Racial Segregation, Landlords, and Urban Decay*

Tom Nicholas† Christophe Spaenjers‡

March 2026

Abstract

We study how racial segregation shaped Manhattan landlords' incentives, rent-setting practices, and maintenance decisions in the middle of the twentieth century. We construct a novel data set linking Census records to real estate transaction data, property images, and renovation records. We show that landlords operating in Black neighborhoods exercised substantial pricing power and faced weak incentives to incur the costs of maintenance or to limit overcrowding. Rents for Black tenants were more closely tied to household size than to measures of housing quality. Moreover, properties with Black tenants generated disproportionately high rental income relative to their market values. These short-term profit incentives had long-run consequences: properties housing Black residents in 1940 experienced fewer renovations and showed greater visual exterior deterioration by 1980. Landlord behavior thus perpetuated racial divides in economic conditions and housing quality.

Keywords: housing; rents; maintenance; urban decay; segregation; Manhattan

JEL classifications: R31, J15, N92

*We are extremely grateful to Asaf Bernstein, Steve Billings, Leah Brooks, Matthijs Korevaar, Thies Lindenthal, Ralph Lerner, Ronan Lyons, Albert Saiz, Anna Scherbina, Chris Stanton, Eva Steiner, and participants at AREUEA and CU Boulder seminars, the Oxford Real Estate Research Conference 2025, the pre-WFA Real Estate Research Symposium 2025, the NBER Development of the American Economy Summer Institute 2025, and the NV FEB (Dutch and Flemish Financial Economists Abroad) 2025 conference for helpful comments and suggestions; to Christine Riggle, Christine Luksic, Heather Oswald and Melissa Murphy at Special Collections & Archives, Baker Library and May Niiya for digitizing volumes of the *Real Estate Record and Builders' Guide*; and to Julian Boilen for access to 1.5 million images of properties in New York City. The Division of Research and Faculty Development at Harvard Business School and the Michael A. Klump Endowed Chair at the Leeds School of Business provided financial support.

†Harvard Business School; tnicholas@hbs.edu.

‡Leeds School of Business, University of Colorado Boulder; christophe.spaenjers@colorado.edu.

1 Introduction

Migration is reshaping cities across the world as population inflows interact with discrimination and other barriers to residential choice, leading to the spatial concentration of migrants and increased pressure on specific neighborhoods (Bryan, Glaeser and Tsivanidis, 2020). Places like Chicago's South Side, the suburbs of Paris, and the inner districts of Brussels have seen the growth of migrant enclaves characterized by high density and low incomes. Understanding the forces determining the future development of such neighborhoods is a central challenge for the design of place-based policies (Glaeser, 2011).

In the first half of the twentieth century, several northern U.S. cities faced a set of circumstances similar to that of migrant-receiving urban areas today. The Great Migration led to mass influx of Black households from the rural and urban South (Derenoncourt, 2022). Discrimination concentrated Black migrants in neighborhoods such as New York City's Harlem. Prior literature has documented the rise of segregation in the early twentieth century, and its effects on housing prices, rents, and access to homeownership across racial household groups (e.g., Cutler, Glaeser and Vigdor, 1999; Boustan, 2010; Akbar et al., 2022; Logan and Parman, 2025). Our paper complements this body of research by studying how the strong demand by constrained Black tenants shaped *landlords'* incentives, investment returns, and maintenance decisions, and through these channels, long-run neighborhood outcomes.

We study these issues in the context of pre-WWII Manhattan. Most newly arriving Black immigrants could realistically only hope to rent in areas with elevated existing concentrations of Black households. In 1930, more than 80% of the Black population in Manhattan resided in the Central Harlem-Morningside Heights neighborhood, creating focal demand.[1] Moreover, turnover was high; the 1940 Census data show that Harlem had the lowest proportion of residents living in the same house five years earlier. Black households in cities also had fewer opportunities for home ownership compared to suburban or rural areas (Collins and Margo, 2001). The upshot was that downward pressure on rents in Black neighborhoods was limited. As noted in a history of Harlem by Tritter (1998), "the massive migration of Blacks to New York supplied landlords with unprecedented demand for their buildings."

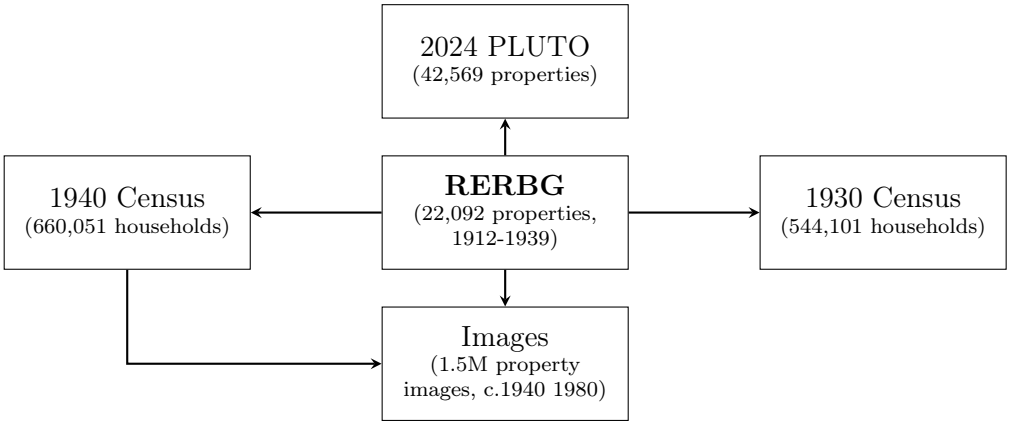
There also exists anecdotal evidence that property owners in Black neighborhoods maximized short-term profits at the expense of longer-run property value preservation. At the start of the twentieth century, when Harlem began transitioning to a Black neighborhood, it offered high-quality housing, and the promise of a thriving Black community, not a "ghetto" (McGruder, 2015). After 1920, however, the neighborhood started suffering from overcrowding. Moreover, historical accounts argue that landlords, even while charging high rents, were "increasingly unwilling to maintain their properties, and allowed them to deteriorate" with "crucially needed

[1] Also, according to the 1940 Census, this neighborhood had the highest proportion of individuals who had migrated from states not bordering New York. Migration rates from states associated with the Great Migration were nearly 11 times higher in Harlem compared to the affluent Upper East Side. The statistics in this paragraph are the authors' own calculations based on the 1930 and 1940 Census records described in detail in Section 2.

maintenance and up-keep declined" (Rosenbaum and Friedman, 2006). Many landlords did not live in the area, and "were no longer interested in the quality of the neighborhood or even the appearance of their own buildings. Their only concern was profit" (Tritter, 1998).

The qualitative data thus describe the pre-WWII Manhattan rental market as one in which Black households faced strong competition in their search for housing, and landlords renting to Black tenants in turn had substantial pricing power—only constrained by tenants' ability to pay—with few incentives to maintain the quality of their properties. In this paper, we test for quantitative evidence of systematic within-city heterogeneity in rent-setting practices, income returns, and maintenance decisions consistent with these observations. We also study the resulting disparities in housing decay in the decades after WWII.

Our analysis is made possible by a major data collection exercise. At the neighborhood level, we digitize statistics on housing conditions for Manhattan from the 1940 Census. At the property level, we construct a new database with the four main components shown schematically below. The different sources are linked using residential addresses, with households aggregated to the property level. (Appendix Table A1 reports matching rates across datasets.)



Outline of Linked Datasets

First, we applied OCR to the digitized corpus of real estate transactions from volumes of the *Real Estate Record and Builders' Guide* (henceforth RERBG) from 1912 to 1939. We observe comprehensive information on the transaction value of properties along with property-level characteristics such as square footage, number of stories, and building type. Our final dataset includes 23,690 transactions for 22,092 properties. Second, we access a dataset of over 720,000 images of properties in New York City taken around 1940 by the Works Progress Administration and NYC Tax Department property tax assessment authority. Manhattan properties were photographed again for tax purposes by the Department of Finance in the 1980s (around 800,000 images), allowing us to examine changes in property features and quality over time. Third, we use the 1930 and 1940 federal Censuses, which provide household rent data and individual-level demographic characteristics, including race, for Manhattan's population. The Census includes

residential addresses, so we can match properties and households, to our geocoded transaction and image data. We aggregate reported rents paid at the property level to estimate the owner's gross rental income. Finally, we link our property transactions, images, and Census data to the 2024 Primary Land Use Tax Lot Output (PLUTO) dataset from the NYC Department of City Planning, which provides additional information on properties, including original construction years. We augment the PLUTO data with data on building alterations collected from the NYC Buildings Information System.

Our empirical analysis using this novel data set is composed of three main parts. First, we study heterogeneity in rent-setting practices and incentives for investments in property quality in Black vs. white neighborhoods prior to WWII. Second, we estimate property-level heterogeneity in rental income returns earned by landlords as a function of tenant race. Third and last, we assesses disparities in maintenance and property quality in the decades after WWII.

The first part of our analysis examines the determinants of rents in Black vs. White neighborhoods in 1940. In line with landlords having pricing power and facing weak market discipline regarding maintenance, we find that rents and vacancies were much less sensitive to housing quality in Black neighborhoods. We also document that rents for Black tenants were more closely tied to household composition, particularly crowding through lodgers, than for White tenants. Taken together, our evidence suggests that landlords renting to Black households encouraged higher occupancies and set rents based on household size and tenants' ability to pay rather than on property quality.

In the second part of the paper, we estimate heterogeneity in rental income returns ("yields") as a function of tenant race. To estimate yields at the property level, we need property valuations. We predict market values of properties at the start of Census years 1930 and 1940, because it is for those years that we will estimate rental income returns. We use two types of valuations. A first valuation algorithm is based on a random forest that allows for non-linearities and interaction effects between hedonic characteristics, and generates the relevant valuations for both Census years for those properties in the RERBG ("random forest-based valuations"). Our second valuation algorithm is based on the property images that we have collected. We follow the approach in Glaeser, Kincaid and Naik (2018), and use the pre-trained ResNet CNN architecture developed by He et al. (2016) to extract high-level image feature vectors for each property in 1940. We then condense these features into 100 principal components. We first confirm that image features are predictive of property prices. We then apply a random forest to our principal components, and generate valuations for all properties in the year-1940 Census for which we have images ("image-based valuations"), not just the properties for which we have information on characteristics from the RERBG data.

As the next step in this part of our analysis, we estimate rental yields and examine their variation. To calculate net rental income (the numerator of rental yields), we adjust the aggregated household rent at the property level, computed earlier using the 1930 and 1940 federal Censuses, for estimates of maintenance costs, insurance premiums, and property taxes. We use

the random forest-based valuations and image-based valuations just described as our estimate of property values (the denominator of rental yields) for 1929 and 1939 so the denominator reflects the cost basis of the property before the rental income is realized. We find that on average rental yields were high, consistent with Jordà et al. (2019), who present aggregate data on U.S. real estate returns, and Lyons et al. (2024), who document elevated gross rental yields across multiple U.S. cities during the twentieth century. More importantly for our research question, we find that Black households lived in properties that generated rental yields for their owners that were substantially higher after controlling for tenant demographic characteristics and property characteristics. The difference in (property-level) rental yield between Black and White household heads is about 0.9 percentage points in our baseline specifications relying on random forest-based valuations for 1940, or 2.0 percentage points using image-based valuations. Using the equivalent valuations for 1930 the difference is 2.5 percentage points. Decomposing the rental yield results shows that properties housing Black residents were valued systematically lower while rents were relatively higher than those paid by White households. Landlords in Black neighborhoods were thus earning substantially higher cashflows from their rental properties than owners of otherwise equivalent properties in White areas.

In the third and last part of the paper, we explore the potential long-term housing inequality implications of high demand by Black households combined with landlords' weak incentives to incur the costs of maintenance or to limit overcrowding in this segregated housing market. We use tract-level housing statistics from the 1960 and 1980 Censuses, and show that the quality of the housing stock decayed more in Black neighborhoods. Next, we show using micro-level data that properties housing only Black residents in 1940 experienced an almost four-year delay in structural alterations that could have enhanced their market value. This delay persists even after controlling for redlining districts, suggesting that any systemic disinvestment associated with redlining does not fully explain the pattern.

To assess long-term changes in property values, we predict 1980 property values by applying the coefficients from a regression of 1940 property values on the principal components of the 1940 image vectors to the 1980 principal components, assuming that the pricing of exterior property attributes does not change too much over time. Properties housing Black residents in 1940 rank significantly lower in value by 1980. We also use our image vectors to analyze property decay over time. We find lower cosine similarity between the 1940 and 1980 principal components of the feature vectors for properties with Black residents in 1940, indicating greater changes in visual characteristics, which we interpret as structural decline for these properties.

Overall, our findings of stark racial disparities in property maintenance and visual patterns tied to structural decline highlight how landlords' behavior perpetuated disparities in property conditions and neighborhood segregation. These patterns suggest that differential incentives for property upkeep contributed to the long-term deterioration of Black neighborhoods, reinforcing economic and racial divides well beyond 1940.

1.1 Related literature

We make four main contributions to the literature. First, a large body of work has studied the rise of segregation in northern cities around the start of twentieth century, and its underlying drivers (e.g., Cutler, Glaeser and Vigdor, 1999; Boustan, 2010; Logan and Parman, 2017). In a recent paper, Akbar et al. (2022) study how the expansion of segregation led to the destruction of Black wealth as rents in newly Black neighborhoods soared and Black households also overpaid for homes.[2] Lee (2022) describes how urban planning policies amplified segregation dynamics, and led to substantial property value decreases in neighborhoods that "flipped" to African American majorities. Our perspective differs in that we focus on how segregation shaped incentives, investment outcomes, and decision-making for (multi-family) rental property investors, and how these in turn had a long-term impact on racial divides in housing quality and neighborhood dynamics.

Second, our work speaks to drivers of long-run urban decay, on which less research exists than on the growth and gentrification of cities. Most existing work has studied urban decline through the lens of urban "flight" and population dynamics more generally (e.g., Cullen and Levitt, 1999; Glaeser and Gyourko, 2005; Boustan, 2010), although some sociological accounts focus on the role of discriminatory practices and racial segregation (e.g., references in Travis (2019)). We document how patterns of segregation and the resulting high demand in Black neighborhoods decreased landlords' incentives to maintain their properties. While in a competitive market rental prices will be a function of property quality characteristics (Rosen, 1974), landlords with pricing power may have lower incentives for quality provision (Spence, 1975). A lower price-quality gradient has previously been shown to reduce maintenance in rent-controlled markets, with spillovers to nearby properties (e.g., Gyourko and Linneman, 1990; Sims, 2007; Rossi-Hansberg, Sarte and Owens, 2010; Autor, Palmer and Pathak, 2014).

Third, a number of recent papers have collected historical data to estimate the long-term returns to real estate (e.g., Jordà et al., 2019; Chambers, Spaenjers and Steiner, 2021; Eichholtz et al., 2021; Lyons et al., 2024; Schmelzing, 2025). It can be challenging to estimate rental yields because price and rent data are often not available for the same properties; also, contractual rents do not always reflect realized income. By linking transaction prices to tenants' reported rent aggregated to the property level, we provide new estimates of realized rental yields in pre-WWII Manhattan, one of the largest real estate markets in the United States. We also study heterogeneity in yields as a function of neighborhood and tenant characteristics. As such, we relate to recent papers arguing that total returns to real estate investments are higher in lower-value market segments (Demers and Eisfeldt, 2022; Damen, Korevaar and Van Nieuwerburgh, 2025) and that renting to poor households is potentially more risky but also more profitable (Desmond and Wilmers, 2019).

[2] There also exists a fast-growing body of work on racial differences in the cost of housing and in both access to and the returns to homeownership in more recent time periods (e.g., Bayer et al., 2017; Christensen, Sarmiento-Barbieri and Timmins, 2021; Kermani and Wong, 2021; Higgins, 2023; Box-Couillard and Christensen, 2024; Diamond and Diamond, 2024; Gupta, Hansman and Mabille, 2024; Ambrose et al., 2025).

Fourth, and finally, our paper contributes to the rapidly expanding field of deep learning (Dell, 2025). Image-based machine learning has become a tool to understand urban change (Naik et al., 2017) and price formation in real asset markets (Glaeser, Kincaid and Naik, 2018; Lindenthal and Johnson, 2021; Aubry et al., 2023), and we contribute to this strand of the literature through our use of historical data.[3] Recent literature has also used computer vision techniques to examine the evolution of cultural and societal norms (Adukia et al., 2023; Voth and Yanagizawa-Drott, 2024). By linking property images to Census data, we shed light on how visual data can enhance our understanding of neighborhood evolution, property investment decisions, and long-term socioeconomic change, and also open up new avenues for studying housing market choices and outcomes at the household level.

2 The Data

2.1 Housing Census Tract Statistics

To capture neighborhood-level housing conditions we construct a tract-level dataset for Manhattan based on the *Housing Bulletin for New York* supplement to the 1940 Census. Census tracts provide a standardized geographic unit designed to approximate relatively homogeneous areas. In 1940, Manhattan was divided into 266 census tracts with a mean of 215 structures and 2,520 units per tract.

Figure 1 shows an excerpt of the underlying data. For each census tract we collect data on vacancy rates, the distribution over age categories, the fraction of occupied units with non-White households, the faction of occupied units with more than 1.5 persons per room (a measure of overcrowding), the percentage of units needing major repairs, the percentage of units having no private bath, and average monthly rents.

Table 1 illustrates differences in housing characteristics between predominantly White and predominantly non-White census tracts, limiting the analysis to tracts with at least 200 units. We see that non-White tracts have an older and lower-quality housing stock, and also see more more overcrowding. Vacancy rates are lower on average in non-White neighborhoods, but the difference is not statistically significant. Rent levels are very similar.

2.2 The Real Estate Record and Builders' Guide

The RERBG includes transactions where confidential information on the actual sale price could be obtained (Grebler, 1955). Market-based transactions represent the prices buyers are willing to pay for specific features, locations, and amenities at any point in time, thereby also providing a basis for estimating the implicit value of property characteristics in valuation models.

[3] While Glaeser, Kincaid and Naik (2018) find that aesthetic appeal (captured by visual data) plays a more significant role in property price determination in wealthier neighborhoods, we will demonstrate how these features can also highlight structural decay in poorer neighborhoods. This underscores the context-dependent importance of visual data in understanding and predicting property values.

Given our focus on the pre-WWII period, we used digitized collections of the RERBG from Columbia University's digital archives for 1912–1922 and extended the dataset to 1939 through our own efforts. We extracted 1,833 pages of transactions from the Columbia archives and digitized an additional 3,864 pages ourselves. We begin in 1912 because, from that year, the layout of the RERBG is more conducive to OCR—see the differences between Panel A and Panel B of Figure 2. Nicholas and Scherbina (2013) manually entered 30 transactions per month from the same source. Modern digitization methods allow us to capture a much larger number of transactions (Panel C).

We ensure that the transactions included in our analysis are actual market sales, as described in the RERBG, excluding those that are non-monetary or between related individuals. From the digitized pages we extract all the relevant information on property transactions including the transaction value, the assessed value, whether the transaction was a foreclosure, the square footage of the lot, the type of property, the construction material, whether it had a store front, and the number of stories. In Panel B of Figure 2 we see that 1026 Park Avenue, in the Upper East side, is a 20 $\times$ 70 lot with a five-story tenement built from stone that had an assessed value of \$31,000-35,000 (land or land plus building) and transacted for \$45,000 in August 1920. We know this property transacted at market price, in contrast to 294-96 Bowery in the Lower East Side, which is designated as a "O C & 100" transaction, which the RERBG codifies as "other consideration & 100." 1026 Park Avenue transacted again in March 1936 and again in February 1937, though the vast majority of properties (87%) in our dataset transact only once (Panel D).

We estimate the total gross square footage of a property by the convention used in the PLUTO database of multiplying the lot area by the number of stories. This method is an approximation, since properties can have setbacks, open spaces, or partial coverage.[4] For the 1026 Park Avenue property, the lot dimensions are square, making the calculation of total gross square footage straightforward. For properties with irregularly shaped lots, we approximate the area by squaring off the average of the lot dimensions.

Our dataset contains 23,690 market-based transactions. Figure 3 shows the location of these properties using the modern ArcGIS locator file for geocoding. Despite potential complications from historical address changes, this method has been proven to yield accurate results (Barr and Tassier, 2016; Gray and Bowman, 2021).

Descriptive statistics on our data are shown in Table 2. The largest share of transactions occurs in the Union Square-Lower East Side neighborhood (15.1%), home to large immigrant communities, followed by Central Harlem-Morningside Heights (14.8%) a major center for Black residents during the early twentieth century. Lower Manhattan, which was already heavily commercial by this time, contains the smallest share of transactions (4.3%), consistent with the fact that commercial properties tend to have longer leases and transact less frequently

[4]"New law" residential buildings in New York City constructed after the enactment of the 1901 New York Tenement House Act were required to have courtyards in order to allow for ventilation and natural light in every unit. "Old law" buildings often did not have courtyards and thus occupied a larger fraction of the lot.

than residential properties, as they are more often owned by institutional investors with longer investment horizons. The most common property type is tenements (47.5%), followed by 'other' dwellings or commercial buildings (22.9%), dwellings (20.6%), and lofts (8.9%). Most properties were constructed with brick (63.1%), while stone and other materials accounted for 13.1% and 23.8%, respectively. 38.9% of properties included a store, which is consistent with Grebler (1955) who shows about 44% of apartment properties contained a store. Building height varied, with 30.7% of properties having six or more stories, indicating the inclusion of particularly large properties in the dataset. Lot sizes averaged 18.73 thousand square feet, while total square footage (lot and building) averaged 81.96 thousand square feet. The average property transaction value is $40,301 (about $700,000 today). Foreclosure transactions account for 34.9% of our data as identified by the text string "FORECLOS" in the RERBG listings; given the prevalence of foreclosures, we will include them in the construction of our indexes while using an indicator variable to account for their specific value.[5] Each property when it transacted was an average of 0.26 miles to the closest subway (subsection 2.5).

2.3 Property Images

We accessed images from a remarkable initiative during the Great Depression. Photographs of Manhattan properties were taken as part of a large-scale effort initiated by the New York City Department of Taxes and supported by the Works Progress Administration (WPA). Known as the Real Property Tax Card Record System, the project aimed to modernize property assessments by documenting all parcels of real estate with tax cards that included photographs, diagrams, and detailed building data (Esperdy, 2004). This was intended to facilitate equitable taxation and administrative efficiency. The project provided employment for about 900 clerical workers and photographers during a period of high unemployment, making it both a practical and relief-driven initiative.

The resulting collection of 720,000 images captures almost every building across the five boroughs from 1939 to 1941. (The five boroughs contained approximately 815,000 parcels of real estate in 1938.) The photographs are particularly useful for understanding the architectural and urban landscape of the time, as they provide detailed, visual records of building types, materials, and uses. While the angle of the photos was sometimes dictated by the building's size and the street's width, necessitating oblique perspectives in certain cases, the images offer insights into the diversity of urban structures with reasonably standard framing perspectives. Designed to assist tax assessors with valuations, they are by construction a visual reflection of underlying property value.

[5] Panel A of Appendix Figure A1 shows that the share peaks around the time of the First World War and during the foreclosure crisis associated with the Great Depression (also see Ghent (2011)). Panel B shows an upward trend in foreclosure property values during the Great Depression. We know savings banks purchased properties at foreclosure auctions, primarily in cases where they acted as the plaintiff. In essence, these banks were acquiring properties to protect their mortgage liens. The RERBG also published maps that detail their frequency and location (see Appendix Figure A2).

While our analysis centers on the buildings themselves as objects of valuation, the photographs also capture subtle elements of urban life, such as lampposts, signage, vehicles, and street level activities, that can serve as proxies for location attributes and amenities, which are often difficult to quantify through property characteristic variables alone. Fortuitously, the Department of Finance undertook a project to retake the tax photographs during the 1980s—a further 800,000 images. Thus we have visual features of properties we can compare from both snapshots in time. A limitation of the images is, of course, that we do not observe interior conditions.

Figure 4 provides examples of the images in our dataset. 415 West 118th Street in the Morningside Heights neighborhood of Manhattan, which transacted for $110,000 in August 1914 in the RERBG, is shown in 1940 and 1980 alongside a current image showing its ownership by Columbia University.[6] The 1980 images are of lower quality because the original 35mm photos were converted into a digital format by the Department of Finance and stored on Laser Video Discs, an early optical storage technology. The next set of 1940 images in Figure 4 shows a property with a store front, as well as images of properties with cars or a person walking by. Each 1940 image includes a placard with an identifier code for the building and lot number, which is included in the filename of the images, which we can use to match with the RERBG and Census data. Some photos, as shown in the bottom panel of Figure 4, contain defects, or in the case of the 1980 collection, a substitute card was placed where an actual photo was missing. These defects can largely be identified and cleaned through their corresponding feature vectors in our visual machine learning analysis (subsection 4.1).

2.4 Census Data and Redlining

We use federal Census data as our source of rent information. In the 1930 and 1940 Censuses, households indicated whether their home was owned or rented. Renters provided the amount of their monthly rent, while homeowners were asked to estimate the potential sale price of their property.[7] The majority of the properties in our dataset were rentals since Manhattan has historically been a rental-dominated market.

What information do the rents convey? Rent control became widespread during the Second World War. Fetter (2016) estimates that approximately four-fifths of the rental housing stock in 1940 was located in areas subjected to federal rent control measures between 1941 and 1946. Before this, rent control laws existed primarily on an *ad hoc* basis and, in Manhattan, relied heavily on judicial discretion. Günnewig-Mönert and Lyons (2023) document how judges were empowered to decide whether rent increases were "reasonable" based on a law introduced in 1920, which was subsequently abolished during the late 1920s. Although agitation for rent control continued throughout our period, we assume for the 1930 and 1940 data we have from

[6] For simplicity, we will refer to the 1939–1941 images as "1940" images and the 1980s images as "1980" ones.

[7] Beyond Census data, comprehensive rent information is available from newspaper listings, as highlighted by the major effort of Lyons et al. (2024) to collect these data for a multitude of U.S. cities.

the Census that rents approximately reflect market prices.

To illustrate our approach to matching price with rent data, consider our example of 415 West 118th Street in the RERBG shown in Figure 4. In 1940 this property was rented by 20 households containing a total of 60 individuals. 54 of the individuals were White and six were Black all of whom belonged to the same household. Household rents ranged between $45 and $70 per month. Three households did not report rent, including the Black household, but we do not find that missing rental data is systematically a function of race.[8] We impute missing rents based on the average rent of the reporting households for each property. We then aggregate at the property level, with annual rents for 415 West 118th Street totaling $12,732.

The richness of the Census data means we will be able to estimate rental yields at the property level conditional on tenant demographic characteristics. The 1940 Census contains data on wages and education which is particularly useful as we are interested in housing affordability, economic conditions, and social inequality. Louise Kilburn, a White 37-year old waitress with a high school education who earned $1,000 during the prior year, lived at 415 West 118th Street with her 11-year old child and 75-year old aunt. Richard Wildman, a naturalized Black resident with a high school education from the West Indies lived there with his spouse, Hazel, and four daughters. Wildman appears to have been the superintendent of this, or another, apartment building based on the occupation string in the Census and his stated earnings are $1,200. We will focus much of our analysis on the 1940 Census because it provides richer demographic variables. In contrast, the 1930 Census offers more limited data, with literacy serving only as a basic indicator of educational attainment. Wages are not observed at all.

Based on the Census data we can see how segregated Manhattan was. Figure 5 shows that the Black population overwhelmingly resided in the Central Harlem-Morningside Heights neighborhood, especially relative to the more dispersed distribution of the White population and other racial groups. We also incorporate spatial data from Home Owners' Loan Corporation (HOLC) maps (1935–1940) from *Mapping Inequality: Redlining in New Deal America* to classify neighborhoods by redlining designations. In 1940, 92% of the Black population of Manhattan resided in a redlined area which were designated as "hazardous" according to perceived investment risks, compared to 55% of Whites.[9]

When geocoding the RERBG and Census addresses, we use the same ArcGIS location file and then match all pairwise combinations of properties based on their calculated distances. Census address strings are often messy, with numbers frequently spelled out as text (e.g., 'One Hundred Eighteenth Street' instead of '118th Street'). Since many Manhattan addresses include numerical components, we first clean the addresses using text algorithms. Given the high density of properties in Manhattan, we apply a strict matching threshold, requiring a

[8] We estimate a linear probability model with an indicator for missing rents as the dependent variable and race indicators as independent variables. The coefficient on Black households heads is -0.012 (s.e. 0.007) and on Other is -0.008 (s.e. 0.010) with White household heads as the baseline.

[9] See Appendix Figure A3 for more details. Appendix Figure A4 plots each geocoded Census observation in the data. Shaded areas in the 1940 map are Grade D areas or "redlining" areas on HOLC maps created during the late 1930s and early 1940s.

distance of zero meters between properties. Descriptive statistics for the matched RERBG-Census data are shown in Appendix Tables A2 and A3 for 1930 and 1940, respectively.

Our final regression datasets include only the large majority of residential properties that only had tenants living in them, as we cannot accurately impute the rental value of owned units in mixed-use properties where systematic differences could exist between rented and owned units. For example, rented units may be smaller, less renovated, or located on less desirable floors, while owner-occupied units might receive more investment in maintenance and upgrades.[10]

2.5 Subways

In cities like NYC, the subway has historically driven urban expansion.[11] We plot the locations of subways which diffused across Manhattan during the early twentieth century. The inaugural line was built by the Interborough Rapid Transit Company and stretched from City Hall in Lower Manhattan to 145th Street and Broadway in Harlem. We use spatial coordinates for NYC subway stations provided by the Metropolitan Transportation Authority. We supplement this with our own data collected on the opening of each station, from the Broadway–Seventh Avenue Line's opening in 1904 to the opening of the East Broadway station on the Sixth Avenue Line in 1936. For each transaction year (between 1912 and 1939) and Census year (1930 and 1940), we estimate the minimum distance from a property to the network of subways operating in that year. In 1912, 41 subways were in operation, increasing to 130 by 1939.

2.6 PLUTO Database and Alterations

As a final step in our data collection, we link all our property-level data to the most recent version of the PLUTO (Primary Land Use Tax Lot Output) dataset, using both geocoding or matching by the building block and lot identifiers. PLUTO contains detailed information about every tax lot in New York City, including Manhattan, and is widely used for urban planning and research. An advantage of this dataset is that it contains information on the approximate year that properties were built, which is a proxy for unobserved quality, including construction

[10] Panels A and C of Appendix Figure A5 shows the vast majority of properties were fully rented in Manhattan in both 1930 and 1940 due to the city's vertical development. The prevalence of multi-unit buildings and tenements, especially in the late nineteenth and early twentieth centuries, catered to renters rather than individual homeowners. Panels B and D present estimated marginal effects of race on property type based on a multinomial logit model. These marginal effects represent the change in the probability of each property type associated with the household head being Black versus non-Black, controlling for property, individual, and spatial characteristics (latitude and longitude). The results indicate that Black households were significantly more likely to reside in rental properties, equally likely as non-Black households to reside in owner-occupied properties, and less likely to live in buildings with mixed ownership and rental households. These figures underscore the predominance of rental housing in Manhattan and racial disparities in access to property types.

[11] The poor tend to live in cities because transportation costs are lower (Glaeser, Kahn and Rappaport, 2008), while property values tend to increase generally with proximity to rail systems due to gentrification (Kahn, 2007) and lower commuting times (Gupta, Van Nieuwerburgh and Kontokosta, 2022). In New York City, Lee (2022) observes a significant increase in segregation through residential sorting by income following the construction of transit infrastructure.

standards, architectural styles as well as ageing and depreciation. The PLUTO database also contains contemporary assessed values. We augment these data with our own collection of administrative data on building alterations from the NYC Buildings Information System.

3 Rents, Occupancy, and Housing Quality

3.1 Rents and Housing Quality

If landlords in Black neighborhoods had more pricing power, and were only constrained by their tenants' ability to pay, then rents should be less responsive to housing quality in those areas. To examine this, we construct a tract-level measure of structural condition in 1940 by taking the first principal component of multiple indicators of housing deterioration from the Census housing data described in Section 2.1 % old units (built pre-1920); % overcrowded units (>1.5 pax/room); % of units needing major repairs; % of units without private bath. This first principal component accounts for 43.4% of the total variance across the individual indicators. All four variables load positively on this component. The resulting index, which we Z-standardize for interpretability, captures an implied maintenance burden, with higher values indicating worse structural condition. Not surprisingly, given the descriptive statistics in Table 1, we find a positive 0.20 correlation between this index and the non-White share across tracts.

We then regress log rents on this condition index, the tract share of non-White households, and the interaction between the two. Figure 6 plots predicted rents for tracts that are all White and all non-White. A one standard deviation improvement in structural condition is associated with a 46.2% increase in rents in all White tracts, but only an 11.7% increase in all non-White tracts. These results imply substantially lower marginal returns to maintenance in Black neighborhoods.

3.2 Vacancy Responses to Housing Quality

To assess whether market discipline differed by race, we examine how vacancy rates respond to housing quality across census tracts in 1940 using the same neighborhood-level data source. Unconditionally, vacancy rates are only slightly lower in non-White neighborhoods (cf. Table 1). However, responsiveness to housing quality differs by race. Figure 7 shows that vacancy rises sharply with the implied maintenance burden in predominantly White tracts but responds only weakly in heavily non-White tracts. Like rents, vacancy rates are substantially less sensitive to housing quality in non-White neighborhoods, implying weaker maintenance discipline faced by landlords in those areas. Landlords had little incentive to improve properties in neighborhoods where tenant demand remained high despite poor housing conditions.

3.3 Household Composition and Rents

We now turn to household size as a mechanism through which landlords with Black tenants exploited their market power, using household-level data from the 1940 federal Census. As shown in Figure 8, Black households had on average more wage earners per household. Table 3 further shows earning lodgers (i.e., lodgers, boarders, or roomers who report positive wage income) are far more prevalent in Black only properties (36.3%) than in White only properties (9.7%). This evidence points to landlords allowing a higher number of occupants per unit, and tenants pooling resources.

Did landlords allow for larger households because it enabled them to charge higher rents? To assess this, we examine how much explanatory power is added by different sets of controls in rent regressions estimated separately by property type. Starting from a baseline specification that includes neighborhood fixed effects only, we sequentially add observable property characteristics and household-level variables, and track the resulting changes in R^2. Results are reported in Table 4. Adding observable property characteristics increases R^2 by a similar amount across White only and Black only properties (8 to 9 percentage points), while household characteristics play a markedly different role by race. In particular, earning lodgers account for little additional variation in rents in White only properties (2.6 percentage points) but explain a large share in Black only properties (15 percentage points). Household wages also explain more variation in rents in Black neighborhoods, though much less than earning lodgers.[12]

Taken together, the results in this section indicate that rents for Black tenants were more closely tied to household composition, particularly crowding through lodgers, than to housing quality or upgrading. Landlords priced rents based on total household size and earnings potential, capturing a larger share of pooled resources in increasingly overcrowded neighborhoods.

4 Rental Yields and Race

We now turn to studying whether landlords earned higher rental income returns ("yields") on properties housing Black tenants, which would provide further evidence of landlord market power. To answer this question, we first generate new property-level valuations (subsection 4.1). We then present our main rental yield estimates (subsection 4.2), which are derived from our property valuations and Census rent data. We then introduce our primary regression specification for examining variation in rental yields (subsection 4.3) and discuss the results using 1940 Census data (subsection 4.4). Next, we conduct robustness checks: substituting 1930 Census data for 1940 (subsection 4.5) and using image-based valuations (subsection 4.6).

[12] If rents in Black neighborhoods were tightly constrained by household resources, dispersion in rents conditional on wages should differ by race. Appendix Figure A6 plots residual log rents from a pooled rent wage regression against wages separately for White and Black household heads. Residual rent dispersion is about 24% lower for Black households than for White households, indicating that, conditional on earnings, Black households faced a narrower feasible range of rents.

4.1 Predicting Property Values for Census Years

We have federal Census data on rents in 1930 and 1940. To estimate the rental income return realized by landlords, we will need to link property-level aggregate rental income to estimates of properties' market values at the start of those Census years. To conclude this section on property pricing, we demonstrate how we can use transaction data from 1912 to 1939 and image data from 1940 to value either those properties in the RERBG (when using information on property characteristics) or all properties in the Census (when using images) in both 1929 and 1939.

4.1.1 Random forest-based valuations

Let us start by describing how we use information on property characteristics. One weakness of using standard hedonic models for valuation purposes is that they impose a (log-)linear and additive structure, while in reality non-linearities and interaction terms might be important for pricing of real assets (Aubry et al., 2023; Koijen, Shah and Van Nieuwerburgh, 2025). We therefore incorporate a random forest approach to capture such complex relationships between features such as location, building type, and square footage in order to improve predictive accuracy of the valuation model. Random forest predictions are aggregated across 100 decision trees, mitigating the risk of overfitting inherent in individual decision trees (Breiman, 2001). Each tree is trained on a random subset of the data and features, with a random subset of predictors at each split to reduce correlation between trees. This ensures that individual trees capture different patterns in the data, improving stability and predictive accuracy.

Since our transaction data includes a temporal component, we train a prediction model and validate it out-of-sample to ensure the model can generalize across different points in time. We start by assessing in general the out-of-sample performance of a random forest model, comparing it to that of a standard hedonic model, with the data split into 80% training and 20% testing sets. As predictors we use property characteristics, the ratio of assessed-to-market value, year of the transaction, latitude and longitude of the property, and a foreclosure indicator. The random forest model consistently outperforms an OLS hedonic model, demonstrating better predictive accuracy with lower average deviations from actual property values, reduced prediction errors, and improved percentage error metrics.[13] These results suggest that while both models effectively capture key predictors, the random forest model provides more precise and robust predictions, likely due to its ability to account for complex, non-linear relationships.

In estimating random forests to predict property values ("random forest-based valuations") for use with the 1930 and 1940 Census rental data, we make two key adjustments to our approach. First, we restrict the training model to use data from only the prior five years, so the training specification reflects more recent market conditions. Second, we exclude foreclosures from the training model, focusing solely on market prices, as these are the values we predict

[13] Appendix Table A4 compares accuracy and predictive performance on the test set. Appendix Figure A7 further illustrates the stronger predictive performance and better fit of the random forest model.

for the cross-section of properties in the RERBG. We then value foreclosed properties using the prediction model as if they transacted at market prices. To minimize missing data, we also drop the assessed-to-market value variable.

4.1.2 Image-based valuations

We now turn to our database of property images as a novel resource for valuation. We do this for two reasons. First, if the images contain sufficient meaningful information about properties' market values, we can put a valuation on *all* properties for which we have rent information in the Census and not just those properties for which we observe a transaction in the RERBG, allowing us to estimate rental yields on a more exhaustive set of properties. This broader scope mitigates selection bias, as transacted properties may systematically differ from those that did not transact (e.g., by having different rent levels). Second, if images correlate with prices, then changes in image features between 1940 and 1980 can tell us something about disparities in long-term price dynamics.

We analyze what the tax assessor saw to extract a quantitative signal about property value from the images. To do this, we employ a modern convolutional neural network (CNN) architecture to represent the images with pixel-based feature vectors. We follow the approach in Glaeser, Kincaid and Naik (2018) and use the pre-trained ResNet model to extract high-level features and patterns present in the image. The 1940 and 1980 images are preprocessed and fed into ResNet to obtain feature vectors as markers of the visual characteristics and structural attributes of the properties. To account for multiple properties potentially appearing in a single image due to the photographer's positioning and perspective (and to focus on the building rather than street-level activity), we apply a mask to prioritize the central region. The resulting feature vectors are then reduced to 100 principal components, capturing over 80% of the variance of the 1940 and 1980 vectors. Importantly, ResNet is not trained to value properties; rather it is used to extract visual features that our models then link to property value. Hence, we avoid look-ahead bias in our empirical analysis. Although historical buildings are not in its training set, ResNet still captures universal characteristics such as edges, textures, and shapes, making it robust for feature extraction.

Visual machine learning models like ResNet are often considered a "black box" because, while they effectively extract features from images, the exact nature of these features and their relationship to specific outputs (like property value in our case) can be difficult to interpret meaningfully (Dell, 2025). We therefore provide visual evidence to illustrate what these feature vectors represent. Figure 9 visualizes the five "nearest neighbors" to each randomly selected 1940 image in the first column using the euclidean distance between the principal components of the feature vectors. Illustratively, proximate properties in vector space also exhibit strong visual similarity.

Moreover, the principal components of the image feature vectors are strongly correlated with property transaction values, as can be illustrated by computing the adjusted R-squared for

linear regression models where we use different sets of variables to predict the log of transaction values. If we run a hedonic model with fixed property and location characteristics, a foreclosure dummy, and year × quarter fixed effects, we get an adjusted R-squared of 27.9%. A model regressing log transaction prices against fixed attributes extracted from historical images using ResNet (i.e., the 100 principal components described above) achieves a lower but still remarkable adjusted R-squared of 19.3%.[14] This compares favorably to the 11.7% reported by Glaeser, Kincaid and Naik (2018) using Google Street View images and modern transaction data from Boston. The fact that our visual features explain more of the variation in sale prices is intuitive as historical images were designed to aid assessors in property valuation whereas images from Google Street View serve broader purposes and are not tailored for evaluating structural or market-related property features.

We then apply a random forest approach to the principal components from the 1940 images, along with latitude and longitude, to predict market values ("image-based valuations") for the full cross-section of properties in the 1940 Census.[15]

4.2 Rental Yield Estimates

We can now estimate rental yields because we aggregated rents over the households living in any property in subsection 2.4, and estimated in subsection 4.1 property values for years preceding Censuses in which rents are observed. Note that, in our data, vacancies are implicitly accounted for, as vacant apartments would not have been enumerated in the Census, and the method we use to aggregate rents at the property level accounts for this. Our yield estimates therefore reflect *effective*—not potential or contractual—rental income.

Beyond estimating gross rental yields, we can also estimate *net* rental yields. We first need to subtract maintenance costs and insurance costs from gross rental income to calculate operating profit.[16] Taxes, calculated as a percentage of the assessed value of the property, are then subtracted to arrive at after-tax profit, which we treat as equivalent to net cashflow. This provides a net measure of income generated from each property after accounting for observable expenses and taxes. (To isolate the intrinsic performance of the property as an asset, we abstract from financing decisions, including mortgage and transactions costs.) We measure each of these components using our 1930 and 1940 datasets. In both years, we set maintenance

[14] Note that we do not include a foreclosure or neighborhood dummies here. Appendix Figure A8 shows a binscatter plot showing a strong correlation between predicted and realized property values between 1912 and 1939, estimated using only the 100 principal components from the 1940 visual features. Combining hedonic characteristics and visual features in a linear regression model explains 30.7% of the variation in log transaction prices, demonstrating the complementarity of hedonic and visual characteristics, while a placebo test replacing meaningful features with random variables achieves only 3.8% with time period fixed effects and 0% without.

[15] Appendix Figures A9 and A10 illustrate strong in-sample correlations between actual and predicted values for properties transacted in 1929 and 1939, supporting the robustness of our approach.

[16] We assume that depreciation is accounted for within maintenance costs. An alternative approach would be to model depreciation as an annualized capital expenditure rate based on the property's market value; however, this is problematic for our time period. Grebler (1955) highlights that capital expenditures were often irregular, with deferred maintenance common during periods of financial distress.

costs at 25% of annual rental income, based on estimates from the *Report of the State Board of Housing of New York*.[17] To estimate assessed values, we use the ratio of assessed value to legal basis (market value), which is 90% in 1930 and 92% in 1940, as reported in the *National Municipal Review*.[18] Insurance costs are estimated at 0.57% of assessed value in 1930 and 0.53% of assessed value in 1940, using data from the *Weekly Underwriter*. Finally, we set the tax rate at $27.20 per $1,000 of assessed value in 1930 and $29.60 per $1,000 in 1940, also based on data from the *National Municipal Review*.

We present mean and median gross and net rental yield estimates for 1930 and 1940 in Figure 10, using as our denominator the random forest-based and image-based property valuations generated in subsection 4.1.[19] The random forest-based valuations imply a mean (median) net yield of 5.9% (4.4%) in 1930 and 13.0% (11.7%) in 1940, while the image-based valuations imply a mean (median) of 10.8% (9.0%) in 1940. Our estimated returns in 1940 are higher than those in 1930, largely due to the comparatively lower property prices in 1940, which contributed to higher rental yields. The results indicate that total real estate returns were primarily driven by the rental yield component, consistent with findings by Jordà et al. (2019), who report a national net rental yield of 5.3% in 1930 and 6.3% in 1940, and Lyons et al. (2024), who document gross rental yields of 13.6% in New York City between 1930 and 1945.[20]

4.3 Regression Specification

We can now examine variation in rental yields based on the race of individuals residing in these properties, as identified in federal Census data. Figure 11 shows that, on average, unconditional net rental yields were higher for properties housing Black residents compared to those housing White residents in 1930 and 1940. We can test more formally for differences in yields by estimating regression specifications.

Our main regression equation, estimated in the next subsection using 1940 Census data on rental properties, takes the following form, where i indexes properties and h indexes households:

[17] In other words, we are using the same relative hypothetical maintenance cost for all properties, and ignore the reality that landlords in Black areas underinvested in maintenance. Assuming a lower maintenance expense on properties housing Black tenants would only *increase* the disparity in yields that we document. In unreported analysis, we find that maintenance expenses would have to have been about two thirds higher in Black neighborhoods for our main coefficients below to change sign.

[18] RERBG sometimes has information on assessed values but not for every transacted property. We will also want to estimate yields for properties that did not transact.

[19] To minimize the impact of data errors on our means, we trim the returns at the 5th and 95th percentiles, as shown in Appendix Figure A11 for our random forest-based valuations. Grebler (1955) reports a range of net investment yields (including the capital component) from -4% to 23% for properties between 1930 and 1950, while our trimmed range for 1940 (excluding the capital component) spans from -1.5% to 49.5%, reflecting a longer tail of returns in our dataset.

[20] Grebler (1955) reports that net income on properties accounted for approximately 35% to 37% of gross income in New York City between 1935 and 1945. Our findings are broadly consistent with this range, with net rental yields representing about 33% of gross income in 1930 and somewhat higher in 1940, where net rental yields were approximately 49% of gross income. In the total return series presented by Lyons et al. (2024), New York City ranks third from the bottom in capital returns among 29 cities between 1930 and 1945, but seventh from the top in the ranking of gross rental returns. This underscores the critical role of rents as a primary driver of real estate income in this location.

$$\frac{\text{Net Rent}_i}{\text{Property Value}_i} = \rho_r \text{Race}_{h,r} + \eta_q \text{Wage Quartile}_{h,q} + \tau_e \text{Education}_{h,e} + \kappa \, \text{Subway Distance}_i + \sum_{j=1}^{J} z_{h,j}\zeta_j + \sum_{k=1}^{K} x_{i,k}\lambda_k + \epsilon_i, \quad (2)$$

We examine variation in inverse hyperbolic sine (IHS) transformed rental yields for property i, which can take on negative values, as a function of both household-level and property-level characteristics. Household-level variables are measured using the attributes of the household head and include indicators for race (Black, other, with White as the baseline). We also include income quartiles and education levels (with a baseline of 0 to 2 years of schooling) for which information was included in the 1940 Census. Additional household-level controls (ζ) include individual characteristics from the Census (age, sex, and employment status).

Property-level variables include the IHS-transformed distance to the nearest subway and a set of additional controls (λ), such as property square footage and type. Because property values are predicted using the same characteristics for the cross-section of RERBG properties, we introduce these variables as controls only in our final specifications. We exclude neighborhood fixed effects from our baseline specifications because racial segregation shown in Figure 5 limits identification. Standard errors are clustered at the property level to account for the hierarchical structure of the specification and potential within-property correlations such as shared characteristics among households.

4.4 Main Results Based on 1940 Census Data

Panel A of Table 5 presents the results of estimating Eq. (2) on the 1940 Census data. Column 1 shows a baseline racial disparity in net rental yields, with Black household heads living in properties that generate a yield for their owner that is about 1.8% (i.e., 1.8 percentage points) higher compared to White household heads.[21] Adding individual controls from the Census reduces this disparity slightly to 1.6% (column 2). Wage and education controls are not strongly correlated with yields, with the estimated racial disparity stabilizing around 1.7% (columns 3–5). Similarly, adding distance to the nearest subway does not change the yield estimate. While proximity to subways may increase property values through improved accessibility, it may also raise rents proportionally, leaving the rental yield unaffected (column 6).

In the final column of Panel A of Table 5, we incorporate property attributes to account for factors beyond household-level characteristics that may explain differences in rental yields. The coefficient for Black household heads is statistically significant but the magnitude is smaller at 0.9%. Note that property characteristics are used to predict property values in the random forest model, which serves as the denominator of rental yields. If these same characteristics

[21] Importantly, our results are not driven by our adjustments for maintenance, expenses, and taxes. As shown in Appendix Table A5, *gross* rental yields are also similarly higher for Black households.

also directly influence net rental yields (the dependent variable), this creates a mechanical link between the predictor and the outcome possibly absorbing part of the variation previously attributed to race. Furthermore, race may be more strongly correlated with certain property characteristics, such as tenement housing, for example, potentially attenuating the effect of race in this specification, especially if adding these controls also reduces identifying variation.

In Panels B and C of Table 5, we use the same specifications on the numerator and denominator of net rental yields to disentangle the factors influencing rents and property values. The net rent specifications in Panel B reflect the rents landlords collect at the property level, revealing patterns that help explain rental yield disparities. We find that properties occupied by Black households generate net rents that are generally larger from properties occupied by White households. The regression coefficients shown in Panel C indicate a consistently negative and statistically significant relationship between race and property values across specifications. Black household heads are associated with much lower property values compared to White household heads. This persistent gap in the market value of the household's residence remains significant even as individual controls, as well as education and wages, and property characteristics are added sequentially in columns 3 to 7. In column 7, which includes the full set of controls, we estimate that properties occupied by Black household heads were valued 8.9% lower than those occupied by White household heads.

Overall, we find that a hypothetical real estate investor would earn higher net rental yields from Black households, primarily driven by consistently lower property values. Black household heads tended to pay elevated rents compared to White household heads with similar characteristics, and lived in properties of lower value.

4.5 Robustness: 1930 Census Data

The main analysis presented in the previous subsection relied on property-level rental income derived from the 1940 Census. As a robustness check, we can repeat our estimations for properties that we match to the Census in 1930, when rental yields were substantially lower on average than in 1940 (Figure 10). In the 1930 Census, we do not observe wages and we can only capture education through an indicator for literacy. The results presented in Table 6 show a significant racial disparity in net rental yields, net rents, and property values. In Panel A, Black household heads are associated with 2.5% to 3.4% higher net rental yields for their landlords, with the coefficient decreasing as additional controls are added, yet remaining significant and of a larger magnitude than the estimates in Table 5. In Panel B, Black household heads are associated with significantly higher net rents, even as controls for literacy, subway distance, and property characteristics are introduced. This pattern is consistent with strong housing demand in 1930, driven by factors such as the Harlem Renaissance (that drew in Black intellectuals, artists, and professionals), the Great Migration, and limited housing supply (Tritter, 1998). Notably, the coefficient on literacy (not shown) in column 3 of Panel B indicates that rents for literate households are about double those for non-literate households, consistent with the

education gradient observed earlier. Meanwhile, Panel C reveals that properties occupied by Black households in 1930 had lower market values in our most demanding specification (column 5), just like in 1940. The consistent disparities we find highlight systemic differences in property valuation and rent-setting practices.

4.6 Robustness: Image-Based Property Values

We can also replicate and extend our 1940 analysis using the image-based property valuations described in subsection 4.1 as the denominator in our rental yield estimation. This approach significantly increases our sample size—nearly fivefold—by incorporating all properties matched to the Census, not just those that transacted. A limitation is that we cannot control for property characteristics beyond what is captured in the images. However, due to the larger sample size, we can estimate rental yields with greater precision and examine finer geographic and demographic patterns that were previously infeasible with the transaction-based estimates in Table 5.

Table 7 shows the results. In column 1 of Panel A, we estimate a racial disparity in net rental yields for Black households of about 2.7%, somewhat higher than the 1.8% reported in Table 5 and this estimate remains stable in columns 2 to 6 with added controls. This increase may reflect the inclusion of non-transacted properties, which adds variation or captures selection effects.

We introduce three new specifications. In column 7 of Panel A, we add neighborhood fixed effects (which we did not do before because of smaller sample sizes) and the racial disparity remains positive and statistically significant, though smaller in magnitude than in earlier specifications. In column 8, where we exclude Central Harlem-Morningside Heights—the area with the highest concentration of Black residents—we find that net rental yields for Black households remain about 2.0% higher. In column 9, which includes neighborhood fixed effects while also excluding Central Harlem-Morningside Heights, the coefficient is smaller in magnitude (1.0%), suggesting that much of the racial difference in rental yields may be driven by neighborhood-level factors, including segregation, rather than property characteristics alone.

In Panel B the larger sample shows smaller racial disparities in net rents to those in Table 5. The coefficient on Black household heads in column 6 is 0.080 compared to 0.148 in the equivalent model in Table 5. When we include neighborhood fixed effects (column 7), the estimate increases, reflecting higher net rents for Black households relative to non-Black households within the same neighborhood. The estimate remains statistically significant, albeit at a reduced level, even after excluding properties in Central Harlem-Morningside Heights (column 8). However, in column 9, the racial difference in net rents is no longer statistically significant, and the point estimate is smaller, suggesting that market dynamics in Central Harlem-Morningside Heights play a crucial role in driving the observed disparities.

In the property value specifications in Panel C, the point estimates are directionally similar to before, suggesting Black households resided in properties with 20.4% lower property values in

column 6 compared to 13.4% lower in Table 5. In column 7, adding neighborhood fixed effects reduces the racial gap in property values, indicating that most of the difference is driven by where Black households live rather than the value of their homes within a given neighborhood. In column 8, we find that racial disparities in cross-sectional property values persist outside of Central Harlem-Morningside Heights. In column 9, where both neighborhood fixed effects are included and Central Harlem-Morningside Heights is excluded, the racial gap disappears, implying that neighborhood sorting, rather than within-neighborhood differences, primarily explains the observed racial disparity in property values.

Crucially, while property values (Panel C) likely reflect long-term market conditions, rents (Panel B) may be more influenced by short-term dynamics, such as tenant turnover, demand fluctuations, and rent-setting practices, which could differ between these groups. As a result, we find that a hypothetical investor would earn higher returns from Black tenants, who pay higher rents relative to property values compared to their White counterparts (Panel A).

5 Long-term Property Decay

When rents are unresponsive to housing quality, and high rents can be generated from properties with low market values, owners may lack incentives to invest in maintenance or renovation.[22] In this section, we turn to the long-run consequences of racial disparities in landlords' incentives. We first use the 1960 and 1980 Census housing statistics to map neighborhood-level decay over time (subsection 5.1). We then turn to the property level, and study how the timing of property alterations by landlords varied as a function of tenant race providing direct evidence on patterns of disinvestment (subsection 5.2). We complement this analysis by providing visual evidence of heterogeneity in decay based on our 1940 and 1980 images (subsection 5.3).

5.1 Mapping Neighborhood-Level Decay

We start this part of our analysis by creating tract-level maps plotting housing condition data from the 1940, 1960, and 1908 Census to visually illustrate structural decline over time. For 1940, we digitize the the original map of the tracts included in the *Housing Bulletin for New York* supplement used before. By digitizing this map in ArcGIS, we can reconstruct the tract boundaries and directly attach the reported tract-level statistics reported to their corresponding

[22] In Appendix B., we construct a real estate price index to investigate capital returns to Manhattan property investors over the 1912–1939 time frame using an adjacent-period hedonic model that allows for time-series variation in characteristics' shadow prices. We find that capital returns to housing were low, consistent with estimates in Jordà et al. (2019) (henceforth JKKST). Whereas JKKST estimate an annualized gross capital return of 0.41% for this period, we find a similar return of 0.34%. Moreover, our index closely tracks the JKKST index over time. We also estimate indexes by neighborhood and building type. We do not find that tenements (low-income housing) or neighborhoods with large populations of Black residents had below-average price changes over our sample period. This is important because, while under-maintenance of rental properties can also be a rational response to deteriorating market conditions (Smith, 1979; Travis, 2019), it does not seem that owners of properties housing Black households suffered capital losses or experienced decreasing demand over our time period.

polygons.[23] For 1960 and 1980, the original census tract boundaries can be used directly. We do not have exactly the same variables across census years, but can work with proxy measures. Figure 12 shows the share of units in 1940 in need of major repairs by census tract; the share of dilapidated units in 1960; and the share of vacant units boarded up in 1980. The figure strongly suggests lower maintenance expenditures by landlords in predominantly Black neighborhoods.

5.2 Property-Level Variation in Alterations

We examine whether properties that housed Black tenants in 1940 were less likely to be maintained by their owners in the decades that followed. Our dataset includes information on construction year from the PLUTO file (see subsection 2.6), which we augment with administrative data on building alterations that we collected independently. To focus on properties unaffected by postwar urban renewal efforts, we restrict the sample to buildings constructed in 1940 or earlier.

Table 8 reports descriptive statistics and difference in means tests for properties we classify by the race of the residents. We identify household race from the 1940 Census by the race of the household head. Properties are not significantly different by the year in which they were built—around 1903. However, properties with 100% Black residents are significantly more likely to never be altered (i.e., there is no alteration permit in our data). Specifically, 62% of properties with 100% Black residents were never altered compared to 43% of properties with less than 100% Black residents. Conditional on being altered, these properties have fewer alteration permits in total and they experience 4.5 year relative delay in the time to the first alteration.

Table 9 examines the timing of these property alterations by racial composition further using both OLS and Cox proportional hazards models. The OLS results in column 1 indicate that properties occupied entirely by Black residents were altered 3.6 years later compared to properties with less than 100% Black residents, controlling for the original year of construction as well as property characteristics, latitude and longitude, and a full set of dummy variables for redlining districts. Column 2 examines the share of Black residents, showing that the delay in alterations increases proportionally with the proportion of Black residents, with properties having a higher share experiencing longer delays. In the Cox specifications, properties entirely occupied by Black residents have a hazard ratio of about 0.5, meaning they are almost 50% less likely to be altered at any given time compared to properties not entirely occupied by Black residents (column 3). This effect is also proportional to the share of Black residents (column 4). Controlling for redlining is crucial in these specifications as it indicates that the observed delays in property alterations are not being confounded by the historical disinvestment associated with redlined areas. This finding aligns with evidence that the financial effects of HOLC maps were likely more limited (Fishback et al., 2024).

[23] This original map is shown as Panel A of Appendix Figure A12; the digitized version is shown as Panel B.

5.3 Visual Property-Level Evidence of Decay

We can also test whether disinvestment disproportionately affected properties that housed Black residents over time. In addition to 1940 property images, we also have images from 1980 (subsection 2.3). These images allow us to examine changes in the physical condition of properties over time and test for greater exterior decay in properties occupied by Black residents in 1940, using two complementary approaches.

First, we predict 1980 property values based on the 1980 images as follows. We take the coefficients from a regression of 1940 property values on the principal components of the image vectors (subsection 4.1), and apply them to the corresponding 1980 principal components extracted from the feature vectors of properties photographed around 1980. This assumes that the value associated with visual features remains relatively stable over time. We can then look at the distribution of changes in estimated property value ranks between 1940 and 1980, both for properties with 100% Black residents in 1940 and other ones. Figure 13 shows that properties with entirely Black residents in 1940 rank lower on average in the 1980 value distribution. A Mann-Whitney U test rejects the null hypothesis of equal distributions.

Second, we compute cosine similarity scores between the principal components of the ResNet-extracted feature vectors for each property's 1940 and 1980 images. Our working assumption is that a lower cosine similarity indicates larger changes in exterior appearance, which is likely to be a sign of property decay. Figure 14 provides two examples from Central Harlem where properties with low cosine similarity exhibit visible signs of degradation, such as boarded-up windows and structural damage.[24] Table 10 shows that properties with 100% Black residents have significantly lower cosine similarity scores, revealing greater changes in visual characteristics, with the effect scaling proportionally with the share of Black residents. These findings underscore the disparities in how properties occupied by Black residents evolved visually between 1940 and 1980.

6 Conclusion

Real estate markets have long played a pivotal role in shaping economic opportunity and social outcomes, functioning both as a driver of wealth creation and a mechanism for reinforcing inequality. Manhattan's real estate market in the middle of the twentieth century offers an important window into how racial segregation can shape housing conditions and longer-run neighborhood change. While most prior research has focused on housing costs borne by households, our approach shifts the perspective to examine landlord economic incentives and financial returns, and their influence on the valuation and maintenance of Black-occupied properties.

By leveraging a rich dataset of property transactions, federal Census records, and property images, we provide evidence of racial disparities in the responsiveness of rents and vacancies

[24] More generally, Appendix Figure A13 highlights the ten properties most similar in 1940 and 1980, while Appendix Figure A14 showcases the ten least similar.

to housing quality. We show that properties occupied by Black residents generated higher rental yields and were less likely to be maintained by landlords over time. Using modern machine learning methods to assess visual indicators of exterior decay, we find that properties in Black neighborhoods experienced greater visual degradation in the second half of the twentieth century. These findings highlight how segregation perpetuated inequality in urban housing markets, with landlord incentives and behavior playing a central role.

References

Adukia, Anjali, Alex Eble, Emileigh Harrison, Hakizumwami Birali Runesha, and Teodora Szasz. 2023. "What We Teach About Race and Gender: Representation in Images and Text of Children's Books." *The Quarterly Journal of Economics*, 138(4): 2225–2285.

Akbar, Prottoy A., Sijie Li Hickly, Allison Shertzer, and Randall P. Walsh. 2022. "Racial Segregation in Housing Markets and the Erosion of Black Wealth." *Review of Economics and Statistics*, 1–45.

Ambrose, Brent W., Xun Bian, Ruoyu Chen, and Hanchen Jiang. 2025. "Rent Control, Rent Overcharge, and Racial Disparity." SSRN.

Aubry, Mathieu, Roman Kräussl, Gustavo Manso, and Christophe Spaenjers. 2023. "Biased Auctioneers." *The Journal of Finance*, 78(2): 795–833.

Autor, David H., Christopher J. Palmer, and Parag A. Pathak. 2014. "Housing Market Spillovers: Evidence from the End of Rent Control in Cambridge, Massachusetts." *Journal of Political Economy*, 122(3): 661–717.

Barr, Jason, and Troy Tassier. 2016. "The Dynamics of Subcenter Formation: Midtown Manhattan, 1861–1906." *Journal of Regional Science*, 56(5): 754–791.

Bayer, Patrick, Marcus Casey, Fernando Ferreira, and Robert McMillan. 2017. "Racial and Ethnic Price Differentials in the Housing Market." *Journal of Urban Economics*, 102: 91–105.

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

Box-Couillard, Sebastien, and Peter Christensen. 2024. "Racial Housing Price Differentials and Neighborhood Segregation." National Bureau of Economic Research w32815, Cambridge, MA.

Breiman, Leo. 2001. "Random Forests." *Machine Learning*, 45(1): 5–32.

Bryan, Gharad, Edward Glaeser, and Nick Tsivanidis. 2020. "Cities in the Developing World." *Annual Review of Economics*, 12(1): 273–297.

Chambers, David, Christophe Spaenjers, and Eva Steiner. 2021. "The Rate of Return on Real Estate: Long-Run Micro-Level Evidence." *Review of Financial Studies*, 34(8): 3572–3607.

Christensen, Peter, Ignacio Sarmiento-Barbieri, and Christopher Timmins. 2021. "Racial Discrimination and Housing Outcomes in the United States Rental Market." National Bureau of Economic Research w29516, Cambridge, MA.

Collins, William J., and Robert A. Margo. 2001. "Race and Home Ownership: A Century-Long View." *Explorations in Economic History*, 38: 68–92.

Cullen, Julie Berry, and Steven D. Levitt. 1999. "Crime, Urban Flight, and the Consequences for Cities." *Review of Economics and Statistics*, 81(2): 159–169.

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

Damen, Sven, Matthijs Korevaar, and Stijn Van Nieuwerburgh. 2025. "An Alpha in Affordable Housing?"

Dell, Melissa. 2025. "Deep Learning for Economists." *Journal of Economic Literature*, 107(1).

Demers, Andrew, and Andrea L. Eisfeldt. 2022. "Total returns to single-family rentals." *Real Estate Economics*, 50(1): 7–32.

Derenoncourt, Ellora. 2022. "Can You Move to Opportunity? Evidence from the Great Migration." *American Economic Review*, 112(2): 369–408.

Desmond, Matthew, and Nathan Wilmers. 2019. "Do the Poor Pay More for Housing? Exploitation, Profit, and Risk in Rental Markets." *American Journal of Sociology*, 124(4): 1090–1124.

Diamond, Rebecca, and William F. Diamond. 2024. "Racial Differences in the Total Rate of Return on Owner-Occupied Housing." National Bureau of Economic Research Working Paper 32916.

Eichholtz, Piet, Matthijs Korevaar, Thies Lindenthal, and Ronan Tallec. 2021. "The Total Return and Risk to Residential Real Estate." *Review of Financial Studies*, 34(8): 3608–3646.

Esperdy, Gabrielle. 2004. "A Taxing Photograph: The WPA Real Property Survey of New York City." *History of Photography*, 28(2): 123–136.

Fetter, Daniel K. 2016. "The Home Front: Rent Control and the Rapid Wartime Increase in Home Ownership." *The Journal of Economic History*, 76(4): 1001–1043.

Fishback, Price, Jonathan Rose, Kenneth A. Snowden, and Thomas Storrs. 2024. "New Evidence on Redlining by Federal Housing Programs in the 1930s." *Journal of Urban Economics*, 141: 103462.

Fishback, Price V., and Trevor Kollmann. 2014. "New Multicity Estimates of the Changes in Home Values, 1920-1940." In *Housing and Mortgage Markets in Historical Perspective*. NBER and University of Chicago Press.

Ghent, Andra C. 2011. "Securitization and Mortgage Renegotiation: Evidence from the Great Depression." *The Review of Financial Studies*, 24(6): 1814–1847.

Glaeser, Edward. 2011. *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*. New York:Penguin Press.

Glaeser, Edward L., Matthew E. Kahn, and Jordan Rappaport. 2008. "Why do the poor live in cities? The role of public transportation." *Journal of Urban Economics*, 63(1): 1–24.

Glaeser, Edward L., Michael Scott Kincaid, and Nikhil Naik. 2018. "Computer Vision and Real Estate: Do Looks Matter and Do Incentives Determine Looks." *National Bureau of Economic Research Working Paper Series*, No. 25174.

Glaeser, Edward L., and Joseph Gyourko. 2005. "Urban Decline and Durable Housing." *Journal of Political Economy*, 113(2): 345–375.

Gray, Rowena, and Rocco Bowman. 2021. "Locating the Manhattan housing market: GIS evidence for 1880-1910." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 54(3): 151–171.

Grebler, Leo. 1955. *Experience in Urban Real Estate Investment—An Interim Report Based on New York City Properties*. New York, NY:Columbia University Press.

Grebler, Leo, David M. Blank, and Louis Winnick. 1956. *Capital Formation in Residential Real Estate: Trends and Prospects*. Princeton University Press.

Gupta, Arpit, Christopher Hansman, and Pierre Mabille. 2024. "Financial Constraints and the Racial Housing Gap."

Gupta, Arpit, Stijn Van Nieuwerburgh, and Constantine Kontokosta. 2022. "Take the Q train: Value capture of public infrastructure projects." *Journal of Urban Economics*, 129: 103422.

Gyourko, Joseph, and Peter Linneman. 1990. "Rent controls and rental housing quality: A note on the effects of New York City's old controls." *Journal of Urban Economics*, 27(3): 398–409.

Günnewig-Mönert, Maximilian, and Ronan C. Lyons. 2023. "Judge for Yourself? The Impact of Controls on Rents in Interwar New York."

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. "Deep residual learning for image recognition." 770–778, IEEE. Las Vegas, NV.

Higgins, Brian E. 2023. "Racial Segmentation in the US Housing Market."

Jordà, Òscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M. Taylor. 2019. "The Rate of Return on Everything, 1870–2015." *Quarterly Journal of Economics*, 134(3): 1225–1298.

Kahn, Matthew E. 2007. "Gentrification Trends in New Transit-Oriented Communities: Evidence from 14 Cities That Expanded and Built Rail Transit Systems." *Real Estate Economics*, 35(2): 155–182.

Kermani, Amir, and Francis Wong. 2021. "Racial Disparities in Housing Returns." National Bureau of Economic Research w29306, Cambridge, MA.

Knoll, Katharina, Moritz Schularick, and Thomas Steger. 2017. "No Price Like Home: Global House Prices, 1870–2012." *American Economic Review*, 107(2): 331–353.

Koijen, Ralph S J, Neel Shah, and Stijn Van Nieuwerburgh. 2025. "The Commercial Real Estate Ecosystem."

Lee, Sun Kyoung. 2022. "When Cities Grow: Urban Planning and Segregation in the Prewar US." Working paper.

Lindenthal, Thies, and Erik B. Johnson. 2021. "Machine Learning, Architectural Styles and Property Values." *The Journal of Real Estate Finance and Economics*, 63: 593–614.

Logan, Trevon D., and John M. Parman. 2017. "The National Rise in Residential Segregation." *The Journal of Economic History*, 77(1): 127–170.

Logan, Trevon D., and John M. Parman. 2025. "Racial Residential Segregation in the United States." *Journal of Economic Literature*, 1(1): forthcoming.

Lyons, Ronan C., Allison Shertzer, Rowena Gray, and David N. Agorastos. 2024. "The Price of Housing in the United States, 1890-2006." National Bureau of Economic Research Working Paper 32593.

McGruder, Kevin. 2015. *Race and Real Estate: Conflict and Cooperation in Harlem, 1890-1920.* Columbia University Press.

Naik, Nikhil, Scott Duke Kominers, Ramesh Raskar, Edward L. Glaeser, and César A. Hidalgo. 2017. "Computer vision uncovers predictors of physical urban change." *Proceedings of the National Academy of Sciences*, 114(29): 7571–7576.

Nicholas, Tom, and Anna Scherbina. 2013. "Real Estate Prices During the Roaring Twenties and the Great Depression." *Real Estate Economics*, 41(2): 278–309.

Rosenbaum, Emily, and Samantha Friedman. 2006. *The Housing Divide: How Generations of Immigrants Fare in New York's Housing Market.* New York University Press.

Rosen, Sherwin. 1974. "Hedonic prices and implicit markets: Product differentiation in pure competition." *Journal of Political Economy*, 82(1): 34–55.

Rossi-Hansberg, Esteban, Pierre-Daniel Sarte, and Raymond Owens. 2010. "Housing Externalities." *Journal of Political Economy*, 118(3): 485–535.

Schmelzing, Paul. 2025. "Housing Returns and the Emergence of the Safe Asset, 1465-2024."

Sims, David P. 2007. "Out of control: What can we learn from the end of Massachusetts rent control?" *Journal of Urban Economics*, 61(1): 129–151.

Smith, Neil. 1979. "Toward a Theory of Gentrification A Back to the City Movement by Capital, not People." *Journal of the American Planning Association*, 45(4): 538–548.

Spence, A. Michael. 1975. "Monopoly, Quality, and Regulation." *The Bell Journal of Economics*, 6(2): 417.

Travis, Adam. 2019. "The Organization of Neglect: Limited Liability Companies and Housing Disinvestment." *American Sociological Review*, 84(1): 142–170.

Tritter, Thorin. 1998. "The Growth and Decline of Harlem's Housing." *Afro-Americans in New York Life and History*, 22(1): 50–70.

Voth, Hans-Joachim, and David Yanagizawa-Drott. 2024. "Image(s)." Centre for Economic Policy Research (CEPR) Discussion Paper DP19219.

Figure 1: Census Tract Housing Data, 1940

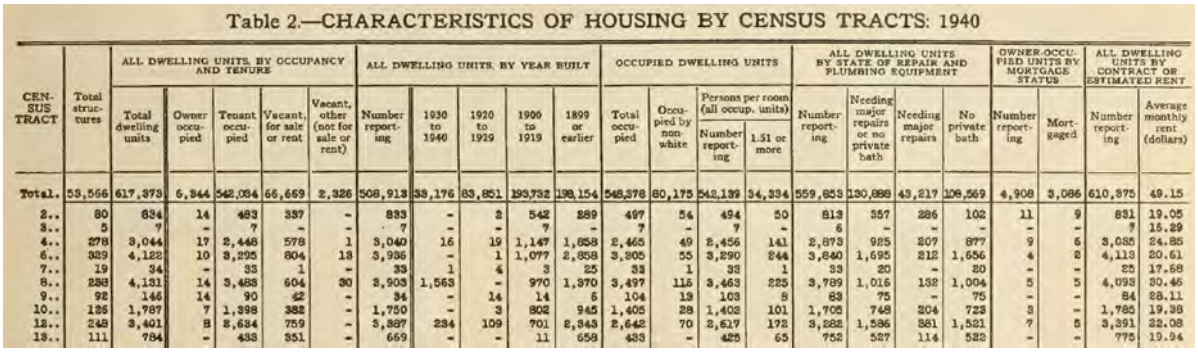


Table 2.—CHARACTERISTICS OF HOUSING BY CENSUS TRACTS: 1940

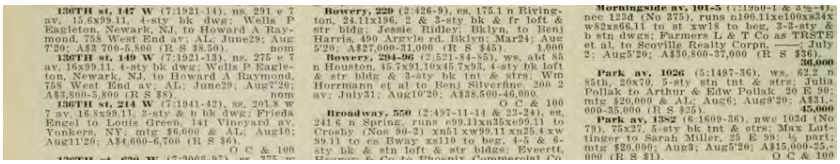
CENSUS TRACT	Total structures	ALL DWELLING UNITS, BY OCCUPANCY AND TENURE					ALL DWELLING UNITS, BY YEAR BUILT					OCCUPIED DWELLING UNITS				ALL DWELLING UNITS BY STATE OF REPAIR AND PLUMBING EQUIPMENT				OWNER-OCCUPIED UNITS BY MORTGAGE STATUS		ALL DWELLING UNITS BY CONTRACT OR ESTIMATED RENT	
		Total dwelling units	Owner occupied	Tenant occupied	Vacant, for sale or rent	Vacant, other (not for sale or rent)	Number reporting	1930 to 1940	1920 to 1929	1900 to 1919	1899 or earlier	Total occupied	Occupied by nonwhite	Persons per room (all occup. units): Number reporting	Persons per room (all occup. units): 1.51 or more	Number reporting	Needing major repairs or no private bath	Needing major repairs	No private bath	Number reporting	Mortgaged	Number reporting	Average monthly rent (dollars)
Total.	53,566	617,373	6,344	542,084	66,669	2,326	508,913	33,176	83,851	193,732	198,154	548,378	60,175	542,139	34,334	559,653	130,888	43,217	108,569	4,908	3,086	610,875	49.15
2..	80	834	14	483	337	-	833	-	2	542	289	497	54	494	50	813	357	286	102	11	9	831	19.05
3..	5	7	-	7	-	-	7	-	-	7	-	7	-	7	-	6	-	-	-	-	-	7	15.29
4..	278	3,044	17	2,448	578	1	3,040	16	19	1,147	1,858	2,465	49	2,456	141	2,873	925	207	877	9	6	3,085	24.85
6..	329	4,122	10	3,295	804	13	3,936	-	1	1,077	2,858	3,805	55	3,290	244	3,840	1,695	212	1,656	4	2	4,113	20.61
7..	19	34	-	33	1	-	33	1	4	3	25	33	1	33	1	33	20	-	20	-	-	25	17.58
8..	288	4,131	14	3,483	604	30	3,903	1,563	-	970	1,370	3,497	115	3,463	225	3,789	1,016	132	1,004	5	5	4,093	30.46
9..	92	146	14	90	42	-	34	-	14	14	6	104	13	103	8	83	75	-	75	-	-	84	28.11
10..	135	1,787	7	1,398	382	-	1,750	-	3	802	945	1,405	28	1,402	101	1,705	748	204	723	3	-	1,785	19.38
12..	248	3,401	8	2,634	759	-	3,387	234	109	701	2,343	2,642	70	2,617	172	3,282	1,586	381	1,521	7	5	3,391	22.08
13..	111	784	-	433	351	-	669	-	-	11	658	433	-	425	65	752	527	114	522	-	-	775	19.94

Notes: Data from the *Housing Bulletin for New York* that was part of the 1940 Census.

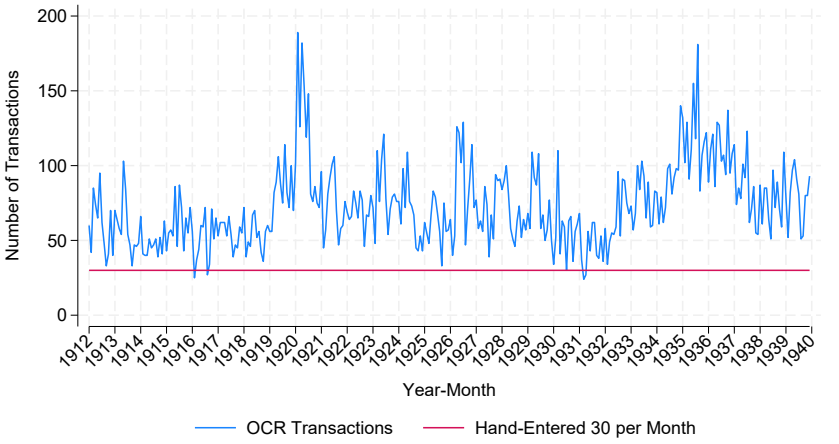
Figure 2: Real Estate Record and Builders' Guide Transactions



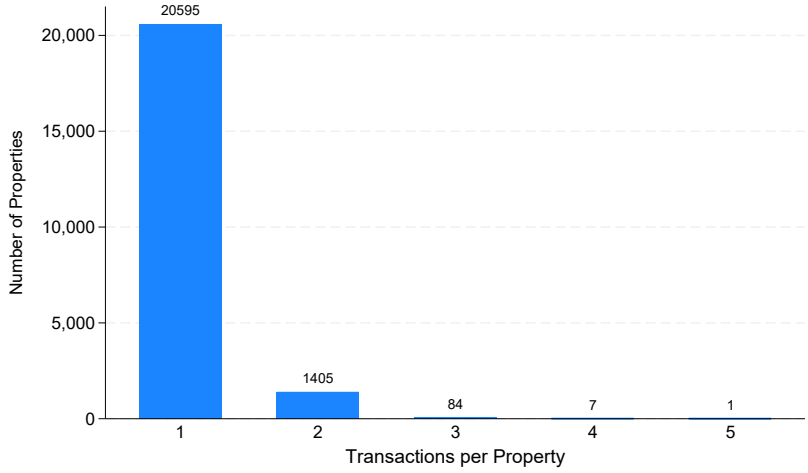
A: Layout (Pre-1912)



B: Layout (1912 Onwards)



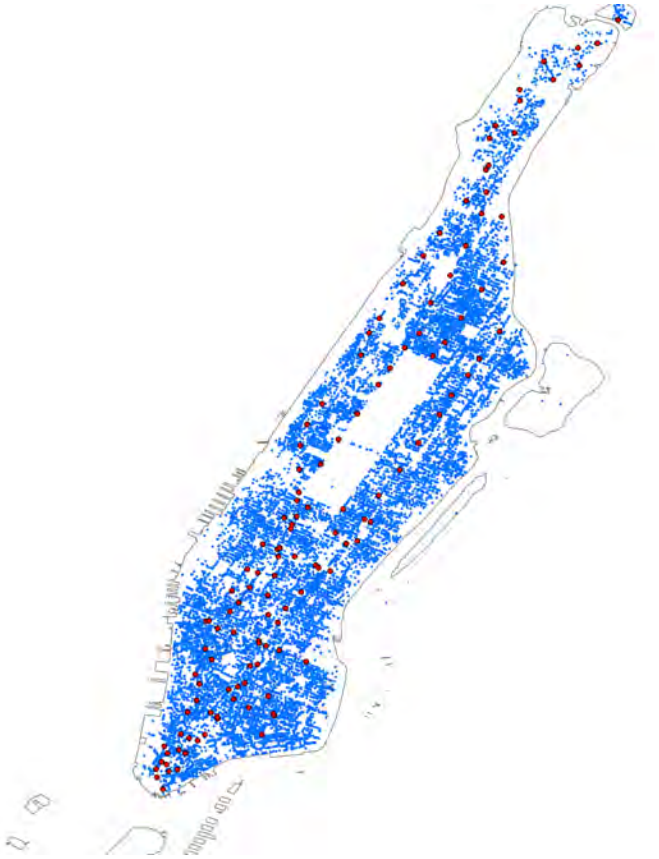
C: Number of OCR Transactions by Year-Month



D: Transaction Frequency, 1912-1939

Notes: Transactions data from 1912 to 1939 from the *Real Estate Record and Builders' Guide*.

FIGURE 3: TRANSACTIONS DATA, 1912-1939



Notes: This map shows transactions data from 1912 to 1939 from the *Real Estate Record and Builders' Guide*. Red circles are the location of subway stations that were operational by 1939. Subway coordinates are provided by the Metropolitan Transportation Authority.

Figure 4: Property Images



415 W 118th Street, 1940, 1980 and 2025

The 1940 images were developed for real property tax assessment by the Works Progress Administration in collaboration with the New York City Tax Department between 1939 and 1941 to aid property assessments. A second exercise was conducted by the Department of Finance between 1982 and 1987. 415 West 118th Street is in the Morningside Heights neighborhood of Manhattan and is currently owned by Columbia University and serves as residential housing for students.

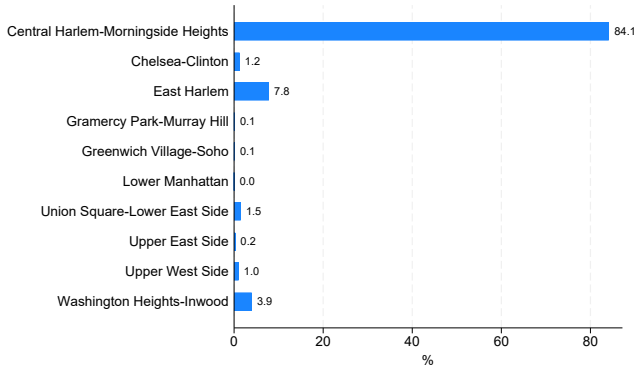


Multiple building types are included, from left to right: 239 East 60th St. (Upper East Side); 432 Lafayette St. (Noho); 366 West 46th St. (Hell's Kitchen); 313 West 88th St. (Upper West Sode).

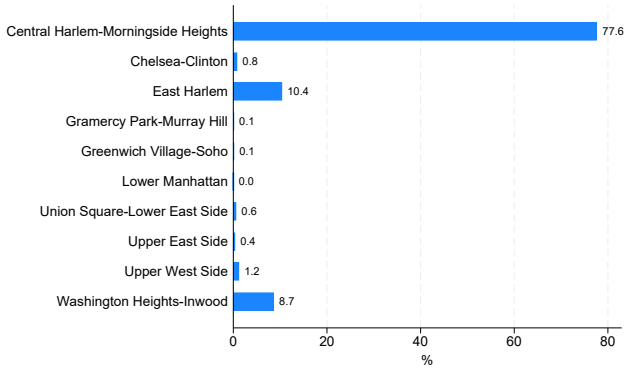


We clean the images where they contain defects (1940 images), or when a substitute card was used instead of a real image (1980 images).

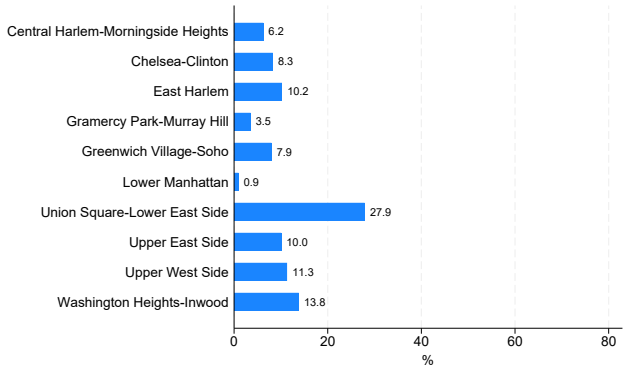
Figure 5: Location of Population by Race, 1930 and 1940



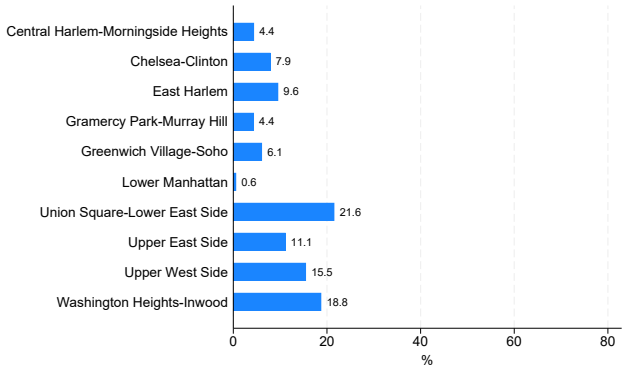
A: Black 1930



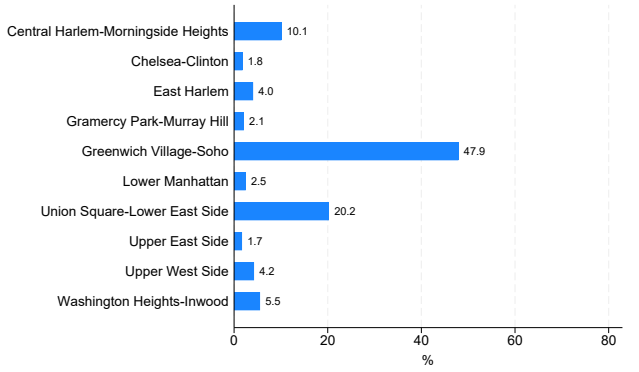
B: Black 1940



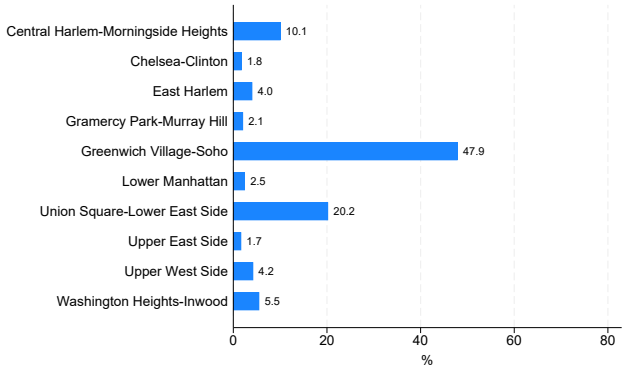
C: White 1930



D: White 1940



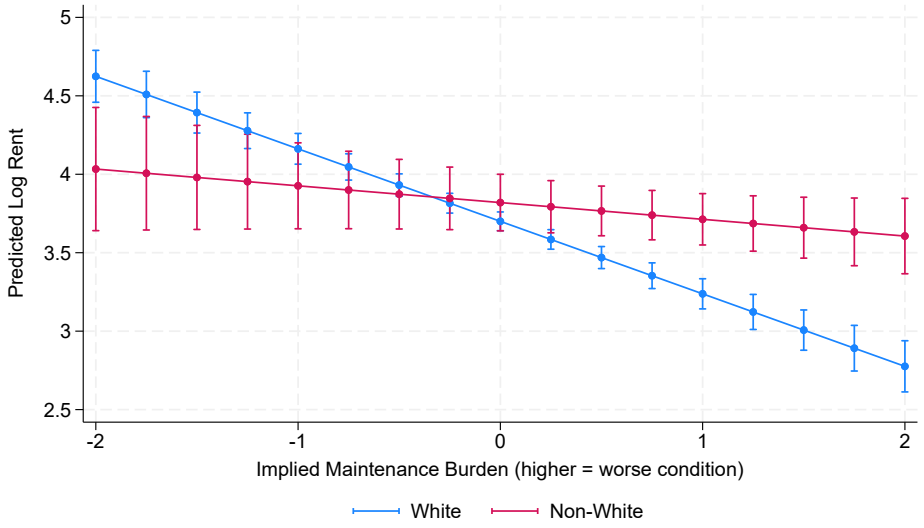
E: Other 1930



F: Other 1940

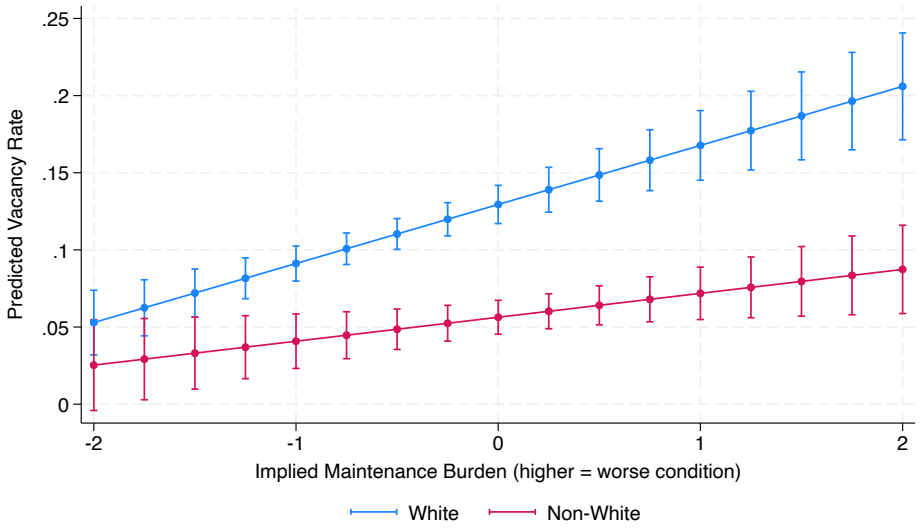
Notes: These figures show the distribution of the population of Manhattan by race based on the 1930 and 1940 Censuses.

FIGURE 6: PREDICTED RENT, MAINTENANCE BURDEN, AND RACE, 1940



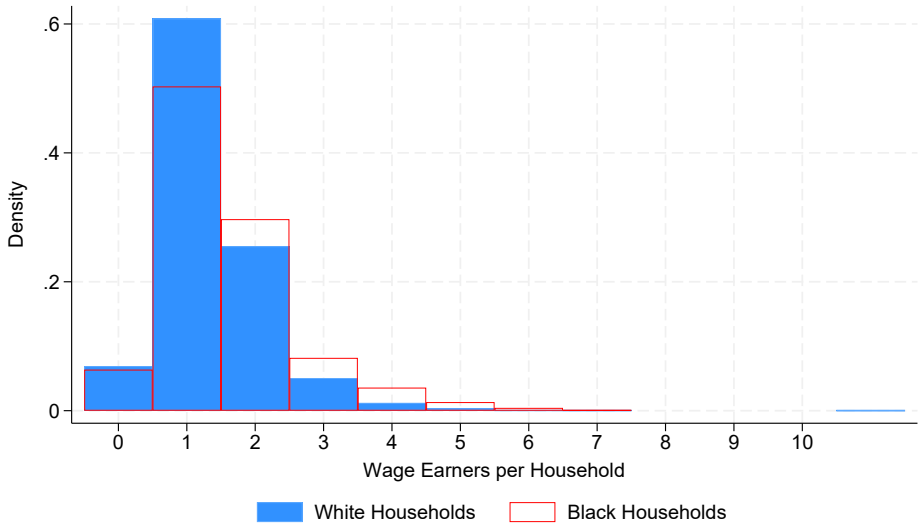
Notes: Predicted values are generated from a regression of tract-level log rent on an index of structural condition (implied maintenance burden), the tract share of non-White households, and their interaction. The condition index is constructed as the first principal component of tract-level indicators of housing deterioration. Higher values of the index indicate worse housing condition. The index is normalized to have mean zero and a standard deviation of one in each year. Predictions are shown for tracts that are all White and all non-White.

FIGURE 7: PREDICTED VACANCY RATE, MAINTENANCE BURDEN, AND RACE, 1940



Notes: Predicted values are generated from a regression of tract-level vacancy rate on an index of structural condition (implied maintenance burden), the tract share of non-White households, and their interaction. The condition index is constructed as the first principal component of tract-level indicators of housing deterioration. Higher values of the index indicate worse housing condition. The index is normalized to have mean zero and a standard deviation of one in each year. Predictions are shown for tracts that are all White and all non-White.

FIGURE 8: THE NUMBER OF WAGE EARNERS PER HOUSEHOLD



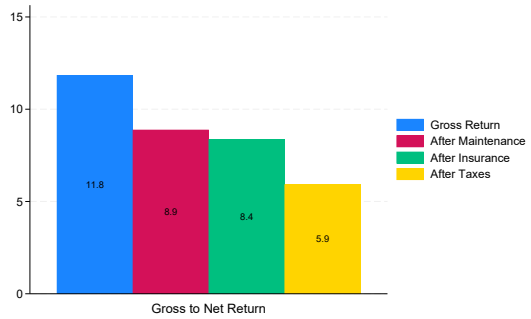
Notes: This figure shows the distribution of the number of wage earners by household based on the race of the household head.

Figure 9: Nearest Neighbors using Feature Vectors, 1940

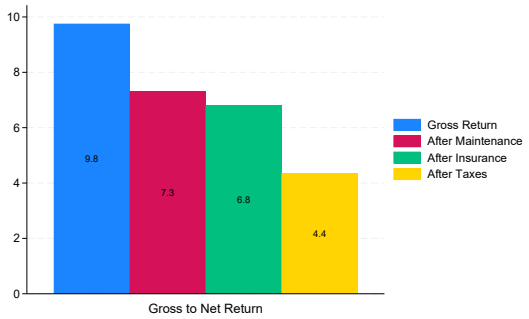


Notes: This figure shows the "nearest neighbors" for 10 randomly selected images developed for real property tax assessment by the Works Progress Administration in collaboration with the New York City Tax Department. Specifically, we calculate the pairwise Euclidean distance between the feature vector of each reference image and those of all other images to identify the five closest matches.

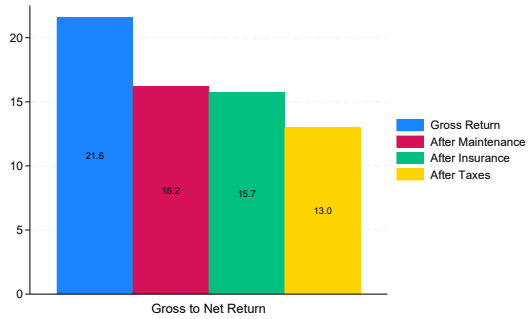
FIGURE 10: ESTIMATING GROSS AND NET RETURNS, 1930 AND 1940



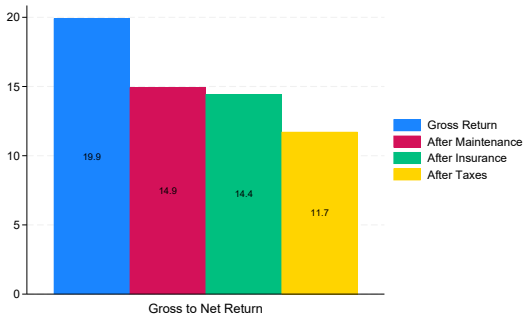
A: 1930 MEAN



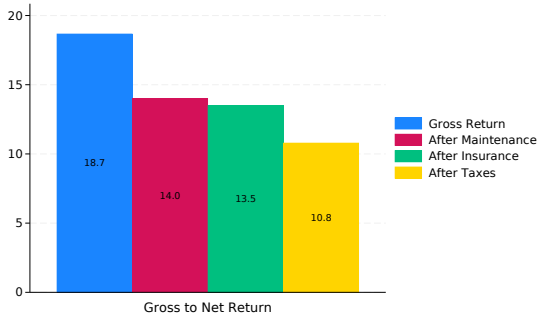
B: 1930 MEDIAN



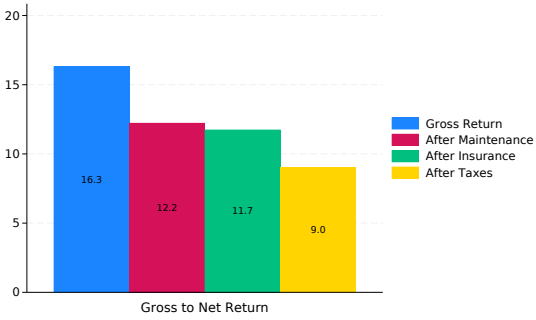
C: 1940 MEAN



D: 1940 MEDIAN



E: 1940 (IMAGES) MEAN



F: 1940 (IMAGES) MEDIAN

Notes: These figures show gross and net returns for properties. All properties are 100% rented. We use the following assumptions:

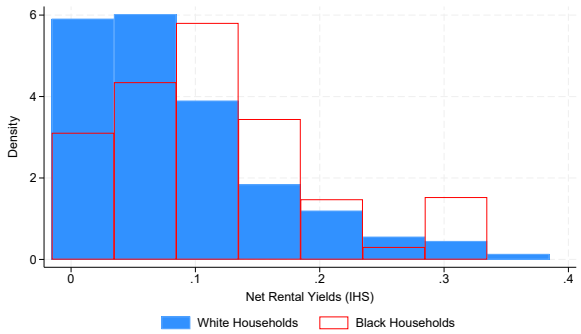
Vacancy Rate: This is already accounted for in our estimates, as vacant apartments within properties would not be enumerated.

Maintenance Rate: In both years, the maintenance cost is assumed to be 25% of annual rental income based on the *Report of the State Board of Housing of New York*. That report shows in newly built lower income properties the maintenance cost was around 24.4% to 24.8%. Grebler (1955) puts maintenance costs as high as 45%.

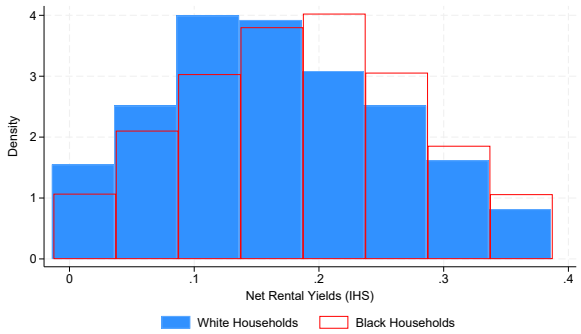
Insurance Cost: Assumed to be 0.57% of assessed value in 1930 and 0.53% of assessed value in 1940, according to the *Weekly Underwriter*.

Tax Rate: Assumed to be $27.20 per $1,000 of assessed value in 1930 and $29.60 per $1,000 of assessed value in 1940, based on the *National Municipal Review*.

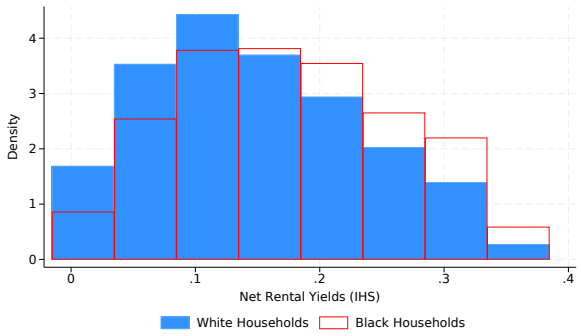
FIGURE 11: DISTRIBUTION OF NET RENTAL YIELDS BY RACE



A: 1930



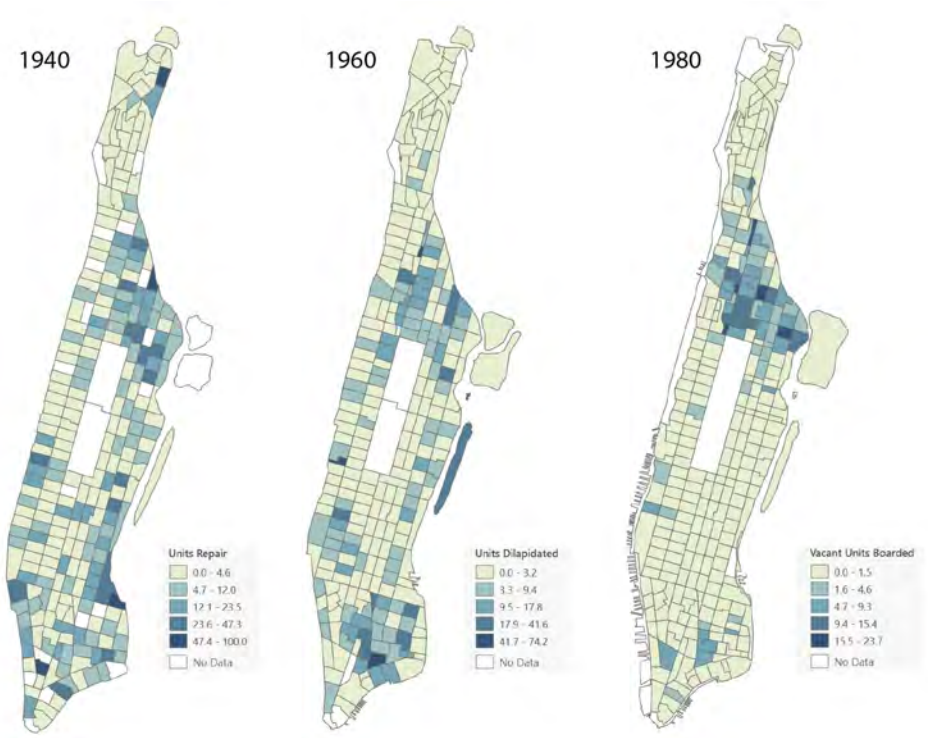
B: 1940



C: 1940 (IMAGES)

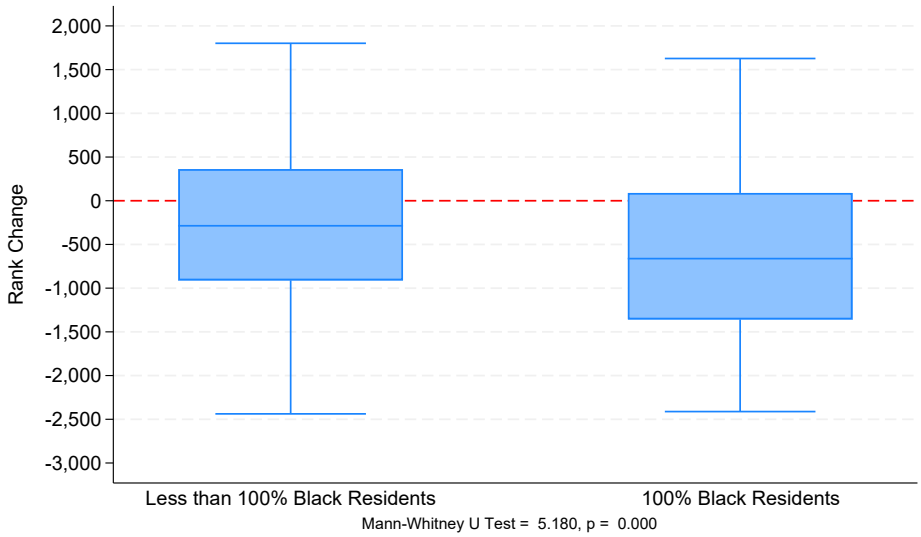
Notes: These figures show the distribution of net rental yields by the racial composition of properties for our 1930 transactions data matched to the Census (Panel A) for our 1940 transactions data matched to the Census (Panel B) and for our 1940 images data matched to the Census (Panel C).

Figure 12: Mapping Neighborhood-Level Decay



Notes: These figures show the proportion of properties at the census tract level in need of major repairs in 1940 (Panel A), dilapidated in 1960 (Panel B), and boarded up in 1980 (Panel C).

Figure 13: Rank Change in Value



Notes: This figure shows a boxplot of the distribution of the change in property value ranks between 1940 and 1980. We predict 1980 property values by applying the coefficients from a regression of 1940 property values on the principal components of the image vectors to the 1980 principal components. The Mann Whitney U test is under the null that the two groups have the same distribution.

Figure 14: Property Decay Between 1940 and 1980



162 West 122 Street, Central Harlem



210 West 112th Street, Central Harlem

Notes: This properties housed Black residents in 1940 and show low similarity scores, based on cosine similarity between feature vectors extracted from ResNet for each property's 1940 image and its corresponding 1980 image.

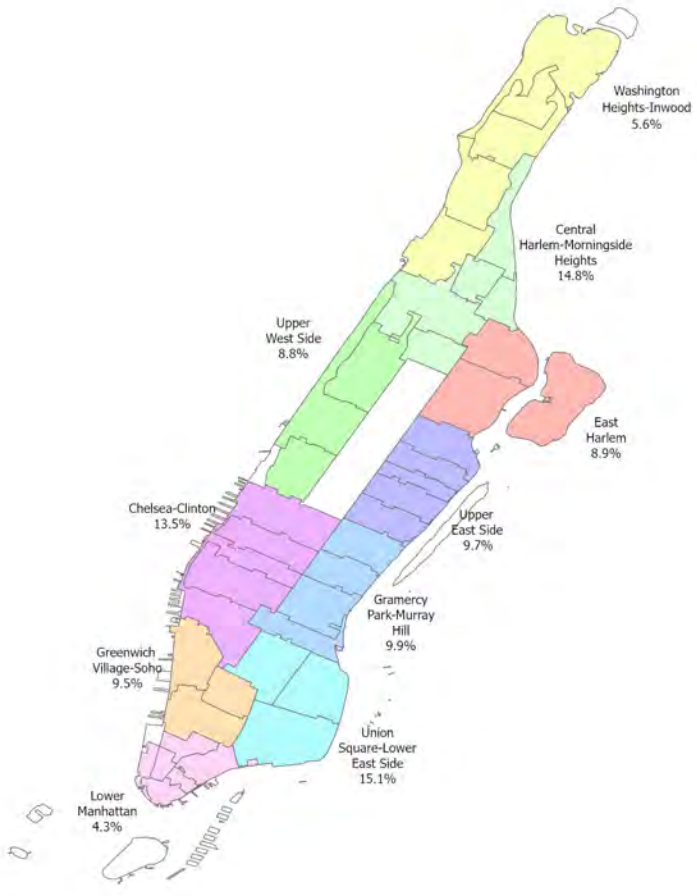
Table 1: Comparison of Housing Characteristics by Census Tract

	Predominantly White			Predom. Non-White			Diff. in Means
	N	Mean	Median	N	Mean	Median	
Vacancy rate	189	0.12	0.10	44	0.11	0.08	-0.01
Percent old (built pre-1920)	180	0.79	0.83	41	0.88	0.97	0.09*
Percent overcrowded (>1.5 pax/room)	188	0.06	0.05	44	0.08	0.06	0.02**
Percent needing major repairs	183	0.07	0.04	42	0.12	0.08	0.04*
Percent without private bath	186	0.20	0.15	41	0.21	0.13	0.01
Average monthly rent	189	52.97	42.77	44	51.18	41.11	-1.78

Notes: This table shows descriptive statistics on housing characteristics based on the census tract data taken from the *Housing Bulletin for New York* that was part of the 1940 Census

Table 2: Transactions Data Descriptive Statistics

	Mean	SD	N
Foreclosure	0.349	0.477	23690
Type: Dwelling	0.206	0.405	23690
Type: Loft	0.089	0.285	23690
Type: Other	0.229	0.420	23690
Type: Tenement	0.475	0.499	23690
Construction: Brick	0.631	0.483	23690
Construction: Stone	0.131	0.337	23690
Construction: Other	0.238	0.426	23690
Feature: None	0.611	0.488	23690
Feature: Store	0.389	0.488	23690
1 Story	0.010	0.098	23690
2 Story	0.027	0.161	23690
3 Story	0.161	0.368	23690
4 Story	0.197	0.398	23690
5 Story	0.298	0.458	23690
6 Stories and Over	0.307	0.461	23690
Square Footage (Lot, in thousands)	18.728	897.790	21825
Square Footage (Lot and Building, in thousands)	81.958	2715.151	19771
Lot Shape	2.458	1.089	21878
Property Transaction Value	40300.970	80769.752	23690
Distance to Nearest Subway (miles)	0.264	0.196	23690



Notes: This table and map show descriptive statistics on properties transacted in the *Real Estate Record and Builders' Guide* between 1912 and 1939 and the share of these transactions by neighborhood.

Table 3: Property Composition by Race of Household Head

Property Type	Number of Properties	Share of Properties	
		Total (%)	With Earning Lodger (%)
White-only Property	4,442	76.97	9.70
Black-only Property	1,012	17.54	36.26
Mixed-race Property	317	5.49	16.72
Total	5,771	100.00	14.75

Notes: Properties are classified based on the race of household heads residing in the property. Earning lodgers are lodgers, boarders, or roomers who report positive wage income.

Table 4: Incremental Explanatory Power of Rent Regressions (ΔR^2)

Added Controls	White Properties	Black Properties	Mixed Properties
Property Characteristics	0.084	0.090	0.064
Family Earners	0.001	0.001	0.033
Earning Lodgers	0.026	0.150	0.040
Household Wage	0.020	0.028	0.030

Notes: Each entry reports the increase in unadjusted R^2 from adding the indicated block of controls to a baseline regression of log rent that includes neighborhood fixed effects. Regressions are estimated separately for White only properties, Black only properties, and properties that housed both Black and White households. Property characteristics include building type, construction type, features, floor area (square footage), and style. Family earners are household members with positive wage income excluding lodgers. Earning lodgers are lodgers, boarders, or roomers who report positive wage income. Household wage refers to the log of the sum of wages at the household level and is included as an additional (non-cumulative) specification.

Table 5: Returns: Random Forest Predictions, 1940

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Net Rental Yields							
Black Household Head	0.018***	0.016***	0.017***	0.017***	0.017***	0.016***	0.009**
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Mean Dep Var.	0.17	0.17	0.17	0.17	0.17	0.17	0.17
R-sq (adj.)	0.01	0.01	0.01	0.01	0.01	0.01	0.10
Panel B: Net Rent							
Black Household Head	0.064	0.053	0.098	0.137*	0.162**	0.148*	0.113
	(0.078)	(0.078)	(0.077)	(0.077)	(0.076)	(0.083)	(0.071)
Mean Dep Var.	8.35	8.35	8.35	8.35	8.35	8.35	8.35
R-sq (adj.)	0.00	0.00	0.01	0.01	0.02	0.02	0.12
Panel C: Property Value							
Black Household Head	-0.177***	-0.182***	-0.139***	-0.122**	-0.099**	-0.134***	-0.089***
	(0.054)	(0.053)	(0.052)	(0.049)	(0.049)	(0.050)	(0.021)
Mean Dep Var.	10.54	10.54	10.54	10.54	10.54	10.54	10.54
R-sq (adj.)	0.01	0.02	0.04	0.10	0.11	0.12	0.77
Observations	14622	14622	14622	14622	14622	14622	14622
Properties	3312	3312	3312	3312	3312	3312	3312
Baseline Individual Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Wage Controls	No	No	Yes	No	Yes	Yes	Yes
Education Controls	No	No	No	Yes	Yes	Yes	Yes
Distance to Subway	No	No	No	No	No	Yes	Yes
Property Controls	No	No	No	No	No	No	Yes

Notes: This table reports coefficients from regressions where the dependent variable is the inverse hyperbolic sine transformed net rental yield. Standard errors clustered by property in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table 6: Returns: Random Forest Predictions, 1930

	(1)	(2)	(3)	(4)	(5)
Panel A: Net Rental Yields					
Black Household Head	0.034***	0.030***	0.029***	0.028***	0.025***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Mean Dep Var.	0.09	0.09	0.09	0.09	0.09
R-sq (adj.)	0.02	0.03	0.04	0.05	0.11
Panel B: Net Rent					
Black Household Head	0.896***	0.847***	0.788***	0.777***	0.497***
	(0.184)	(0.181)	(0.184)	(0.184)	(0.175)
Mean Dep Var.	7.49	7.49	7.49	7.49	7.49
R-sq (adj.)	0.01	0.01	0.01	0.02	0.10
Panel C: Property Value					
Black Household Head	0.023	0.002	-0.032	-0.035	-0.061***
	(0.059)	(0.058)	(0.058)	(0.059)	(0.021)
Mean Dep Var.	11.16	11.16	11.16	11.16	11.16
R-sq (adj.)	0.00	0.01	0.04	0.05	0.74
Observations	9404	9404	9404	9404	9404
Properties	2705	2705	2705	2705	2705
Baseline Individual Controls	No	Yes	Yes	Yes	Yes
Literacy Control	No	No	Yes	Yes	Yes
Distance to Subway	No	No	No	Yes	Yes
Property Controls	No	No	No	No	Yes

Notes: This table reports coefficients from regressions where the dependent variable is the inverse hyperbolic sine transformed net rental yield, net rent, or property value. Standard errors clustered by property in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table 7: Returns: Random Forest Predictions Using Images, 1940

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Net Rental Yields									
Black Household Head	0.027***	0.024***	0.025***	0.025***	0.025***	0.024***	0.012***	0.020***	0.010**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.005)	(0.005)
Mean Dep Var.	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15
R-sq (adj.)	0.01	0.02	0.02	0.02	0.02	0.02	0.05	0.02	0.05
Panel B: Net Rent									
Black Household Head	0.109***	0.067**	0.075***	0.076***	0.082***	0.080***	0.123**	0.124**	0.087
	(0.026)	(0.027)	(0.027)	(0.027)	(0.027)	(0.028)	(0.053)	(0.059)	(0.062)
Mean Dep Var.	7.96	7.96	7.96	7.96	7.96	7.96	7.96	7.95	7.95
R-sq (adj.)	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01
Panel C: Property Value									
Black Household Head	-0.212***	-0.215***	-0.211***	-0.204***	-0.202***	-0.203***	-0.036***	-0.080***	0.004
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.011)	(0.012)	(0.011)
Mean Dep Var.	10.26	10.26	10.26	10.26	10.26	10.26	10.26	10.29	10.29
R-sq (adj.)	0.10	0.10	0.11	0.12	0.13	0.13	0.22	0.04	0.14
Observations	62876	62876	62876	62876	62876	62876	62876	55373	55373
Properties	15762	15762	15762	15762	15762	15762	15762	13470	13470
Baseline Individual Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wage Controls	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Education Controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Distance to Subway	No	No	No	No	No	Yes	Yes	Yes	Yes
Property Controls	No	No	No	No	No	No	No	No	No
Neighborhood FE	No	No	No	No	No	No	Yes	No	Yes
Neighborhood Drop	None	None	None	None	None	None	None	CHMH	CHMH

Notes: This table reports coefficients from regressions where the dependent variable is the inverse hyperbolic sine transformed net rental yield, net rent, or property value. CHMH refers to the Central Harlem-Morningside Heights neighborhood. Standard errors clustered by property in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table 8: Property Alteration Outcomes by Race

Variable	(1) Less than 100% Black Residents	(2) 100% Black Residents	(3) Difference (p-value)
Year Built	1903.45	1904.12	0.67
	(21.32)	(11.44)	(0.64)
Never Altered	0.43	0.62	0.19***
	(0.49)	(0.49)	(0.02)
Total Alterations	2.22	1.50	-0.72***
	(1.56)	(0.79)	(0.07)
First Alteration Year	1960.35	1964.87	4.51***
	(15.51)	(15.11)	(1.13)
Mean Alteration Year	1967.11	1967.53	0.42
	(12.53)	(13.89)	(1.03)

Notes: This table reports difference in means tests for alterations-related variables for properties in Manhattan. Standard deviations in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table 9: The Timing of Property Alteration by Race

	(1)	(2)	(3)	(4)
Year Built	-0.011	-0.011	0.995***	0.995***
	(0.018)	(0.018)	(0.001)	(0.001)
100% Black Residents	3.561***		0.525***	
	(1.303)		(0.047)	
Share Residents Black		4.158***		0.513***
		(1.270)		(0.045)
Mean Dep Var.	22.39	22.39	35.52	35.52
R-sq (adj.)	0.02	0.02		
Observations (Properties)	1479	1479	2820	2820
Estimation	OLS	OLS	COX	COX

Notes: This table reports OLS coefficients from regressions where the dependent variable is the time, in years, since 1940 when the first recorded alteration occurred for each property and Cox proportional hazards models, where the time to the event of a property alteration is measured in years relative to the baseline year of 1940. All specifications control for property characteristics and spatial characteristics as well as a full set of dummy variables for redlining districts. Robust standard errors in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Table 10: Cosine Similarity Between 1940 and 1980 Property Images

	(1)	(2)
Year Built	0.004***	0.004***
	(0.001)	(0.001)
100% Black Residents	-0.360***	
	(0.061)	
Share Residents Black		-0.359***
		(0.059)
Mean Dep Var.	-0.04	-0.04
R-sq (adj.)	0.03	0.03
Observations (Properties)	2120	2120

Notes: This table reports OLS coefficients from regressions where the dependent variable is the cosine similarity between feature vectors extracted from ResNet for each property image in 1940 and its corresponding image in 1980. All specifications include a full set of dummy variables for redlining districts. Robust standard errors in parentheses. *p<0.1, **p<0.05, ***p<0.01.

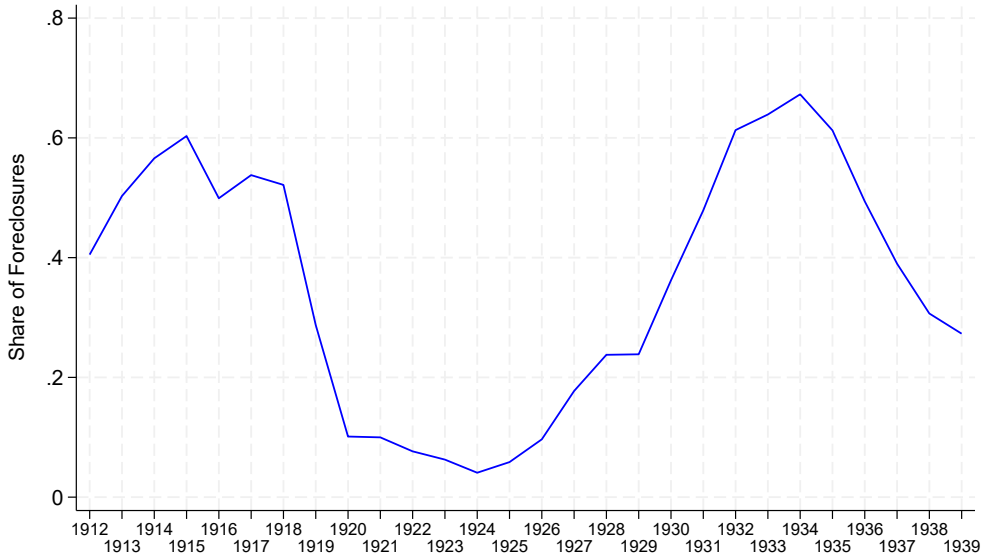
Online Appendix

Racial Segregation, Landlords, and Urban Decay

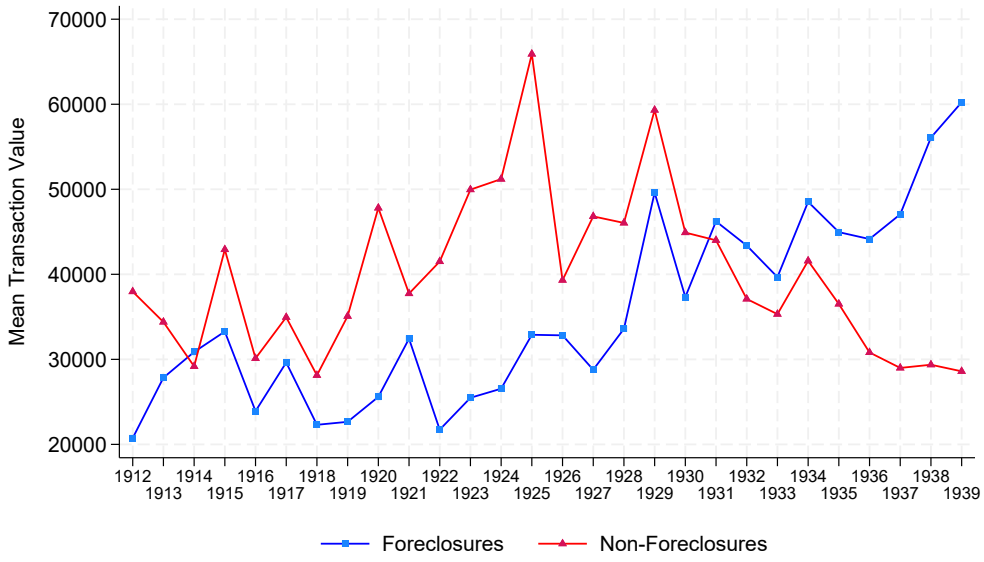
Tom Nicholas Christophe Spaenjers

Appendix A. Additional Tables and Figures

FIGURE A1: FORECLOSURES, 1912-1939



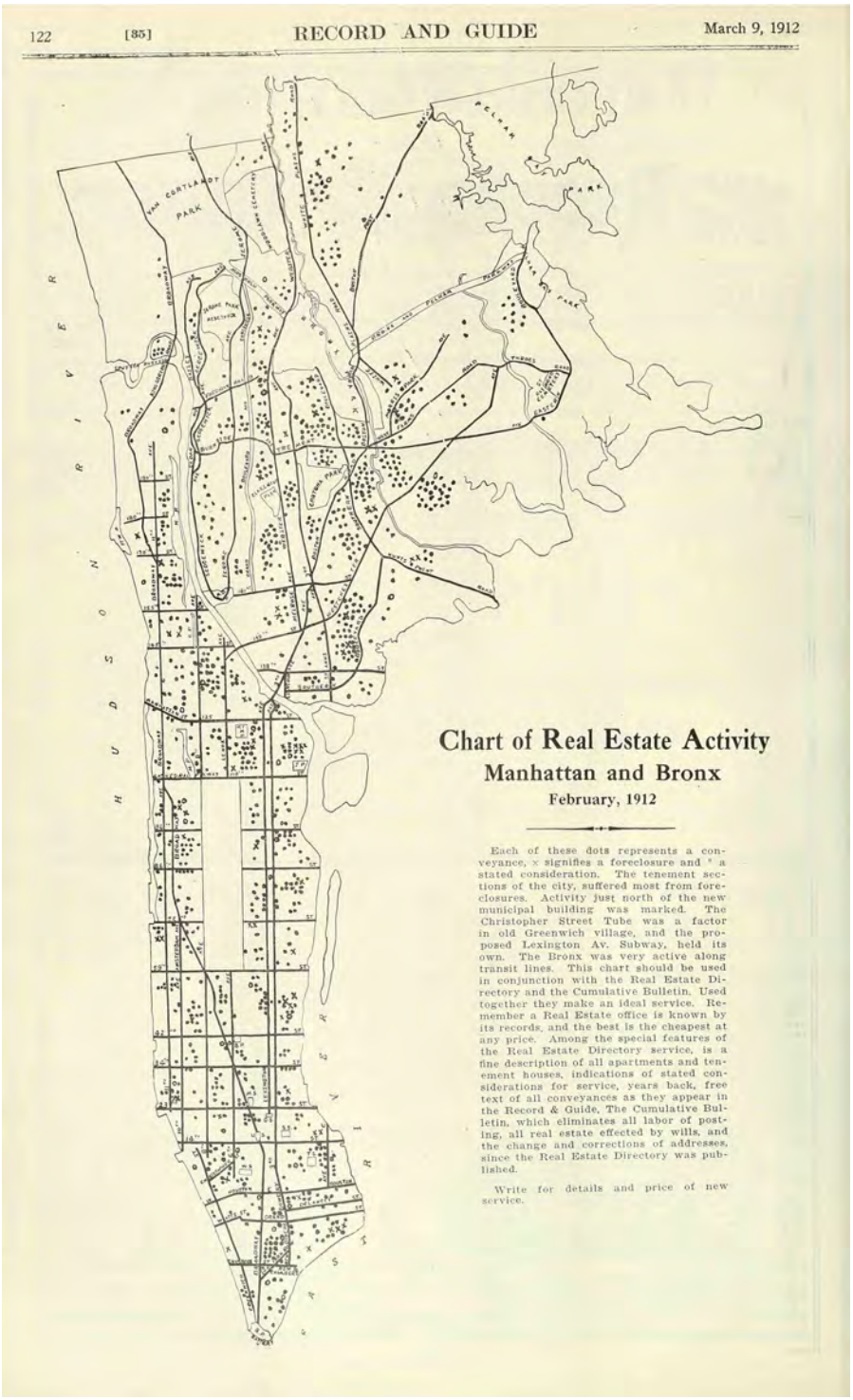
A: SHARE



B: VALUE

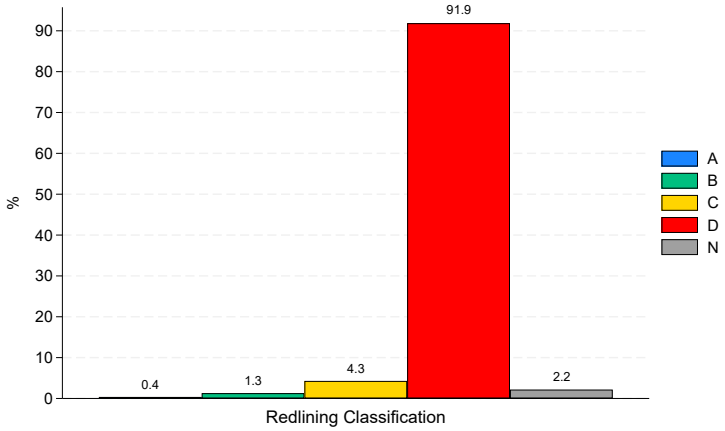
Notes: These figures show the number and average value of foreclosed property transactions between 1912 and 1939 from the *Real Estate Record and Builders' Guide*.

FIGURE A2: VERIFYING FORECLOSURE COUNTS

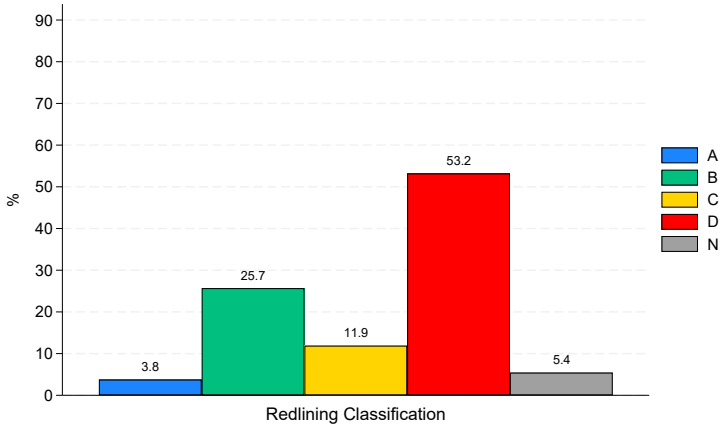


Notes: This figure show the number of foreclosed property transactions in 1912 from the *Real Estate Record and Builders' Guide*.

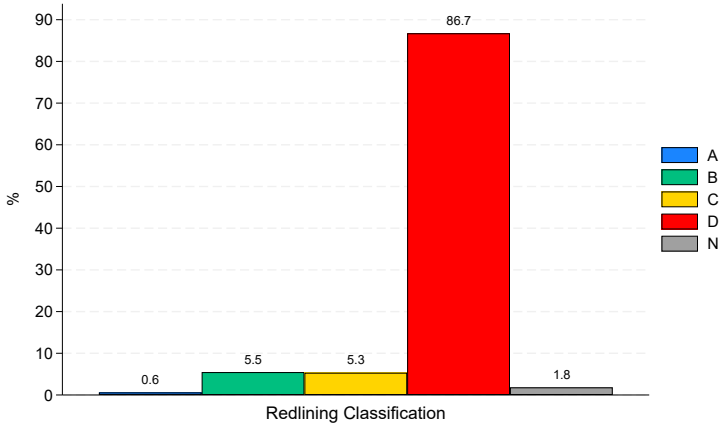
Figure A3: Location of Population by Redlining Area, 1940



A: Black



B: White



B: Other

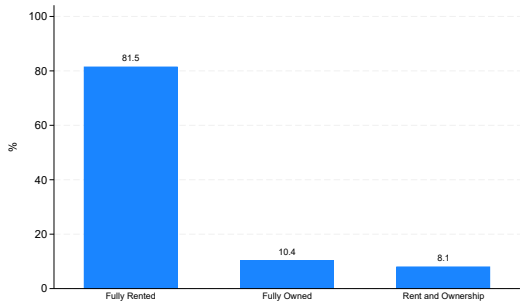
Notes: These figures show the distribution of the population of Manhattan by race based on the categories A (best), B (still desirable), C (declining), and D (hazardous) in redlining maps, originally developed by the Home Owners' Loan Corporation to assess the "risk" associated with mortgage lending in neighborhoods. The category "N" represents unclassified areas.

FIGURE A4: PROPERTIES IN THE CENSUS, 1930 AND 1940

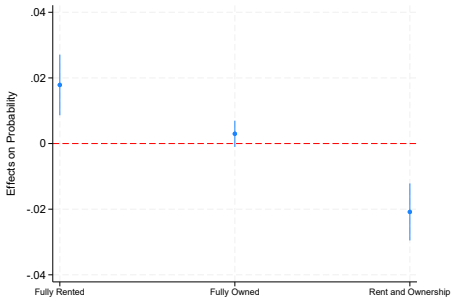


Notes: These maps shows geocoded locations of residential properties from the 1930 Census (top map) and the 1940 Census (bottom map). Shaded areas in the 1940 map are Grade D areas or "redlining" areas on HOLC maps created during the late 1930s and early 1940s.

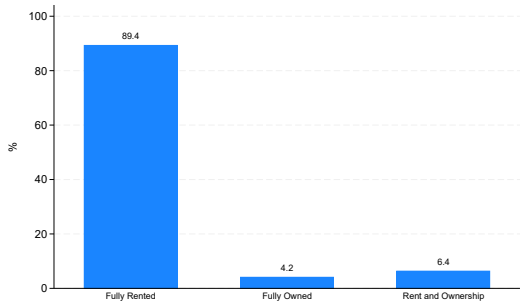
Figure A5: Property Types, 1930 and 1940



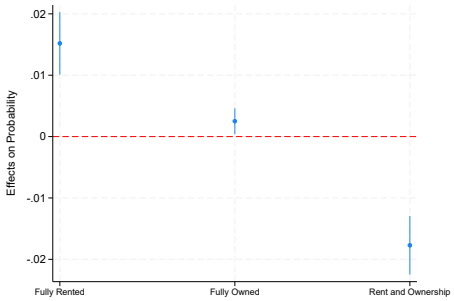
A: 1930 Property Type



B: 1930 Property Type by Race



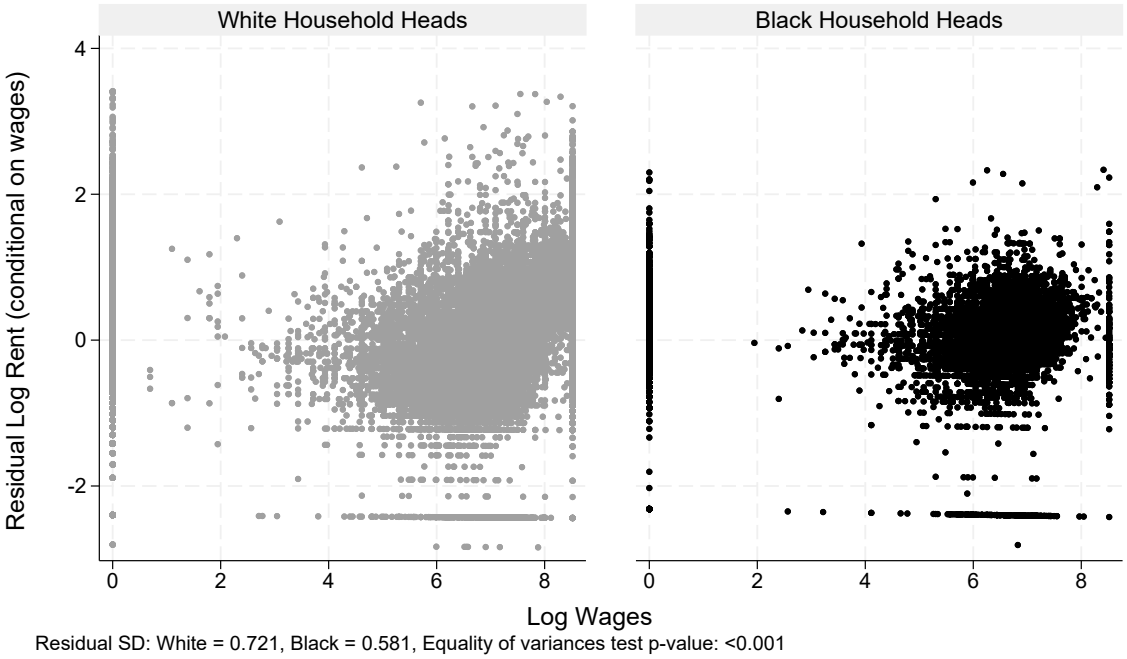
C: 1940 Property Type



D: 1940 Property Type by Race

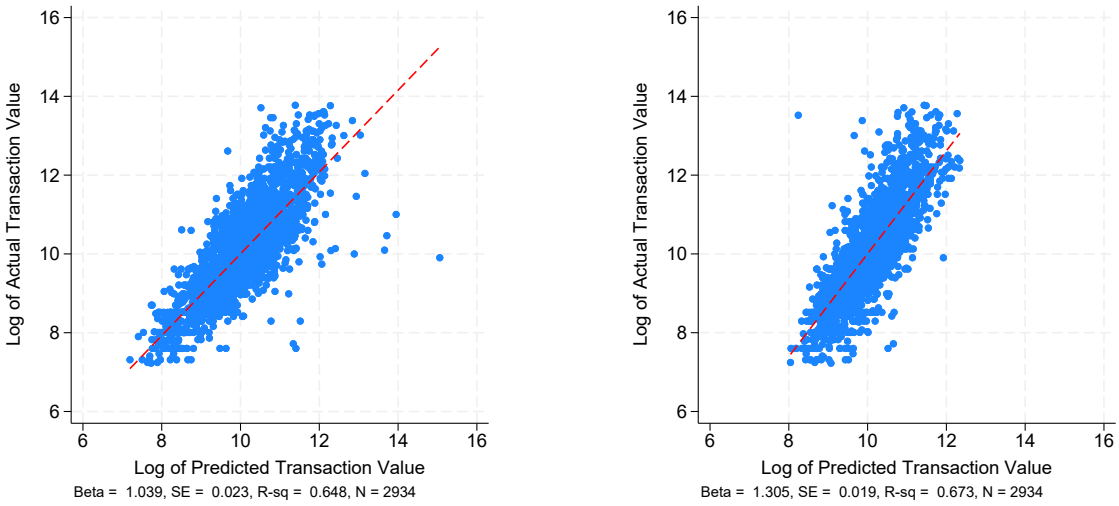
Notes: These figures show the share of properties linked from the *Real Estate Record and Builders' Guide* to the 1930 and 1940 Census that are classified as fully rented, fully owned or a mix of rent and ownership. Marginal effects of race on property type based on a multinomial logit. Marginal effects show the change in the probability of each property type associated with the household head being Black vs. non-Black. Specification controls for property characteristics, individual characteristics, and spatial characteristics.

Figure A6: Conditional Rent Dispersion by Race



Notes: This figure plots residual log rents from a pooled household level regression of log annual rent on log annual wages of the household head, a Black household head indicator, and their interaction, with residuals plotted against log wages separately by race of the household head.

Figure A7: Relation Between Predicted and Actual Property Value

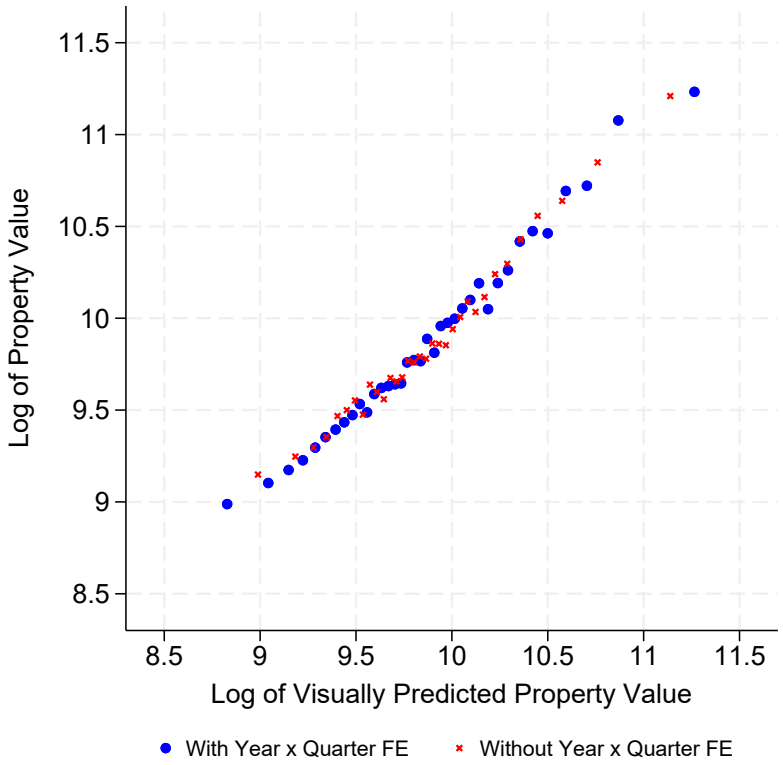


A: OLS

B: Random Forest

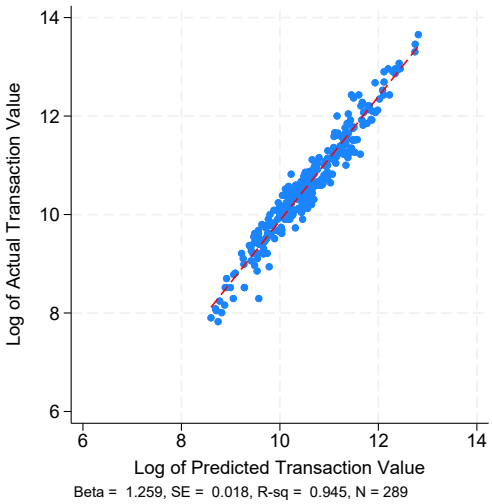
Notes: These figures show scatter plots of predicted and realized property value transactions estimated using an 80/20 training-test split for transactions occurring between 1912 and 1939. The model specification is described in the notes to Table A4.

FIGURE A8: USING VISUAL CHARACTERISTICS TO PREDICT PROPERTY VALUES

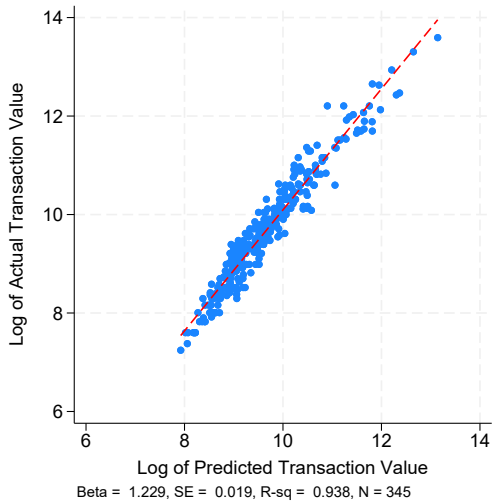


Notes: This figure shows a binscatter plot of predicted and realized property value transactions between 1912 and 1939 estimated using *only* the 100 principal components derived from visual features from 1940 photos of these properties extracted using ResNet. Plots are shown with and without year × quarter fixed effects.

FIGURE A9: PREDICTED PROPERTY VALUES IN-SAMPLE, 1929 AND 1939



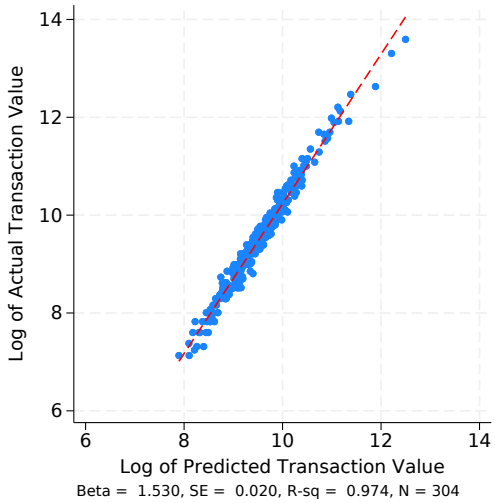
A: 1929 RANDOM FOREST



B: 1939 RANDOM FOREST

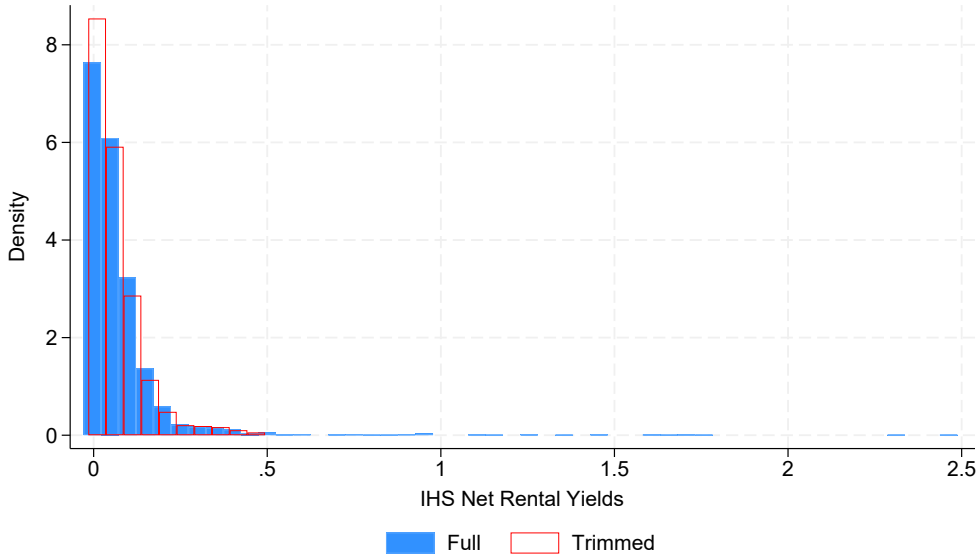
Notes: These figures show scatter plots comparing predicted and actual property transaction values, estimated using transactions that occurred in the five-year window before the target years of 1929 and 1939. Specifically, we use data from 1925–1929 and 1935–1939 to predict transaction values for the full cross-section of properties observed transacting between 1912 and 1939, as if they had occurred in 1929 and 1939, respectively. The predicted values are then plotted against the actual transaction values observed for these target years. We exclude foreclosures from the training model but include properties that experienced a foreclosure between 1912 and 1939 in the predicted set, treating them as if they transacted at market prices.

FIGURE A10: PREDICTED PROPERTY VALUES IN-SAMPLE USING IMAGES, 1939

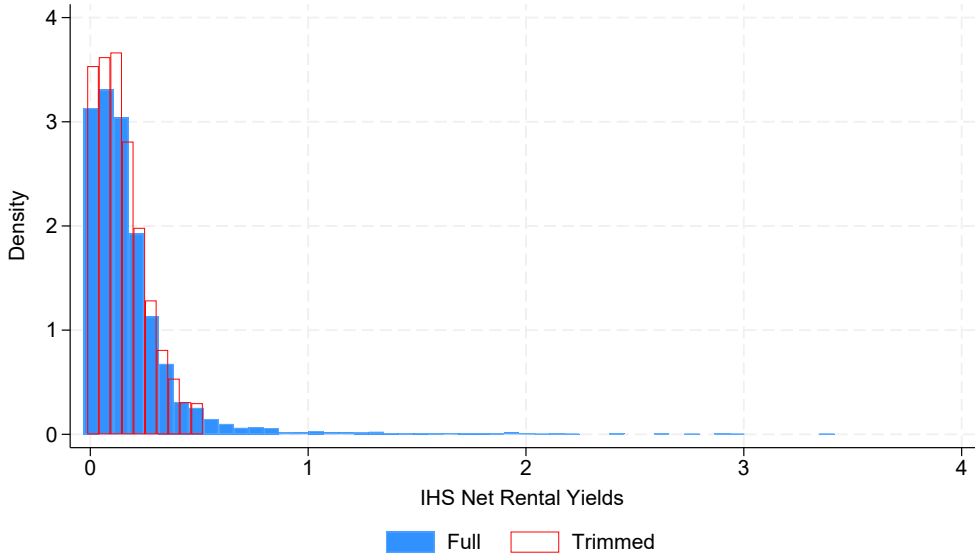


Notes: These figures show scatter plots comparing predicted and actual property transaction values, estimated using transactions that occurred in the five-year window before the target year of 1939. Specifically, we use image vector data from 1935–1939 to predict transaction values for the full cross-section of properties observed transacting between 1912 and 1939, as if they had occurred in 1939. The predicted values are then plotted against the actual transaction values observed for the target year. We exclude foreclosures from the training model but include properties that experienced a foreclosure between 1912 and 1939 in the predicted set, treating them as if they transacted at market prices.

Figure A11: Trimming Net Rental Yields



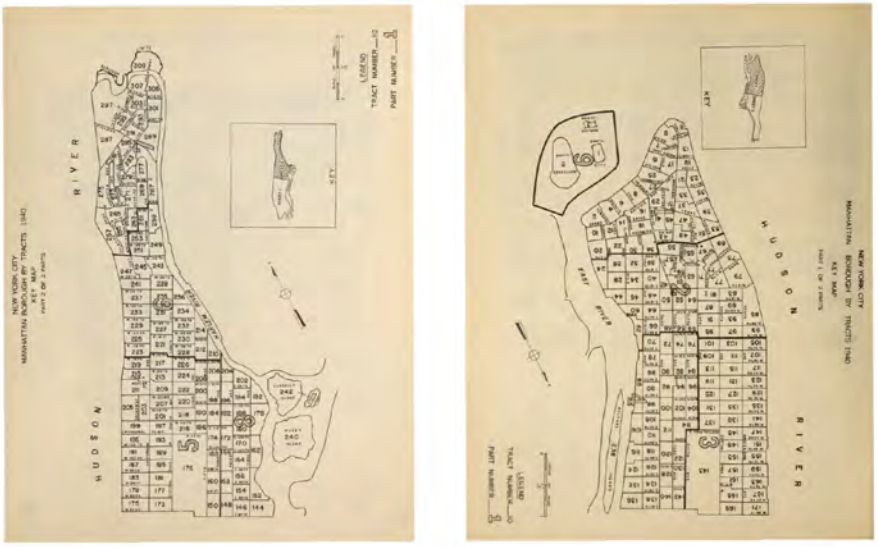
A: 1930



B: 1940

Notes: These figures show the full distribution of returns compared to the distribution trimmed at the 5th and 95th percentiles.

Figure A12: Digitizing 1940 Map of Census Tracts



A: Original Map



B: Digitized Version

Notes: These figures show the original maps in the *Housing Bulletin for New York* supplement to the 1940 Census (Panel A) and our digitized version (Panel B).

Figure A13: Highest Similarity Scores: Black Properties 1940 and 1980



Notes: This figure shows the 10 Black properties with the highest similarity scores, based on cosine similarity between feature vectors extracted from ResNet for each property's 1940 image and its corresponding 1980 image.

FIGURE A14: LOWEST SIMILARITY SCORES: BLACK PROPERTIES 1940 AND 1980



Notes: This figure shows the 10 Black properties with the lowest similarity scores, based on cosine similarity between feature vectors extracted from ResNet for each property's 1940 image and its corresponding 1980 image.

Table A1: Sample Construction and Match Rates

	Count	% of Universe
Panel A. Census Properties (Address Level)		
Census properties (distinct addresses)	71,334	100.0
Geocoded	59,265	83.1
Geocoded, score > 95	55,487	77.8
Matched to RERBG	8,303	11.6
Estimation sample (matched to RERBG, fully rented)	3,312	4.6
Matched to Images	34,485	48.3
Estimation sample (matched to Images, fully rented)	15,762	22.1
Panel B. RERBG Properties		
RERBG properties	22,412	100.0
Geocoded	22,092	98.6
Geocoded, score > 95	15,633	69.8
Matched to Census	8,303	37.0
Estimation sample (fully rented)	3,312	14.8
Panel C. Image Properties		
Image properties	71,772	100.0
Geocoded	68,216	95.1
Geocoded, score > 95	66,832	93.1
Matched to Census	34,485	48.0
Estimation sample (fully rented)	15,762	22.0

Notes: The table reports sample construction and matching rates from the perspectives of census addresses, transaction properties, and image properties. The 1940 Census enumerates households rather than properties; throughout, properties are defined as distinct residential addresses (structures). Score is an ArcGIS 0-100 match-quality score for geocoded properties. Estimation samples are (1) matched to RERBG transaction records and (2) matched to property images. Final estimation samples are restricted to fully rented properties only.

Table A2: Descriptive Statistics, 1930

	Household Head is White		Household Head is Black		Household Head is Other	
	Mean	SD	Mean	SD	Mean	SD
Panel A: Property-Level						
Type: Dwelling	0.076	0.265	0.057	0.232	0.024	0.154
Type: Loft	0.016	0.127	0.008	0.089	0.000	0.000
Type: Other	0.114	0.317	0.110	0.313	0.024	0.154
Type: Tenement	0.794	0.405	0.825	0.380	0.952	0.215
Construction: Brick	0.833	0.373	0.814	0.390	0.880	0.328
Construction: Stone	0.011	0.105	0.004	0.059	0.000	0.000
Construction: Other	0.156	0.363	0.183	0.387	0.120	0.328
Feature: None	0.458	0.498	0.601	0.490	0.277	0.450
Feature: Store	0.542	0.498	0.399	0.490	0.723	0.450
1 Story	0.006	0.076	0.004	0.066	0.000	0.000
2 Story	0.009	0.093	0.004	0.066	0.000	0.000
3 Story	0.059	0.236	0.047	0.212	0.024	0.154
4 Story	0.137	0.343	0.050	0.217	0.157	0.366
5 Story	0.511	0.500	0.655	0.476	0.663	0.476
6 Stories and Over	0.278	0.448	0.240	0.427	0.157	0.366
Square Footage (Lot, in thousands)	48.348	2296.567	18.622	230.058	3.243	1.878
Square Footage (Lot and Building, in thousands)	111.084	4593.378	94.554	1150.321	16.447	11.595
Lot Dimensions	2.431	0.999	2.209	0.783	2.964	1.005
Property Value (Random Forest)	47623.794	48257.080	43952.632	34396.281	32799.340	23198.284
Property Value (OLS)	48281.524	44733.451	46473.581	38888.068	38857.832	15961.112
Property Rent (12 Months)	8312.993	14160.545	8791.059	9578.787	5048.964	5537.125
Distance to Nearest Subway 1929 (miles)	0.282	0.141	0.269	0.114	0.226	0.077
Panel B: Individual-Level						
Age	45.900	13.854	38.734	10.257	41.614	11.380
Male	0.739	0.439	0.667	0.471	0.964	0.188
Female	0.261	0.439	0.333	0.471	0.036	0.188
Employment Status: At Work	0.726	0.446	0.815	0.389	0.843	0.366
Employment Status: Unemployed	0.110	0.313	0.114	0.319	0.096	0.297
Employment Status: Not in Labor Force	0.164	0.370	0.071	0.257	0.060	0.239
Non-Literate	0.135	0.342	0.017	0.129	0.205	0.406
Literate	0.865	0.342	0.983	0.129	0.795	0.406

Table A3: Descriptive statistics, 1940

	Household Head is White		Household Head is Black		Household Head is Other	
	Mean	SD	Mean	SD	Mean	SD
Panel A: Property-Level						
Type: Dwelling	0.129	0.336	0.188	0.391	0.063	0.245
Type: Loft	0.017	0.131	0.005	0.073	0.014	0.118
Type: Other	0.133	0.339	0.102	0.303	0.049	0.217
Type: Tenement	0.720	0.449	0.704	0.456	0.873	0.334
Construction: Brick	0.814	0.389	0.748	0.434	0.866	0.342
Construction: Stone	0.009	0.097	0.006	0.079	0.000	0.000
Construction: Other	0.177	0.382	0.246	0.431	0.134	0.342
Feature: None	0.510	0.500	0.669	0.471	0.190	0.394
Feature: Store	0.490	0.500	0.331	0.471	0.810	0.394
1 Story	0.001	0.032	0.004	0.061	0.000	0.000
2 Story	0.007	0.083	0.005	0.073	0.021	0.144
3 Story	0.071	0.257	0.159	0.366	0.070	0.257
4 Story	0.146	0.353	0.081	0.273	0.148	0.356
5 Story	0.457	0.498	0.491	0.500	0.254	0.437
6 Stories and Over	0.318	0.466	0.260	0.439	0.507	0.502
Square Footage (Lot, in thousands)	33.299	1557.145	9.157	115.752	5.797	13.368
Square Footage (Lot and Building, in thousands)	115.078	3302.523	46.574	578.085	33.036	80.612
Lot Dimensions	2.370	0.970	2.182	0.683	2.775	1.113
Property Value (Random Forest)	30933.132	43417.595	23179.540	28440.369	17700.208	15365.595
Property Value (OLS)	48111.765	847370.454	27007.855	94456.453	28997.217	43569.838
Property Rent (12 Months)	7805.456	12477.000	5835.612	5999.990	4641.887	4431.332
Distance to Nearest Subway 1939 (miles)	0.228	0.130	0.179	0.066	0.207	0.103
Panel B: Individual-Level						
Age	47.439	14.388	41.772	11.375	41.979	11.522
Male	0.714	0.452	0.586	0.493	0.972	0.166
Female	0.286	0.452	0.414	0.493	0.028	0.166
Employment Status: At Work	0.620	0.485	0.648	0.478	0.754	0.432
Employment Status: Unemployed	0.106	0.308	0.108	0.311	0.127	0.334
Employment Status: Not in Labor Force	0.274	0.446	0.243	0.429	0.120	0.326
Wage	938.741	1277.828	553.369	730.290	664.148	741.150
Years of Schooling	7.636	4.526	7.392	3.382	4.070	4.943

Table A4: Property Value Prediction: Out-of-Sample Tests

	Model Performance			
	MAE	MedAE	MPE	MedPE
OLS	27158	6478	-31.3	-6.8
Random Forest	24834	5874	-22.9	-3.6

Notes: This table presents model performance metrics for predicting property transaction values, estimated using an 80/20 training-test split for transactions occurring between 1912 and 1939. The model specification predicts the log of property value based on the ratio of value to assessed value, square footage, physical dimensions, indicators for property type, construction type, architectural style, property features, foreclosure classification, latitude and longitude and year. Pricing errors have been unlogged.

Table A5: Gross Returns, 1940

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Black	0.024***	0.021***	0.022***	0.022***	0.023***	0.022***	0.012**
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Mean Dep Var.	0.27	0.27	0.27	0.27	0.27	0.27	0.27
R-sq (adj.)	0.01	0.01	0.01	0.01	0.01	0.01	0.10
Observations	14622	14622	14622	14622	14622	14622	14622
Properties	3312	3312	3312	3312	3312	3312	3312
Baseline Individual Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Wage Controls	No	No	Yes	No	Yes	Yes	Yes
Education Controls	No	No	No	Yes	Yes	Yes	Yes
Distance to Subway	No	No	No	No	No	Yes	Yes
Property Controls	No	No	No	No	No	No	Yes

Notes: This table reports coefficients from regressions where the dependent variable is the inverse hyperbolic sine transformed gross rental yield. Standard errors clustered by property in parentheses. *p<0.1, **p<0.05, ***p<0.01.

Appendix B. Hedonic Model and Capital Return Estimates

Because properties trade infrequently, their values are typically unobserved. Market transaction prices offer the most reliable basis for valuation (needed to estimate rental yields) and the estimation of capital gains over time, but we need to account for the heterogeneous nature of real estate. In this section, we will first present a simple hedonic model that sheds light on the drivers of cross-sectional price variation in our dataset of RERBG transactions. This analysis will serve as a stepping board for our construction of price indexes.

B.1 A Simple Hedonic Model

Let $V_{i,t}$ denote the observed market price of property i sold at time t, and let $v_{i,t}$ represent its natural logarithm. Assume that each transaction includes M measurable characteristics that fully describe the property (as outlined in Table 2) and remain constant over some period of time: $\mathbf{z}_i = [z_{i,1}, z_{i,2}, \ldots, z_{i,M}]$.

Using all these characteristics, along with quarter-by-year dummies, we can estimate a standard hedonic regression using our RERBG observations:

$$v_{i,t} = \delta_{q,t} D_{q,t} + \sum_{m=1}^{M} z_{i,m}\gamma_m + \epsilon_{i,t}, \quad (1)$$

where $D_{q,t}$ is a dummy variable for quarter q of year t, and $\delta_{q,t}$ represents the corresponding coefficient for the quarter-by-year effect. These dummies explicitly control for temporal variation in property values at a quarterly level, ensuring that the coefficients on the characteristics reflect within-period differences (rather than picking up broader market trends correlated with changes in the distribution of characteristics over time). This model enables us to estimate the market value of any property at a given time based on its specific attributes and the fitted coefficients.

Table B1 reports the γ_m coefficients corresponding to the property characteristics, neighborhood fixed effects, and location characteristics in $z_{i,m}$. In column 1 we report the coefficients for our full time period 1912-1939 and then we report estimates of sub-period effects for the 1910s, 1920s, and 1930s in columns 2, 3, and 4. The coefficient on the log of square footage is positive and significant though less than unity across all periods suggesting diminishing marginal returns to property size in larger buildings where much of the additional square footage may be allocated to less desirable spaces, such as hallways, storage, or elevator shafts. The coefficients on stories capture the positive aggregate value of taller buildings (the baseline of six stories or over).

Deviations from a standard rectangular shape influence property values. A one-unit increase in the number of dimensions is associated with a 6.1% increase in property value. Irregular lots could be beneficial for commercial or mixed-use developments, for example, driving up prices relative to standard lots. In the full 1912-1939 period having a store is associated with a 14% higher property value ($e^{0.131} - 1 = 0.139$). The effect is largest in the 1930s where properties with a store command a 20.1% higher price compared to those without additional features.

Building type significantly affects property values. Compared to tenements (perhaps due to their association with poor living conditions, lack of amenities, and historical use as affordable housing), dwellings, lofts, and other building types command substantial transaction premiums, with 'other' building types showing the largest positive effects. Construction material also plays a role; compared to brick (the base category), stone construction adds a small premium, while properties made from other materials are valued less.

Neighborhood fixed effects reveal substantial variation. Affluent neighborhoods like Gramercy Park–Murray Hill and the Upper East Side consistently exhibit significant premiums compared to Union Square–Lower East Side, with the largest premiums observed in the 1930s. Predominantly Black neighborhoods, specifically Central Harlem– Morningside Heights and East Harlem, generally show lower property values, though the disparity does not seem to dramatically change over time. Distance to the nearest subway station negatively impacts property values across all periods, emphasizing the importance of transit accessibility. A 10% increase in the distance to the nearest subway station is associated with a 1.5% to 1.7% decrease in property value across periods.

Finally, the coefficient for foreclosure is negative and significant across all periods, with foreclosed properties being sold at an average discount of about 20.3% from 1912 to 1939 compared to non-foreclosed properties. The discount is largest in the 1920s where foreclosed properties sold for approximately 43.2% less than comparable non-foreclosed properties. Given the preponderance of foreclosures in our data, we include these properties in our estimation of capital returns but account for the lower associated transaction values.

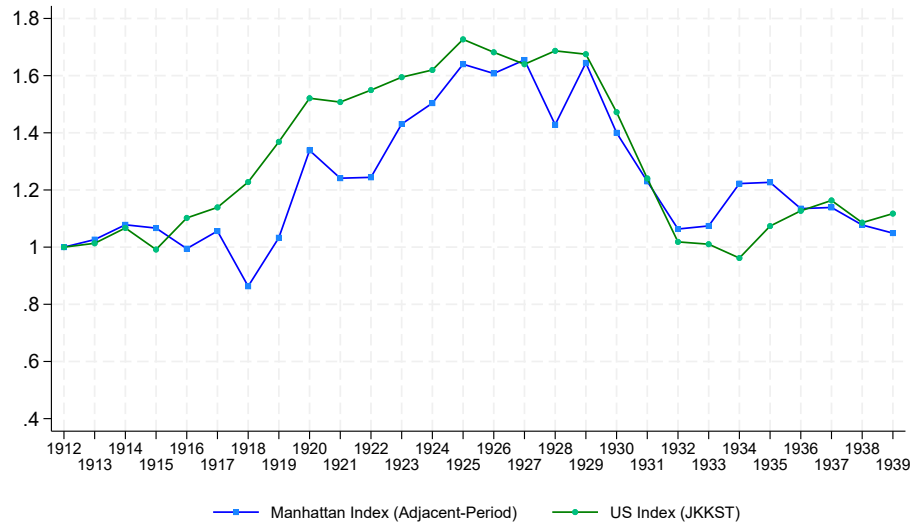
B.2 Aggregate, Property Type, and Neighborhood Returns

We now move to constructing price indexes, comparing them to estimated aggregate capital returns to housing from Jordà et al. (2019), while exploring heterogeneity by property type and neighborhood. We modify the simple hedonic model introduced in the previous subsection in two ways. First, to ensure interpretability and robustness over longer time periods, we replace quarter-by-year effects with year dummies. This allows us to estimate indexes that reflect the relative change in property values across years, measuring the annual capital return on a normalized initial investment of $1. Second, one drawback of a hedonic model estimated on pooled data (i.e., data for the full sample period or pooled over a decade) is that characteristics' shadow prices are constrained to be constant over relatively long periods of time. In reality, the relative attractiveness and thus pricing of, for example, certain neighborhoods may show faster variation. We therefore switch to using an adjacent-period method, which pools transactions from overlapping two-year time frames to estimate property values relative to adjacent periods. This approach allows for changes in the pricing of characteristics while retaining the ability to estimate annual price changes. In this adjacent-period method, we generate an index by applying the chain rule, linking cumulative relative estimated changes in property values across the overlapping two-year periods.

Figure B1 presents our key findings on capital returns, highlighting the close correspondence between our estimates and the nation-wide capital return series from Jordà et al. (2019). (The U.S. house price index in Jordà et al. (2019) is taken from Knoll, Schularick and Steger (2017), who themselves rely on multi-city single-family house price indexes from Grebler, Blank and Winnick (1956) and Fishback and Kollmann (2014) for the relevant period.) They estimate an annualized gross capital return of 0.41% between 1912 and 1939, whereas we estimate a return of 0.34% using our adjacent-period index. This correspondence is striking because of the differences in methodologies, granularity, and geographic focus of the data. Our results for Manhattan are highly consistent with their findings, implying that the capital return at the national level also holds true in a major localized urban market.

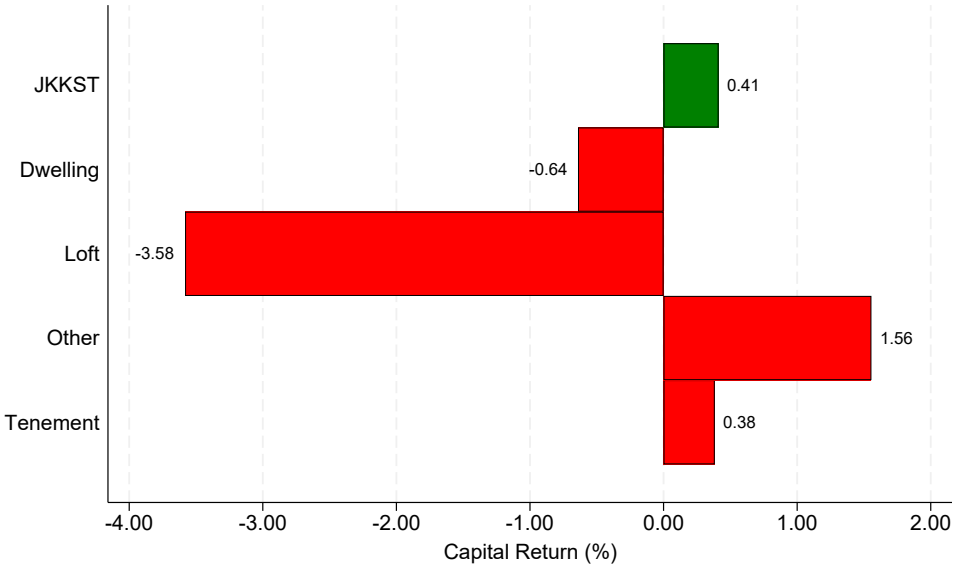
Figure B2 explores capital returns further by property type and neighborhood, revealing some heterogeneity in the real estate market. Different building types appear to cater to distinct buyer or investor groups, leading to variations in price growth. Yet, property investors owning tenements or properties in the neighborhoods with substantial populations of Black households—Central Harlem-Morningside Heights, East Harlem, Washington Heights-Inwood—did not seem to have experienced below-average capital gains over this period. Figures B3 and B4 show estimated price indexes by property type and neighborhood, respectively.

FIGURE B1: BENCHMARKING MANHATTAN INDEX AGAINST U.S. TRENDS

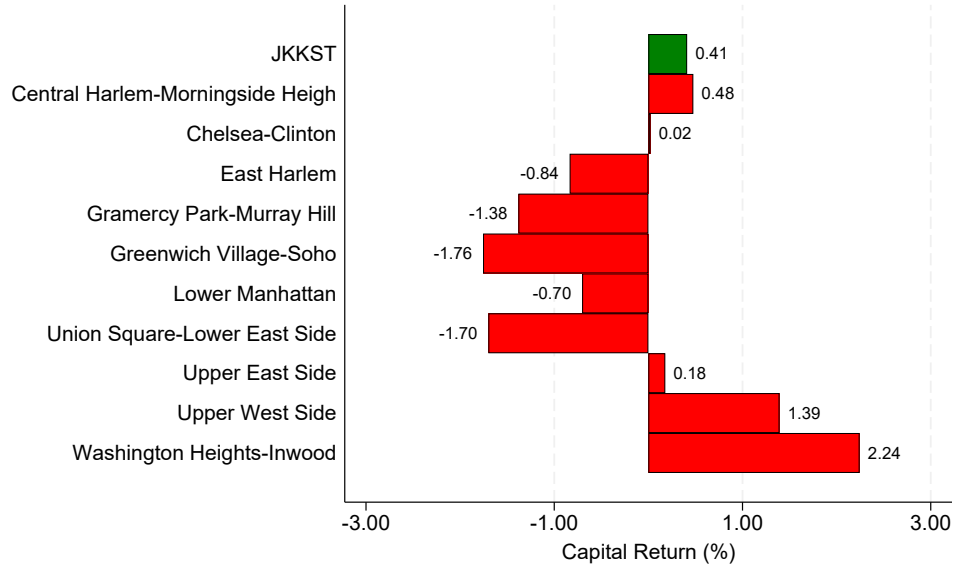


Notes: This figure shows indexes constructed from property transactions between 1912 and 1939 from the *Real Estate Record and Builders' Guide*. Our returns are estimated using hedonic models pooling data over adjacent years (e.g., 1912-1913, 1913-1914). The indexes are normalized to $1 at the start of the period. The JKKST series is constructed using the housing capital return series for the USA in Jordà et al. (2019).

FIGURE B2: CAPITAL RETURNS BY PROPERTY TYPE AND NEIGHBORHOOD



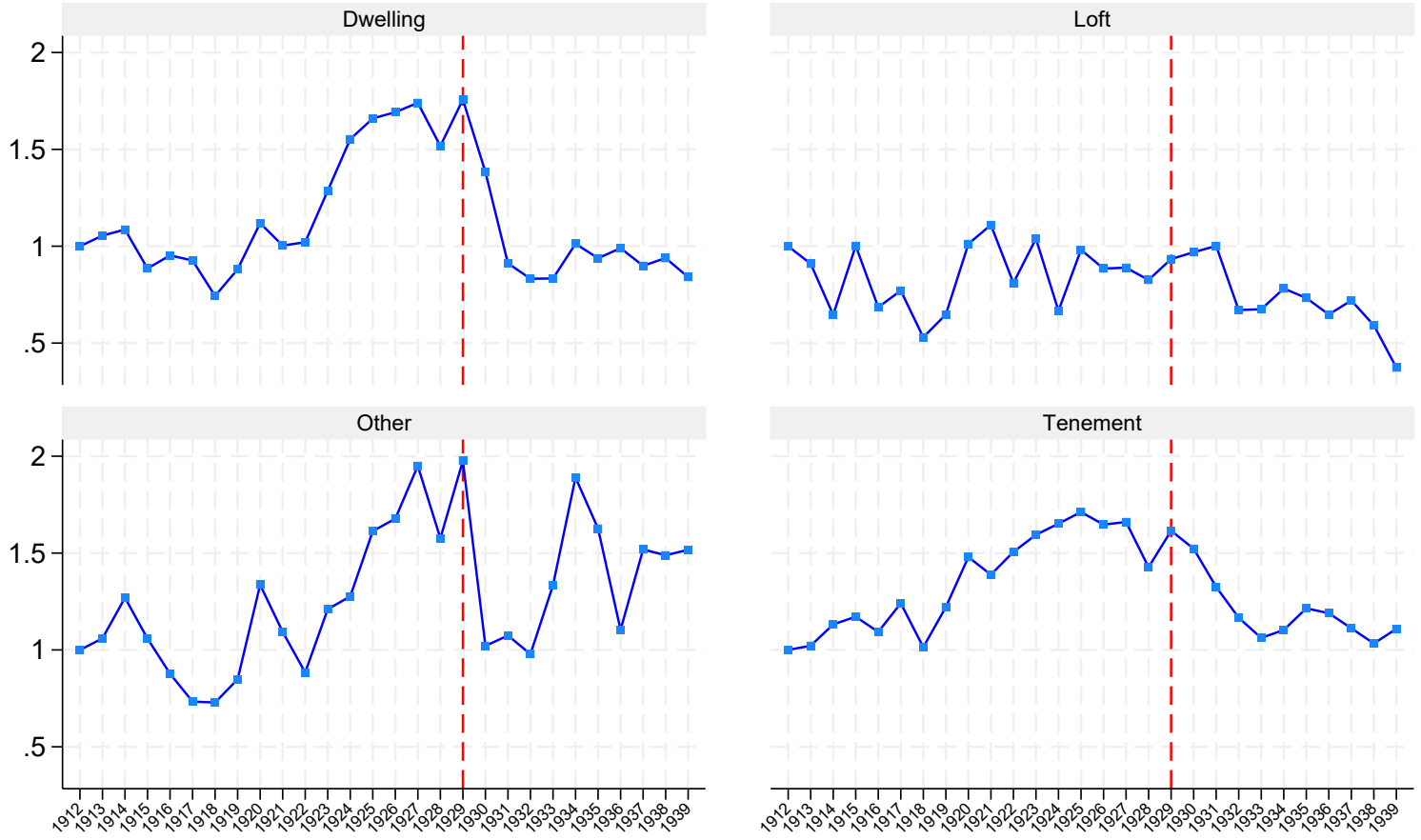
A: PROPERTY TYPE



B: NEIGHBORHOOD

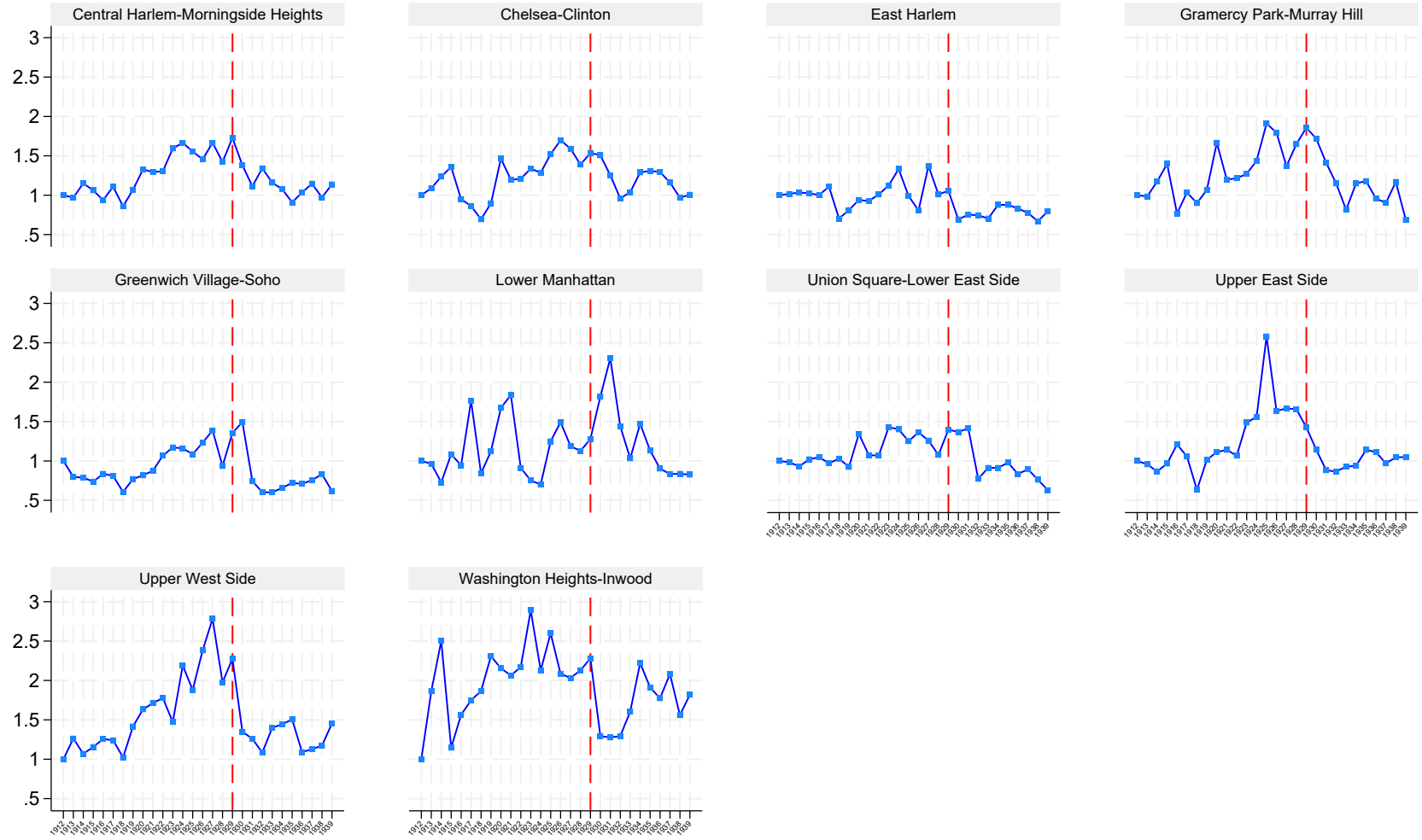
Notes: These figures show estimated annualized capital returns between 1912 and 1939. Our returns are estimated using hedonic models pooling data over adjacent years (e.g., 1912-1913, 1913-1914). The JKKST estimates are constructed using the housing capital return series for the USA in Jordà et al. (2019).

FIGURE B3: INDEXES BY PROPERTY TYPE



Notes: This figure shows adjacent-period hedonic price indexes constructed by building type. The indexes are normalized to one at the start of the period.

Figure B4: Indexes by Neighborhood



Notes: This figure shows adjacent-period hedonic price indexes constructed by neighborhood. The indexes are normalized to one at the start of the period.

Table B1: Hedonic Regressions - Main Periods

	(1) 1912-1939	(2) 1910s	(3) 1920s	(4) 1930s
Log Square Footage	0.338***	0.314***	0.282***	0.438***
	(0.019)	(0.037)	(0.028)	(0.030)
Lot Shape	0.061***	0.028	0.056***	0.074***
	(0.009)	(0.020)	(0.013)	(0.013)
Foreclosure=1	-0.227***	-0.152***	-0.565***	-0.146***
	(0.017)	(0.028)	(0.035)	(0.027)
Building Type (base=Tenement)				
Dwelling	0.263***	0.289***	0.203***	0.304***
	(0.023)	(0.041)	(0.035)	(0.042)
Loft	0.369***	0.451***	0.373***	0.293***
	(0.031)	(0.061)	(0.049)	(0.050)
Other	0.680***	0.641***	0.661***	0.663***
	(0.032)	(0.064)	(0.055)	(0.052)
Construction Material (base=Brick)				
Other	-0.376***	-0.284***	-0.369***	-0.379***
	(0.055)	(0.082)	(0.086)	(0.141)
Stone	0.143***	0.143***	0.189***	0.098***
	(0.017)	(0.031)	(0.028)	(0.030)
Additional Features (base=None)				
Store	0.131***	0.078**	0.094***	0.183***
	(0.018)	(0.034)	(0.030)	(0.031)
Number of Stories (base=Six and Over)				
1	-0.121	-0.619***	-0.154	0.217
	(0.095)	(0.169)	(0.161)	(0.150)
2	-0.357***	-0.614***	-0.473***	-0.013
	(0.058)	(0.110)	(0.090)	(0.098)
3	-0.559***	-0.795***	-0.658***	-0.244***
	(0.039)	(0.075)	(0.060)	(0.066)
4	-0.313***	-0.467***	-0.380***	-0.129**
	(0.033)	(0.063)	(0.053)	(0.053)
5	-0.230***	-0.364***	-0.253***	-0.100**
	(0.027)	(0.051)	(0.045)	(0.044)
Neighborhood (base=Union Sq.-Lwr. E. Side)				
Central Harlem-Morningside Heights	-0.096***	-0.176***	-0.105**	-0.058
	(0.026)	(0.046)	(0.042)	(0.046)
Chelsea-Clinton	0.362***	0.236***	0.325***	0.429***
	(0.028)	(0.052)	(0.043)	(0.052)
East Harlem	-0.070***	-0.094**	-0.194***	0.012
	(0.026)	(0.047)	(0.044)	(0.047)
Gramercy Park-Murray Hill	0.423***	0.234***	0.352***	0.573***
	(0.031)	(0.062)	(0.048)	(0.055)
Greenwich Village-Soho	0.018	0.012	-0.032	0.060
	(0.030)	(0.057)	(0.047)	(0.053)
Lower Manhattan	0.242***	0.178*	0.108	0.490***
	(0.050)	(0.091)	(0.077)	(0.092)
Upper East Side	0.364***	0.177***	0.294***	0.484***
	(0.030)	(0.058)	(0.046)	(0.055)
Upper West Side	0.277***	0.098	0.272***	0.350***
	(0.033)	(0.063)	(0.050)	(0.056)
Washington Heights-Inwood	0.035	-0.106	-0.013	0.110
	(0.040)	(0.070)	(0.065)	(0.067)
Location				
Log Miles to Nearest Subway	-0.152***	-0.136***	-0.169***	-0.149***
	(0.012)	(0.021)	(0.020)	(0.020)
Mean Dep Var.	9.887	9.745	10.171	9.700
R-sq (adj.)	0.288	0.273	0.259	0.277
Observations	19771	4891	7361	7519
Year × Quarter FE	Yes	Yes	Yes	Yes

Notes: This table reports coefficients from regressions where the dependent variable is the log of the property transaction value. Robust standard errors in parentheses. *p<0.1, **p<0.05, ***p<0.01.