau=0}$	–	–	–
$\gamma_{\tau=-10}$	–	0.008	0.009
		(0.013)	(0.009)
$\gamma_{\tau=-20}$	–	0.030	0.012
		(0.020)	(0.011)
$\gamma_{\tau=-30}$	–	0.013	0.011
		(0.016)	(0.014)
$\gamma_{\tau=-40}$	–	0.006	-0.017
		(0.015)	(0.013)
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	0.779	0.784	0.780

D.4 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 construct hedonic house sales price indexes by neighborhood and by year. Based on the demographic changes of population census and trend-break analysis, I create an indicator variable for neighborhoods that underwent the tipping process.

Table 6: Treatment Effects for the Railway & Tipping for Housing Prices

	(1) log(house sales price)	(2) log(house sales price)
Transit access	0.121**	0.116**
	(0.052)	(0.052)
Tipping		-0.356***
		(0.087)
Neighborhood	Yes	Yes
Year	Yes	Yes
Observations	3397	3397
R-Squared	0.483	0.485

Standard errors in parentheses

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

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.5 Urban Wage Premium from Panel Data

$$\text{Earnings}_{it} = \alpha_i + x'_{it}\beta + d_t + u_{it} \qquad (4)$$

where i denote an individual i at time t; $earnings_{it}$ denote individual i's earning at time t, α_i is individual i's time-invariant characteristics, β is a time-variant vector of coefficients of a characteristic's marginal contribution to individual i's earnings, and d_t is a time fixed effect. The unobserved heterogeneityu_{it} is assumed to be normally distributed. I create an indicator variable for individuals living in minority-neighborhoods that "tipped."

Table 7: Urban Premium for Rural to NY migrants

	(1) Income	(2) Income	(3) Income
Urban (i.e. New York City)	6.680***	6.708***	6.708***
	(0.050)	(0.053)	(0.053)
Urban x African American		-0.336***	-0.170
		(0.129)	(0.144)
Urban x African American x Minority Nghd			-0.296***
			(0.113)
Observations	932718	932718	932718
R-Squared	0.046	0.046	0.046

Standard errors in parentheses

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