Essays on Inequality in Cities

by

Laura Weiwu

B.A., B.S., Stanford University (2018)

Submitted to the Department of Economics
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

May 2024

Authored by: Laura Weiwu
Department of Economics
January 15, 2024

Certified by: David Autor
Ford Professor of Economics
Thesis Supervisor

Certified by: Dave Donaldson
Class of 1949 Professor of Economics
Thesis Supervisor

Certified by: David Atkin
Professor of Economics
Thesis Supervisor

Accepted by: Isaiah Andrews
Professor of Economics
Chair, Departmental Committee on Graduate Students

Essays on Inequality in Cities

by

Laura Weiwu

Submitted to the Department of Economics
on January 15, 2024, in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

Abstract

This dissertation investigates the equity consequences of one the largest and most ambitious public works programs in the world—the Interstate highway system—on inequality across groups and the intergenerational outcomes of children.

In my primary dissertation project described in chapters one and two, I quantify how Interstate road infrastructure construction during the 1960s increased racial inequality in American cities. As a mechanism for why these disparities arise, I explore the central role of exclusionary institutions that limited access to residential neighborhoods for minority families and how they shape the distributional effects of Interstate highway policy.

I develop a quantitative model with rich spatial heterogeneity to capture how in equilibrium, institutions interact with social preferences and the housing market to influence inequality in highway impacts. The model structure crucially enables distinguishing the extent to which segregation is de jure, versus due to economic differences or social homophily, to measure which barriers are most meaningful for the residential choices of Black families. I estimate the model using spatially-granular historical records constructed from restricted Decennial Censuses for 25 cities and archival road network maps that I digitized for 71 cities. The empirical variation for estimation derives from quasi-random placement of Interstate routes relative to historical comparison roads and from spatial discontinuities in where racial exclusion prevailed across locations.

With the empirical estimates and the model framework, I find that highways produced commuting benefits that accrued largely to suburban neighborhoods and environmental harms concentrated in central areas, where the Black population resided, leading to their losses from Interstate highways. Exclusionary institutions are the primary force behind segregation in the 1960s and account for the majority of the racial gap in Interstate impacts. When I eliminate constraints on where Black households are permitted to live, rather than being harmed by highways, the Black population achieves gains. This shift in welfare occurs because they are no longer confined to the urban core, where highway costs outweigh benefits, and the findings highlight the key role of institutions in shaping the disparate incidence of place-specific shocks.

In my secondary dissertation project described in chapter three, I measure the consequences of place-based interventions on children's long-run outcomes, continuing with the setting of the Interstate highway system.

The first step to making this research possible is building an infrastructure of intergenerational linkages encompassing 100 years of economic events. At the U.S. Census Bureau, I have played a leading part in creating novel parent-child linkages that cover the universe of the U.S. population born between 1964 and 1979—on the order of tens of millions of individuals. Construction uses machine learning models, string-comparison methods from natural language processing, and immense

administrative tax datasets.

Using these linkages, I document that upward mobility for Black children was depressed during this period, where children from families in the top-quintile of parental income were more likely to enter the bottom-quintile in adulthood rather than remain in the top-quintile. I then investigate how the Interstate highway network affected the economic mobility of these households. Place-based policies, such as transportation infrastructure, aim to benefit areas they directly target, and highways do so by improving a neighborhood's access to employment. However, when a policy is sufficiently large, migration in equilibrium alters the peer composition of both the locations households seek out and the locations they depart from. In the case of the Interstate network, as this migration was highly differential by economic status (higher educated, White families left central areas for suburbs with increases in connectivity), inequality was amplified across space—suburban neighborhoods experienced job access increases and enhanced peer externalities while central neighborhoods faced solely declines in peer externalities. These direct and indirect impacts then subsequently influence intergenerational mobility by race.

Thesis Supervisor: David Autor

Title: Ford Professor of Economics

Thesis Supervisor: Dave Donaldson

Title: Class of 1949 Professor of Economics

Thesis Supervisor: David Atkin

Title: Professor of Economics

Acknowledgements

I have been extremely fortunate to have spent the past several years in the doctoral program in Economics at MIT surrounded by brilliant advisors, wonderful friends, and stimulating intellectual conversations.

My advising committee consists of the three (best) Davids in the department, who have each contributed a unique and special part to my growth as an economist. I deeply admire their talent, kindness, and wit, and hope to emulate these traits when I am an advisor to future students. They have spent countless hours providing academic and life coaching, listening to my presentations, and carefully reviewing my writing. I am eternally grateful.

I am thankful for the unconditional love provided by my family. My twin brother, Laurence, who I have shared a whole life with and who is one of the most good-hearted people I know. My mom for teaching me the importance of self-sufficiency. My late grandfather for always being so patient with me and instilling in me the importance of an education.

My friends have also made a huge difference in my life and have taught me so much as a person during my time in Cambridge. I will miss you all dearly when I leave for the West Coast. My friends Akhila and Minjia for listening to many intimate personal details and the highs and lows of relationships, family, and graduate school. Natalie, Matt, Will, Vanna, Melissa, and Udai for many fun gatherings (and introducing me to trapeze). To Mike, for all the years we spent together and your support. My friends, co-authors, and officemates Vincent, Brandon, Tishara, Nancy, Garima, Viola, Isa, Dean, Ray, Deivy, Vishan and many others for making research so much more fun and creating a community in the department. My friend Ridge who understands the perils of academia and for decades of friendship since middle school.

I am very appreciative of the advice and assistance provided by the many economists at the Census that helped with data assembly and provided feedback: Martha Stinson, Sean Wang, Katie Genadek, Jonathan Rothbaum, Todd Gardner, and Dani Sandler.

Many mentors along the way have been instrumental in reaching this point. In particular from Stanford, Rebecca Diamond, Matthew Gentzkow, and Pascaline Dupas who believed me and pushed for me to come to this program. Teachers throughout my life, managers from internships, and parental figures have encouraged and motivated me to go down this path.

Biographical Notes

Laura Weiwu was born in Beijing, China and immigrated to the United States at the age of four. She spent most of her childhood in Houston, Texas and moved to California to attend college at Stanford University. After graduation, she will join UC Berkeley as an Assistant Professor at the Haas School of Business.

Table of Contents

Chapter 1 – Unequal Access: Racial Segregation and the Distributional Impacts of Interstate Highways in Cities – Empirics

1 Introduction

The Interstate highway system is one of the defining infrastructure projects of American history and dramatically transformed cities across the country. Immense sums of public investment have been channeled into Interstate development to achieve benefits from facilitating commuting between locations (Department of Transportation, 2021). But along with these benefits, highways were deeply destructive and produced substantial costs through local environmental pollution, noise near routes, and the splitting of existing communities (Mohl, 2004; Currie and Walker, 2011).[1]

I investigate the distributional consequences of the Interstate system, which conventional wisdom points to as having promoted American economic growth in aggregate—yet, a few crucial facts allude to highly heterogeneous impacts. Coinciding with highway construction in the 1960s, segregation reached extraordinary levels as pervasive legal and extralegal institutions excluded Black Americans from homogeneous White neighborhoods (Massey and Denton, 1993; Cutler et al., 1999; Rothstein, 2017). This segregation concentrated Black families in the center of cities, which Interstate roads were designed to intersect. As political advocacy by lower-income, minority neighborhoods was often ineffectual in averting construction, disadvantaged groups in central areas disproportionately bore the costs (Glaeser and Ponzetto, 2018; Brinkman and Lin, 2022). Commuting improvements further appeared largely in suburban areas that prohibited the entry of minority families. Combined with differential car usage, the commute benefits of highways then accrued unequally by race.

A long history of analysis across various disciplines has examined how the Interstate system acutely increased inequality in cities, e.g. Caro (1974), Jackson (1985), and Rose (1990). However, considerable challenges have stood in the way of a systematic assessment. This paper overcomes those challenges by undertaking the collection of spatially-granular commuting statistics for the entire U.S. using restricted Census microdata and newly digitized road maps. These unique historical data enter a quantitative spatial framework that richly captures the reallocation of households and the equilibrium evolution of neighborhoods and workplaces after highway construction. I propose a novel mechanism for why, in particular, racial disparities emerged: Black households were constrained by *exclusionary institutions* from leaving central areas, which interacted with the Interstate system to create stark inequality in impacts. To quantify the strength of this mechanism, I concurrently address the challenge of distinguishing institutional forces of segregation from competing *economic forces* through housing affordability and *social forces* through preferences for same-race neighbors.

Taken together, I find via the model framework that the Black population faced losses from Interstate highways while the White population garnered gains. Within race, disparities by class are

[1] The Interstate Highway System originally cost \$114 billion, around \$500 billion in 2020 dollars, and is to date the most expensive infrastructure project in U.S. history. Investment in highways has continued with the Infrastructure Investment and Jobs Act of 2021 providing \$550 billion to fund infrastructure. To address the local costs of highways, \$1 billion is set aside to reconnect communities divided by the Interstate highway system. An additional \$21 billion is allotted to environmental remediation with a focus on environmental justice (Department of Transportation, 2021).

minimal. Exclusionary institutions are a key determinant of segregation and inequality in impacts—absent institutional barriers, Black households would have resided farther from the central business district and benefited more from Interstate highways, reducing the racial gap in impacts. While White families migrated outwards in reaction to the positive and negative consequences of highways, enlarging their gains, Black families were initially concentrated and remained in the urban interior. Residential exclusion, which I document favors White households by preventing racial integration and housing competition, is a central force for why Black households refrained from migrating to suburban areas despite the wide-ranging repercussions of Interstate highways.

To reach these conclusions, the paper is organized in three parts. The first characterizes the empirical variation provided by two natural experiments. With quasi-random placement of Interstate highways, I measure population responses across neighborhoods, which feeds into the estimation of parameters for the costs and benefits of highways and a subset of the forces of segregation. I then exploit spatial discontinuities in where exclusionary barriers prevailed across neighborhoods and sharply establish that institutions are an important determinant of racial segregation. In the second part of the paper, I develop the model features and discipline the magnitude of the channels in the spatial equilibrium framework using the two sources of quasi-experimental variation. Lastly, I employ the framework to conduct distributional analyses across race and class and explore the importance of discriminatory institutions for inequality in highway impacts. The three parts are presented linearly, although each piece informs and is intertwined with the others in the rest of the paper.

Following that order, I show first empirically that the Interstate system introduced substantial costs and benefits to cities. I leverage restricted microdata from the 1960 and 1970 Decennial Censuses, which include the previously under-studied Journey to Work survey, to construct the first historical measures of commute flows in 25 cities. These flows, crucially, are disaggregated into two categories of race (White/Non-White) and two categories of class i.e. education (less than high school/high school graduate). Non-White is treated as synonymous with Black.[2] As commuting time surveys with broad coverage were never administered for this period, I generate commute time matrices at high spatial resolution across 49 million bilateral pairs for changes from Interstate development. This calculation uses road maps I digitized from Shell Atlases for 71 cities and a database on when each Interstate segment was built to focus on the 1960 to 1970 period (Baum-Snow, 2007).

In long differences between 1960 and 1970, I document declines in population, rental prices, and racial composition (percentage White) by highways, which are informative of the local costs and equilibrium responses for neighborhoods. The population shifts are driven by White migration as I find only a small Black population response, and the differential migration by race contributes to the changes in racial composition.

Non-random placement of Interstate routes, often in disadvantaged areas, conflates selection on trends with highway costs. I attain sharper identification by creating comparison areas and instruments for the location of interstate routes. In the Shell Atlases, I separately categorize major roads as

[2]In 1960, 11.4 percent of the population was Non-White and 10.5 percent of the population was Black (Census Bureau, 1961). Only 3.5% of the population was Hispanic (with Spanish origin surname) in 1960 and were enumerated under the White category as the Census did not ask respondents about ethnicity until 1980. Black households thus comprise almost all of the Non-White population.

candidates for Interstate construction. Not all were converted, and these untouched historical roads serve as natural control groups. Geographic features such as the location of historical railroads, ports, canals, rivers, and the central business district which highways aimed to connect are used as controls to compare neighborhoods with similar propensity of receiving an interstate road. As two separate instruments, I digitize planned engineering maps for 100 metropolitan areas that were less subject to political influences and construct a Euclidean ray network to intersect cities in the plans where neighborhoods coincidentally located between cities are treated (Chandra and Thompson, 2000). With this multi-fold empirical strategy, I find declines by highways can be interpreted as causal.

I next show that the population grew in suburban areas where connectivity increased as evidence for the commute benefits of highways. Because peripheral growth was ongoing during the 20th century and did not originate solely from Interstate construction,[3] I control for distance from the central business district and exploit variation within the suburbs relative to the comparison roads (Borusyak and Hull, 2023). I employ the same identification strategy described above for identification of where commuting benefits emerged. Measures of Commuter Market Access (CMA), a model-implied aggregator of commute costs, summarize the commuting impacts of highways (Donaldson and Hornbeck, 2016; Tsivanidis, 2022). Consistent with the population response to highway costs, the response to the benefits is highly differential by race with substantial White migration and near zero Black migration. Equilibrium responses through changes in rents and racial composition to CMA improvements are of smaller magnitude compared to the large movement outwards of White households.

Why do Black households not respond to either the costs or benefits of Interstate highways? Spatial frictions from institutions may be a central force for their immobility. As a proxy for formal and informal institutional barriers, I draw upon redlining maps from the Home Owners' Loan Corporation (HOLC) which evaluated the credit risk of neighborhoods and traced longstanding racial and economic divisions in cities (Nelson et al., 2020). The lowest grades were issued to racially diverse neighborhoods, deeming them "redlined." These maps represent practices by private actors, real estate agents, and government officials to preserve the homogeneity of uniformly White neighborhoods through means of violence, refusal to sell or rent housing, and collective animosity.

In line with prior evidence, I find the maps are highly effective at conveying where Black households were permitted to live and that racial composition discontinuously shifts across the redlining borders (Fishback et al., 2020; Aaronson et al., 2021). Including redlining fixed effects and measuring Black responsiveness to CMA within redlined areas, population responses are no longer zero and are in fact significantly positive. These results, while suggestive, indicate the limited response by the Black population stems from barriers that inhibit free movement across neighborhoods.

The reduced form facts, though transparent, are nevertheless unable to separate feedback between the equilibrium responses and lack the structure needed to assess welfare. In the second part of the paper, I extend Ahlfeldt et al. (2015) and Tsivanidis (2022) to develop a within-city model of neighborhoods that has a novel focus on the role of institutions for heterogeneous impacts by race.

By design, the model incorporates the empirical facts of household reallocation across locations

[3] Increasing crime in the central city, desegregation of school districts, the Great Migration, and subsidies for suburban development were parallel contributors to suburbanization (Jackson, 1985; Cullen and Levitt, 1999; Boustan, 2010; Baum-Snow and Lutz, 2011; Boustan, 2012).

and the endogenous evolution of housing prices, racial composition, and wages. The sources of sorting are captured via residential elasticities for spatial mobility by race, differential sensitivity to prices through housing consumption shares, and endogenous neighborhood amenities through heterogeneous preferences for racial composition (percent White). Institutional barriers are time-invariant to highway policy and hence contained in fundamental amenities (immutable in equilibrium, commonly thought of as "natural" amenities) where discrimination lowers amenities of neighborhoods. Direct impacts of highways appear through commute time reductions between bilateral pairs of locations and the decline of fundamental amenities from local costs that decay over distance from Interstate routes. In equilibrium, housing prices change elastically, and wages adjust at firms as employment reallocates in reaction to reductions in commute costs.

Importantly, the model's expressions for residential choice govern how segregation arises *in levels* and *in changes* in response to shocks. One set of parameters determines both the forces of segregation that initially concentrated the Black population prior to highway development and how the Interstate system affects sorting with feedback mechanisms for welfare. However, a challenge that looms large is disentangling institutional barriers from economic and social forces.

The Interstate shock provides variation for estimating some of the non-institutional forces: residential elasticities and racial preferences, following evidence on differential mobility by race to CMA improvements. Reminiscent of the reduced form equation, the structural equation implies residential elasticities are far larger for White households. The same logic informs the creation of instruments for racial composition to estimate racial preferences (Davis et al., 2019). Given high White mobility and low Black mobility to the highway shock, the model predicts declines in percent White in the urban core. With the simulated predictions as shifters, I find White households strongly prefer White neighbors while racial preferences for Black households are also homophilic but much weaker. These preferences do not conflate institutional barriers as estimation includes redlining fixed effects, using variation within redlined areas where institutions are less likely to bind. They also do not represent preferences for socioeconomic status as a rich set of controls accounts for socioeconomic factors.

Housing price sensitivity is calibrated to the Consumer Expenditure Surveys, and I find less-educated, Black households are more sensitive to prices. With the above set of parameters, I decompose population differences, i.e. across the redlining borders, into the institutional, economic, and social components where the institutional component is contained within the structural residual of fundamental amenities. In the next section, I return to the fundamental amenities to isolate institutional segregation as a core mechanism for disparities from Interstate highways. But as time-invariant terms, they are not prerequisites for measuring highway policy impacts.

In the third part of the paper, I combine the structure of the model and key parameters to evaluate the effects of Interstate highways on inequality. In a general equilibrium counterfactual, I find that Interstate construction lowers welfare by -1.04% for Black households and raises welfare by 2.9% for White households. I compare these numbers to a calculation of direct impacts where location choices are fixed. As the Black population resided in areas near Interstate routes and commuted with cars at a lower rate compared to the White population, both direct effects push for increases in inequality with losses of -1.6% for the Black population and gains of 1.7% for the White population. Accounting for equilibrium effects only slightly raises welfare for the Black population but greatly increases it for the White population. This widening of the racial gap is a result of the lower reallocation of Black households, who remained in central areas where costs outweigh benefits, as general equilibrium outcome adjustments play a smaller role and are somewhat offsetting for welfare.

Finally, I show that institutional segregation is a primary mechanism behind racial disparities in highway impacts. I document that racial composition sharply shifts over borders of redlining maps, but not all of the discontinuity can be attributed to institutions. Housing prices vary and sorting from racial preferences can reinforce any differences along the border. In a border discontinuity design and using the previously estimated parameters, I find that 65% of the massive rise in population (140 log point increase) for Black households crossing into redlined neighborhoods is unaccounted for by prices, racial preferences, or socioeconomic variables. The remaining component represents residual fundamental amenities containing exclusionary institutions. For White households, the 50 log point decrease in population entering redlined areas is fully explained by their racial preferences against Black neighbors. No residual discontinuity remains, and the identification assumption of fundamentals being continuous across the border is evidently satisfied.[4]

I simulate removing institutions at the border, and in this new environment, constructing Interstate highways leads to less unequal impacts: Black households are more spatially dispersed and consequently are harmed less by Interstate development. Relaxing barriers additionally reduces welfare for White higher-educated households, motivating why exclusionary institutions were originally established. However, the racial gap for highway impacts does not close substantially. While the border discontinuity design has the advantage of allowing for a testable identification assumption, the change in racial composition at the border, while sizable, fails to capture the stark degree of segregation more broadly. White neighborhoods in close proximity to minority areas are less exclusionary than those in outlying suburbs fortified from contact with Black households.

Yet, a lesson is learned from the border design, namely that much of racial differences in fundamentals result from institutions. With this insight, I extrapolate to examine barriers writ large. I make the stronger assumption that fundamental amenities, commonly considered to be natural amenities, should not be valued differentially by race *for all neighborhoods*. Setting Black fundamentals equal to White fundamentals (which are unaffected by housing discrimination), I provide Black households the same access to neighborhoods that White households possess, a hypothetical upper bound on how far exclusionary barriers can be eliminated. Under this arrangement, I find that the Black population gains by 1% from Interstate highways so that all groups benefit from highway development, and the racial gap in welfare impacts closes by a striking 54%.

Related Literature – This paper builds on a literature in quantitative spatial economics (see Redding and Rossi-Hansberg (2017) for a review) that proceeds from an earlier body of work on urban models (Rosen, 1979; Roback, 1982). Most closely, it extends the frameworks of Ahlfeldt et al. (2015)

[4] I probe this assumption in several tests. I remove physical barriers of large roads, railroads, and highways from the sample to only measure social barriers. I further drop areas near school district borders that may fall along the borders of redlining maps. After gathering several datasets on natural amenities, I find that land cover types of open water and wetlands are continuous along the border, supporting the identification assumption of no change in fundamental amenities.

and Tsivanidis (2022) and applies these powerful tools to disentangle the various sources of segregation. Sorting via endogenous amenities has been studied recently by Diamond (2016) and Almagro and Dominguez-Iino (2020) with earlier work summarized in Kuminoff et al. (2013). The departure of this paper is to map institutions, a central determinant of spatial inequality by race, into model fundamental amenities where heterogeneity in these fundamentals has been greatly under-explored.

This paper also contributes to a rich literature on the impacts of transportation infrastructure. Duranton and Turner (2012), Duranton et al. (2014), Allen and Arkolakis (2014), Donaldson and Hornbeck (2016), and Donaldson (2018) examine the benefits of transportation improvements through their reduction of travel or trade frictions. Directly related to the setting of this paper are several studies on the Interstate highway system (Baum-Snow, 2007; Michaels, 2008; Baum-Snow, 2020; Brinkman and Lin, 2022). Recent work by Bagagli (2022) and Mahajan (2023) provides reduced form evidence on heterogeneous racial responses but does not measure welfare. This paper is the first to quantify the distributional impacts in a comprehensive general equilibrium framework, made possible with novel disaggregated data and granular archival maps that span the entire country.[5]

The final literature tied to this paper is an interdisciplinary one on racial inequality and *de jure* segregation influencing the economic access of Black Americans (Kain, 1968; Hirsch, 1983; Wilson, 1987; Bullard, 1993; Massey and Denton, 1993; Rothstein, 2017). I find that highways increased segregation, as is the case with railroads in Ananat (2011), and their effects are further shaped by pre-existing divides. Related to the methods of this paper, Bayer et al. (2007) employs the border design to study racial preferences for neighbors. I apply the same design instead to test for the presence of institutional exclusion. I find discrimination entails welfare costs, in line with experimental research by Christensen and Timmins (2021, 2022), but I do so in an *observational* setting. I thus uncover far larger magnitudes for the importance of discrimination, and crucially, I go one step further—I examine how racial exclusion *interacts* with highway infrastructure. This interplay between institutions and policy extends to other place-based policies beyond transportation infrastructure, and it animates why there exist profoundly disparate impacts by race.

Summary – The remainder of the paper is structured as follows. Section 2 describes the historical context and data sources. Section 3 presents reduced form evidence on neighborhood changes in cities. Section 4 characterizes the model. Section 5 derives the estimating equations and interprets estimation results. Section 6 conducts welfare and counterfactual analyses. Section 7 isolates the role of institutional segregation for welfare. Section 8 concludes with policy implications.

[5] In other social science disciplines that follow empirical or qualitative approaches, there is evidence of unequal impacts by Hirsch (1983), Rose and Mohl (2012), Avila (2014), Connolly (2014), Rothstein (2017), Nall (2018), Nall and O'Keeffe (2019), and Trounstine (2018).

2 Historical Context and Data on Interstate Highways and Inequality

2.1 The Interstate Highway System

The Federal-Aid Highway Act of 1956 initiated the monumental construction of 42,800 miles of Interstate freeways, which at the time, became the largest and one of the most advanced road networks in the world. President Dwight D. Eisenhower had several motivations for its development. Although the Interstate system would be used for defense if necessary, its primary purpose was to support automobile traffic and stimulate economic growth.[6] Post World War II, several federal interventions spurred the population to flee the urban core and heighten road congestion. The GI Bill promoted homeownership for millions of veterans, and the Federal Housing Act of 1949 expanded mortgage insurance for newly developed suburbs while simultaneously funding renewal of downtown areas. Interstate roads were one factor among many catering to the new social order of cities (Rose, 1990).

The harms and inequities of highways were, however, soon apparent. Urban commentators Lewis Mumford, Jane Jacobs, and Patrick Moynihan criticized how highways displaced neighborhoods, altered the style of urban living, and polluted the nearby environment. Freeway revolts successfully shifted the course of several routes and in some cases permanently terminated their construction.[7] Yet, revolts were only successful for certain neighborhoods, as without the support of influential actors, disadvantaged populations were regularly disregarded in favor of business interests (Rose and Mohl, 2012). In Miami, Interstate 95 razed the Black residential area of Overtown when business leaders aimed to clear out perceived slums to redevelop the central district (Connolly, 2014). City planners likewise targeted highways towards low-income housing for urban renewal, and displacement triggered racial turnover in adjoining working-class neighborhoods (Hirsch, 1983).

Decennial Censuses by Race and Education – To measure disparities in the incidence of the Interstate network, I collect the location of residences and workplaces (which together form the bilateral commuting pair) separately by race and education from restricted Decennial Censuses in 1960 and 1970. Crucially, this decade covers 51% of the construction of the network. Residential units are census tracts, and workplace units are Place of Work Zones, newly assembled as the intersection of county and municipality codes using the 1960 Census Journey to Work questionnaire. I limit the sample to 25 of the largest cities as some cities have few Place of Work zones, and those with available data are listed in Table G.34. Appendix 7 contains more details on data construction.

Race is split into White and Non-White since finer cuts leave too few counts for smaller geographies, and Non-White is considered equivalent to Black as the Black population comprised almost all of the Non-White population in 1960. Education is split into high school graduates for the higher-educated and without a high school degree for the less-educated. Attached to these individuals (and also to residences and workplaces) are wages and quality-adjusted housing prices.

[6] General Clay, the head of the Interstate financing committee, stated "It was evident we needed better highways. We needed them for safety, to accommodate more automobiles. We needed them for defense purposes, if that should ever be necessary. And we needed them for the economy. Not just as a public works measure, but for future growth."

[7] A prominent example is the Lower Manhattan Expressway (I-78) which was shut down after advocacy by Jane Jacobs against metropolitan planner Robert Moses.

For some of the later evidence, I study an expanded set of cities, not just the 25 with Decennial microdata, using tract-level aggregates from IPUMS-NHGIS derived from the Decennial Censuses (Manson et al., 2017). Covering 1940 to 1990, the longer panel of the dataset allows for examining pre-existing differences before and long-run changes post Interstate construction.

Commuting Networks – Commute cost reductions are one of the central impacts of highways, but commuting time surveys were conducted irregularly and for a limited set of cities.[8] Sufficiently large samples are essential for distributional analysis, so I build commute time matrixes using historical road network maps with assumed travel speeds. I digitize Shell Atlases in 1951 and 1956 for 71 cities and categorize roads into superhighways and other major roads, assigned slightly different speeds following historical travel surveys (Rumsey, 2020; Gibbons and Proctor, 1954; Walters, 1961). The Interstate network is set to the speed limit of each segment (in the range of 55-65 mph), and to only examine routes constructed between 1960 and 1970, I exploit panel variation in when different Interstate segments were built using the PR-511 database from Baum-Snow (2007). Commute times are then generated in ArcGIS Network Analyst for 49 million bilateral pairs using the Interstate network as of 1960 and 1970 overlaid on the historical road network.

The lack of data on public transit complicates the analysis of other transport modes, where usage can differ substantially by race. I retrieve reported commute times for other modes from the 1980 Decennial Census, the first census to survey travel time, and non-parametrically estimate times and mode shares (in 1960 and 1970) over bins of bilateral distance and distance from the central business district (CBD) for residences and workplaces. Share-weighted times enter any group-level analysis.

Summary Statistics on Racial Inequality – With these novel datasets, in Table 1, I calculate summary statistics in 1960 which indicate large differences by race *conditional on education* in characteristics related to the Interstate highway system. For example, 49% of Black higher-educated workers commute by car compared to 66% for White higher-educated workers, so upgrades in road speed likely benefit the White population more. Among the higher educated, Black workers are located approximately 3.6 miles closer to the CBD, where commuting improvements are muted, compared to White workers.

Black workers also reside 0.8-0.9 miles closer to a highway within education, which can be due to political influences leading to unequal route placement, as described in the historical context. In Table 2, Columns 1–3 indicate highways were built further from more-educated, higher-income, and White areas at baseline in 1950. This correlation is partially, but not all, driven by highways being designed to intersect the central city, which was more disadvantaged (see Columns 5–6). The geographic inequality in placement then influences the distribution of costs across demographic groups.

Are the racial differences in location explained by economic characteristics? As shown in Table 1, wages, rents, and home values of Black higher-educated workers are comparable to those of White less-educated workers. Yet, the two groups still experience vast differences in location relative to highways and the central city. They also reside in neighborhoods with stark contrasts in racial com-

[8] Brinkman and Lin (2022) use travel surveys for Chicago and Detroit in 1953 and 1956, respectively. However, these surveys only report aggregated flows and are not suitable for evaluating distributional impacts. They are also for years when few segments of Interstate highways were completed.

position. On average, Black higher-educated workers live in neighborhoods that are 42 percent White versus 94 percent White for White less-educated workers. Prices alone are thus not a central determinant of segregation, confirming past findings (Logan, 2011; Bayer et al., 2021; Aliprantis et al., 2022).

2.2 Institutional Segregation

What mechanisms underlie the striking differences in where Black and White families live? Prior to and during the era of highway construction, government and private actions limited the residential choices of Black families. Although certain obstacles were dismantled before the Interstate system, other practices endured.[9] Between 1960 and 1970, the period under consideration, segregation reached its peak as Black Americans migrated from the rural South into crowded, racially-mixed neighborhoods of the urban North (Cutler et al., 1999; Boustan, 2010). Segregation began to decline with the Civil Rights of 1968 authorizing provisions to combat housing discrimination, leading to its colloquial title: the Fair Housing Act. Yet enforcement was uneven, and only through decades of advocacy by fair housing organizations did segregation lessen to the lower levels of the modern day.

Exclusionary institutions that imposed the geographic separation of the races encompass both state law and individual behavior. The two can often be inextricably linked, and an example from Rothstein (2017) illustrates the murkiness of the division. In 1957, shortly after an African American couple—Bill and Daisy Meyers—moved into the White suburb of Levittown, Pennsylvania, 600 demonstrators gathered to pelt their home with rocks while state troopers stood by passively. The state, though it was charged with providing equal protection, neglected to restrain these acts of individual discrimination and as a result, sanctioned the racial violence. In this sense, widespread private prejudice can become institutionalized because it is abetted by government policy.

Despite their importance, hostile forms of discrimination are hard to quantify, so to make progress, this paper uses federal maps to proxy for a complex mix of government and private exclusion.

HOLC Redlining Maps – To discernibly measure where residential discrimination occurred, I employ maps from the Home Owners' Loan Corporation (HOLC), a federal agency formed in 1933 post the Great Depression to appraise the risk of neighborhoods for mortgage refinancing and purchase. These maps were created in consultation with local lenders and reflected existing racial and economic characteristics (Harriss, 1951). Historians at the American Panorama project have digitized them for public use, and high-risk neighborhoods displayed in red are those deemed "redlined" (Nelson et al., 2020). As examples, maps from Los Angeles and New York City are depicted in Figure B.1.

Concurrently with the HOLC as a part of the New Deal, the Federal Housing Administration (FHA) initiated its own practice of "redlining" by denying mortgage insurance to Black families, majority-Black neighborhoods, and socially or racially mixed areas (Hillier, 2003).

[9] In the 1968 Kerner Commission Report to President Lyndon B. Johnson, commission members write "What white Americans have never fully understood — but what the Negro can never forget — is that white society is deeply implicated in the ghetto. White institutions created it, white institutions maintain it, and white society condones it." Policies such as racial zoning established White-only districts through local ordinances, and restrictive covenants placed discriminatory language in property deeds to prevent the sale of homes to anyone outside of the Caucasian race. Racial zoning was outlawed in the Supreme Court case Buchanan v. Warley in 1917, and restrictive covenants were ruled unenforceable in Shelley v. Kraemer in 1948.

While the HOLC maps are not the same as those by the FHA, there exists historical evidence that the FHA was influenced by the HOLC and a correlation exists between the maps for Chicago (Jackson, 1980; Aaronson et al., 2021).[10] Findings from Faber (2020) and Aaronson et al. (2021) suggest the FHA directly increased segregation. Existing racial divisions were often perpetuated by the agency through the encouragement of real estate agents to preserve racially homogeneous neighborhoods, e.g. with restrictive covenants in property deeds that prevented sales to non-White households (Jones-Correa, 2000). However, many of the lines followed past discriminatory patterns, and recent work by economic historians argues not all of the segregation associated with the maps can be placed on policies by the FHA or the HOLC (Fishback et al., 2020). For the purposes of this paper, the HOLC borders delineate institutional segregation from both government actions (through federal redlining) and private behaviors (through exclusionary practices such as violence or refusal to sell homes to Black families), where as evidenced above, the two are challenging to separate.[11]

Summary Statistics on Segregation – In spite of some limitations, I provide descriptive evidence that the HOLC maps are highly informative for where the Black population lived and suggest where institutions prohibited them. Figure B.2 shows Black households are far more clustered near the CBD compared to White households. Redlined neighborhoods presented in the same figure are also located close to the center in a pattern strikingly similar to the distribution of the Black population. These two facts are consistent with the heavy concentration of the Black population in redlined areas. For example, Table A.2 lists that in 1950, 93% of the Black population in Chicago lived in redlined areas compared to 32% for the White population. After 1950, these neighborhoods continued to be the residences of most Black families as Figure B.3 shows the racial composition of the median non-redlined tract remained close to 100% White until 1970. Indeed, I find in Figure B.2 that while White households began to shift outwards after 1950, the Black population was still predominantly located in neighborhoods within 10 miles of the central city.

However, redlined areas were not wholly Black, and most were racially integrated. As indicated in summary statistics Table A.3, the majority of the population in redlined neighborhoods was White. These simple descriptive statistics notably reveal an asymmetry in the residential locations of the two racial groups. While White households were located across all neighborhoods, Black households lived in a reduced set that were often concentrated in older downtown areas.

Although these statistics are highly suggestive, they do not rule out economic differences or social preferences of homophily as contributors to segregation. This paper aims to provide compelling evidence that institutions are indeed a dominant force, which then interacts with Interstate highways to widen racial inequality. In the next section, I provide evidence on the geographic distribution of

[10]FHA maps for most cities have unfortunately been lost. Fishback et al. (2022) finds additional maps for Baltimore City, Maryland; Peoria, Illinois; and Greensboro, North Carolina.

[11]While the qualitative literature documents cases of discriminatory pricing preventing Black households from living in majority-White neighborhoods as in Taylor (2019), I do not find that Black families faced substantially higher prices in non-redlined neighborhoods for similar quality housing. In Appendix Table A.1, after accounting for neighborhood fixed effects and housing quality controls, Black households face 3% higher rents in non-redlined areas and 8% higher rents in redlined areas. Because the race differential is not greater in non-redlined areas, the concentration of Black households in redlined areas can not be explained by lower price discrimination in redlined areas. However, the existence of a positive race differential in rents suggests the presence of some discrimination.

the costs and benefits of Interstate highways, and this spatial dimension is especially critical for the equity consequences given that the Black population was concentrated in central areas.

3 Motivating Evidence on Neighborhood Changes in Cities

I present several empirical facts to illustrate how Interstate highways impacted cities through their localized costs, commuting benefits, and the subsequent equilibrium responses in the characteristics of neighborhoods. These results motivate the key mechanisms I incorporate into the quantitative model and inform which sources of quasi-experimental variation enter parameter estimation.

3.1 Population and Equilibrium Responses to Highway Costs and Benefits

With a long differences specification, I measure changes in population, which revealed preference logic implies is correlated with the cost and benefits of highways, and equilibrium responses in neighborhood rents and racial composition. Neighborhoods are tracts denoted by i.

$$\Delta Y_i = \beta_1 \log DistHW_i + \beta_2 \log DistCBD_i + \beta_3 Redlined_i + \mathbf{X}_i \eta + \gamma_{m(i)} + \epsilon_i \tag{1}$$

The time horizon is over 1960 to 1970 (and occasionally 1950 to 1960 for cities with earlier highway construction, stacked into one sample) using the expanded set of cities in the public-use dataset.[12]

I examine differences over distance from highways $DistHW_i$ as representing the costs and over distance from the central business district $DistCBD_i$ as representing the benefits. A redlined indicator captures heterogeneity in redlined versus non-redlined areas and leads the empirical variation for changes related to highway costs and benefits to come from within the two types of neighborhoods.

Results are presented in Table 3 Panel A where no controls are yet specified ($\mathbf{X}_i$ is empty).[13] City fixed effects $\gamma_{m(i)}$ for each city m are included to exploit variation within cities. In later specifications, I discuss which controls are fruitful to incorporate for identification. Standard errors adjust for spatial correlation following Conley (1999) within a radius of 1 kilometer.

I find relative declines by highways in population, rental prices, and percentage White for racial composition, displayed in log-log form as elasticities in Columns 1–3. These declines are indicative of the harms of highways reducing the desirability of adjacent areas. Additionally, I document growth in population, rental prices, and percentage White in suburban areas which become better connected by the Interstate highway system, although I do not yet make any claims on causality. The shifts in racial composition stem largely from White households leaving areas by highways and the central city, as shown in Columns 4 and 5. The latter column further indicates Black households did not

[12] Tracts in cities where less than 10% of the mileage of Interstate highways was built in 1960 are in the 1960 to 1970 sample. Tracts in cities where less than 10% of the mileage of Interstate highways was built in 1950 (occasionally some cities began construction on Interstate highways before the Federal Highway Act of 1956) but more than 10% was built by 1960 are in the 1950 to 1960 sample. Cities here are Core Based Statistical Areas and include both Metropolitan Statistical Areas and Micropolitan Statistical Areas.

[13] One control included in all specifications is the gradient = $DistCBD_i / DistHW_i$ to account for how $DistHW_i$ is mechanically lower near the city center. This is also addressed by including $DistCBD_i$ in the estimating equation.

move away from highways as the coefficient is not statistically significant, and while they migrated slightly into the suburbs, they did so at a far lower rate compared to White households.[14]

Feedback channels link these three neighborhood characteristics. As income is correlated with race, a positive relationship between changing rents and changing racial composition may be due to differential responsiveness to price changes i.e. non-homotheticity in housing consumption. Preferences for racial composition further reinforce sorting. For example, when an area becomes less White in response to a direct shock, the feedback effect of homophilic preferences precipitates more out-migration of White households. These population responses then transmit into housing price changes. Relevant to the quantitative model, the equilibrium system should aim to characterize how the channels are determined simultaneously and measure the degree of importance for each.

Comparing magnitudes of outcomes, the changes in rents and racial composition are smaller than the White population response, so for the impacts of the Interstate system, the large reallocation of White households may be more meaningful for their gains than adjustments in these two characteristics. The low reallocation of Black households also has important implications for their welfare—if they remain in central areas, where costs from highways are high and benefits are minimal, they are unlikely to gain from Interstate highways.

Heterogeneity by Redlining – I find in the bottom row of Table 3 Panel A that redlined areas experienced substantial inflows of Black households and outflows of White households, leading to a drop in the percentage White of redlined neighborhoods. Although not directly related to the Interstate system, this result shows Black households moved predominantly into redlined neighborhoods.

In addition to large changes over time in the racial composition of redlined areas, there exist sizable cross-sectional differences between redlined and non-redlined neighborhoods. As depicted in Figure 1, percentage White sharply shifts over the border of the HOLC maps with around a 20 percentage point decline crossing into redlined neighborhoods. Along with the earlier summary statistics in the historical context, these findings illustrate how the Black population, in flows and in levels, was concentrated in redlined areas in the center of cities.

Historical Comparison Areas – Non-random placement of Interstate routes can lead changes by highways to be contaminated by selection on trends. To obtain cleaner identification, I create comparison areas that were likely to have received an Interstate highway. I follow guidance from the 1944 report, *Interregional Highways*, which recommended highway engineers: (1) build along existing roads with heavy traffic since a primary goal was to combat congestion, and (2) account for topographic features and other transportation methods.[15]

I consider super-roads from the digitized Shell Atlases as Interstate candidates, where those not converted to highways are the counterfactual control routes to be compared against. This strategy

[14] I examine the additional outcome of home values in Appendix Table A.6 and find no significant changes by highways, so they are not considered to be a central impact of Interstate highways.

[15] The introduction to *Interregional Highways* states that the "recommended system follows in general the routes of existing Federal-aid highways" and Interstate development would occur through "the improvement of a limited mileage of the most heavily traveled highways." The section Principles of Route Selection in Cities in *Interregional Highways* states there should be "desirable coordination of highway transportation with rail, water, and air transportation."

addresses a concern posed by Borusyak and Hull (2023) that transportation infrastructure tends to non-randomly impact areas depending on its location relative to existing markets, which can be alleviated if counterfactual networks are specified. As an illustration of how highways often followed the historical road networks, in Figure 6a, I overlay the Interstate system on the historical network for the Boston area. As is evident, the two are closely aligned, but several historical roads were never re-built as Interstate highways, and these are the roads that serve as the control routes.

Additionally, maps on historical railroads, canals, steam-boat navigable rivers for the late 19th century, and natural features such as bodies of water, shores, and ports are retrieved from Atack (2015, 2016, 2017) and Lee and Lin (2017). These geographic characteristics are all considered factors that influenced highway placement.

Using these various maps, in Table 3 Panel B, I include log distance from the super-roads, railroads, canals, rivers, etc. as controls $\mathbf{X}_i$ in Equation 1. The estimating equation thus compares tracts near areas with historically higher levels of traffic to areas that ultimately received Interstates. While in previous research, historical routes have often been employed as instruments such as in Duranton and Turner (2012), past infrastructure influences subsequent economic development, leading to a violation of the exclusion restriction (Donaldson and Hornbeck, 2016). The strategy of this paper is to instead purge these past economic influences by including them as controls.

I find that relative to Panel A, the results are essentially unchanged—declines by highways in population, rents, and racial composition are of the same magnitude as in the specification without geographic controls. While somewhat surprising, distance from the CBD and redlining fixed effects were previously included, so some of the non-exogenous placement was already partially absorbed.

Planned Maps as Instrumental Variables – Even conditional on geographic features, final placement of highway routes may be biased due to local political factors, e.g. protests in high-income areas.

One strategy is to employ transportation plans designed before external influences as instruments. Transportation engineers were often indifferent to local socioeconomic conditions, so their planned maps yield cleaner variation (Rose and Mohl, 2012). I digitize plans created by state engineers for 100 metro areas in the 1955 *General Location of National System of Interstate Highways* (informally called the "Yellow Book") (Brinkman and Lin, 2022). These maps are consolidated with a 1947 plan from Baum-Snow (2007) I re-digitized at finer spatial scales to create a regional and metropolitan planned network. Examples for Atlanta, GA and Cleveland, OH are depicted in Figure B.4.

Still, transportation planners may not have been fully neutral in their route choices. In a second strategy, I construct an Euclidean ray network that connects cities in the planned maps with straight lines, similar to the "inconsequential units" approach of Chandra and Thompson (2000), Faber (2014), Morten and Oliveira (2018), and Banerjee et al. (2020). Neighborhoods coincidentally between cities are treated by the Interstate highway system, so the variation from the Euclidean ray network is more quasi-random than that from the plans. Figure 6b-2d plots the two instruments next to the Interstate network for Boston and shows they are often adjacent.

To test for instrument validity, I examine pre-trends in Table A.4 for changes between 1940 to 1950 or 1950 to 1960 depending on the timing of Interstate construction. As required for identification, the location of the planned routes and Euclidean rays is not correlated with demographic or economic changes before Interstate development, conditional on geographic controls. In Columns 8–9 of Table 2, there is also no cross-sectional correlation between the plans and rays with 1950 baseline characteristics after including controls. Finally, I estimate the strength of the first-stage both in log distance and with indicators for being within 1 mile of the highway or the plans and rays (described in greater detail in Appendix 3). As shown in Table A.5, F-statistics are all above 100.

Table 3 Panels C and D present IV estimates that instrument distance from highways with distance from the planned routes and Euclidean rays. I find that on the whole, the IV results concur with those measured with OLS as magnitudes are similar for all the outcomes of population (including separately by race), rental prices, and percentage White. Although coefficients are slightly larger with the ray instruments, the estimates do not differ greatly in their qualitative findings. As the results are consistent across OLS, with the controls, and instrumenting with the plans and rays, declines by highways can be interpreted as causal and a consequence of the local harms of Interstate highways.

In Appendix Figures B.7 and B.8, I present results that are similar in form to the long-differences specification where I non-parametrically plot changes over distance from the highway in 3 bins over distance from the CBD. These figures depict the curvature of the decline near Interstate roads, and given that the OLS coefficients are comparable to the IV estimates, they can also be interpreted as indicative of the costs of highways across neighborhoods.

3.2 Population Elasticities to Commuter Market Access

In the preceding section, I documented that population growth in suburban areas was considerable at the time. However, this finding cannot be attributed to treatment effects of the Interstate highway system as ongoing factors in American cities pushed for outmigration to the periphery e.g. school desegregation and rising crime as well as government policies described in the historical context (Cullen and Levitt, 1999; Boustan, 2010). Consequently, I analyze population responses to commuting benefits with a measure of connectivity that is microfounded by the spatial model presented later.

Commuter Market Access (CMA) summarizes commute frictions across locations with heterogeneity by race $r \in \{B, W\}$, for Black and White, and education $g \in \{L, H\}$, for less-educated and high-educated. In the set-up, residential tracts i are connected to workplaces j paying group-specific wages ω_{jgr} with commute costs d_{ijgr} that can be differentially affected by transportation infrastructure.

$$CMA_{igr} = \left(\sum_j \omega_{jgr} / {d_{ijgr}}^{\phi} \right)^{\frac{1}{\phi}}$$

CMA_{igr} for neighborhood i aggregates over workplaces, accounting for wages and commute costs with substitution elasticity ϕ, and increases when commute costs are reduced or wages are raised.

With this definition and the Decennial microdata, I measure how population L_{igr} responds to improvements in CMA from 1960 to 1970 separately by race through elasticity β_r. Within CMA, the parameter ϕ is set to 3 which falls in the middle of estimates from the literature (Ahlfeldt et al., 2015;

Morten and Oliveira, 2018; Severen, 2021). [16]

$$\Delta \log L_{igr} = \beta_r \, \Delta \log CMA_{igr} + \mathbf{X}_i \mu_r + \psi_{m(i)} + v_{igr}$$

Importantly, I control for *distance from the CBD* to exploit variation in concentric rings around cities, i.e. within the suburbs, and eliminate the correlation stemming from pre-existing suburbanization (as CMA increases the most in the suburbs). This variation is also relative to the comparison roads as following Borusyak and Hull (2023), I take a control function approach and construct CMA where the possible counterfactual shocks of large historical roads are converted into Interstate highways (shown in Figure B.6). Borusyak and Hull (2023) argues this re-centering addresses non-exogenous exposure where areas connected by highways are systematically different in the existing spatial network. Further included in the specification are the geographic controls of railroads, canals, etc. (plus distance to the CBD), all interacted with race, and city fixed effects.

Results are reported in Table 4 with standard errors clustered at the tract level and Conley (1999) standard errors in brackets. Population elasticities are in the range of 1.2–1.4 for the White population across Columns 1–3 where Column 1 includes only the geographic controls, Column 2 adds city fixed effects, and Column 3 includes the control variable of CMA where historical large roads are built as highways. Because the estimates do not vary greatly, non-exogenous exposure to the highway shock does not seem to be driving the findings. Elasticities for the Black population are in the range of 0.1–0.4 and are significantly lower than for White households; including city fixed effects in Column 2 leads the coefficient to no longer be significant. Consistent with the previous results, Black households do not respond to the benefits of highways, nor do they respond to the costs.

Instruments for Estimation – However, CMA changes not only from commute costs but also from wages. Employing only commuting variation from highways shock, I define the instrument $\mathbb{Z}^{HW}_{igr}$ $= \frac{1}{\phi}\,(\log \sum_j \omega_{jgr,1960}/{d^{HW}_{ijgr,1970}}^{\phi}) - \frac{1}{\phi}\,(\log \sum_j \omega_{jgr,1960}/{d_{ijgr,1960}}^{\phi})$ to exploit the panel nature of when Interstate segments were built with $d^{HW}_{ijgr,1970}$ containing the segments built by 1970. In this measure, wages are fixed to 1960 levels, thus ignoring endogenous adjustments that would bias estimation. Furthermore, I build two additional instruments where the change in commute costs comes from the plans or the Euclidean rays by replacing $d^{HW}_{ijgr,1970}$ with d^{Plans}_{ijgr} and d^{Rays}_{ijgr}.

I report IV estimates in Columns 7–9 and find that Kleibergen-Paap rk Wald and Cragg-Donald Wald F-statistics are all far above 10.[17] Compared to the Interstate instrument, elasticities with the planned and ray instruments for the White population are larger in magnitude, possibly because of

[16] $\Delta \log CMA_{igr} = \frac{1}{\phi}\,(\log \sum_j \omega_{jgr,1970}/{d_{ijgr,1970}}^{\phi} - \log \sum_j \omega_{jgr,1960}/{d_{ijgr,1960}}^{\phi})$. In this observed CMA measure, I use wage changes from 1960 to 1970 as well as commute cost changes. Since commute times are all computer generated, the change in commute costs comes from the addition of the segments of the Interstate highway system built between 1960 and 1970 as well as changes in mode of transport weights by race and education between 1960 and 1970 (all groups increase their car usage). The functional form for d_{ijgr} is detailed in the model section. $\omega_{jgr} = T_{jgr}(w_{jgr})^{\phi}$ is scaled wages as explained in the model section.

[17] In Appendix Table A.7, I report the first-stage regressions between CMA and corresponding measures with the Interstate, plan, and ray instruments. Because CMA includes both wage and commute cost changes, the first-stage coefficients on the planned and ray CMA instruments are lower than the first-stage coefficients reported for placement in Table A.5. However, when wages are fixed to 1960 levels, the first-stage coefficients for CMA look similar to those for placement as then the variation only comes from Interstate highway construction.

the negative selection of constructed routes. Coefficients are smaller with the instruments than with OLS, so the previously estimated reduced form elasticities may include responses to endogenous wage changes rather than solely the Interstate shock. Elasticities for the Black population are imprecisely estimated as standard errors are large. Yet overall, the population responses to CMA appear to result from causal treatment effects of Interstate highways.

Institutions as Mechanism – The previous descriptive evidence suggested that institutional barriers could be driving the varying responses by race, especially for the low mobility response of the Black population. I explore heterogeneity in how the population elasticities are shaped by this factor by including redlining by race fixed effects in Columns 4–6, which repeat Columns 1–3 and add the fixed effects. With variation within types of neighborhoods, population elasticities for White households are reduced in size to around 1, so some of their earlier estimated response to CMA improvements was across types e.g. by moving from redlined to non-redlined neighborhoods in the suburbs.

For Black households, elasticities are now in the range of 0.3–0.6 and statistically significant across all specifications (though still lower than for White households). Their dampened overall elasticities without redlining fixed effects mask how Black households respond to commuter access changes within redlined areas. These findings highlight how institutions create spatial frictions that inhibit the Black population from leaving centrally located, redlined neighborhoods for suburban, non-redlined neighborhoods. Importantly, the outlying neighborhoods are those where highway benefits are the largest and simultaneously, where costs are more muted.

Additional Results – (i) I estimate how the equilibrium outcomes of rents and racial composition respond to CMA in Table A.9 Columns 1 and 3. Consistent with the long differences, elasticities for rents and racial composition are smaller compared to the White population response and presumably play a smaller role in the welfare assessment. Because these equilibrium responses in turn affect population responses through feedback channels, I probe their importance for the population elasticities by controlling for rents and racial composition successively in Columns 2 and 4 as conducted in Adão et al. (2019). They do not appear to be a large determinant of the population responses to CMA. (ii) I construct two additional Borusyak and Hull (2023)-proposed CMA controls in Table A.8 Columns 1 and 2 where the planned routes and rays are converted into Interstates, and estimates remain unaffected when adding the controls. (iii) Pooling population elasticities by race is a fair approximation as in Table A.8 Column 4, I do not find substantial heterogeneity within race by education.

3.3 Discussion

While the reduced form evidence is informative for some of the responses to Interstate highways, it captures both direct and indirect effects through endogenous reallocation and adjustments in prices, racial composition, and wages. In the next section, I lay out a quantitative urban model that is rich enough to encompass all of these channels and carefully consider the various forces at play.

This model guides the derivation of estimating equations to measure the two direct highway impacts of commuting benefits and localized costs and to disentangle the sources of segregation that shape the population responses and welfare impacts. These specifications follow a structure reminiscent of the empirical evidence. Accordingly, parameter estimation captures many of the key empirical facts previously documented such as how mobility is highly differential by race and how the equilibrium objects of rents and racial composition influence location decisions.

Notably, the model's expressions for residential choice are instructive for why segregation arises both *in levels* and *in changes* in response to shocks. One set of parameters enables distinguishing the sources of segregation prior to Interstate construction as well as predicting how the Interstate highway system impacts segregation with accompanying feedback effects for welfare. The former application is especially important for understanding the role of institutional barriers in explaining the clustering of the Black population in central neighborhoods. This insight then determines inequality in impacts from highway policy because the spatial distribution of costs and benefits is highly differential across central versus suburban areas.

Chapter 2 – Unequal Access: Racial Segregation and the Distributional Impacts of Interstate Highways in Cities – Quantitative Model

4 A Quantitative Model of Cities

I develop a general equilibrium framework that extends previous advances from Allen and Arkolakis (2014), Ahlfeldt et al. (2015), and Tsivanidis (2022) by defining the forces of segregation and mapping each to components of classic urban models. To assess the distributional impacts of highway policy, the model features heterogeneity in race and education. Neighborhoods are linked via commuting networks, and transportation infrastructure lowers the costs of traveling on these networks which feeds into adjustments in the rest of the model. Using Decennial microdata, 25 cities enter the quantitative analysis, and I investigate the within-city impacts for each in a closed city set-up.

4.1 Model Features

Workers are differentiated by education $g \in \{L, H\}$ for less-educated and higher-educated groups and by race $r \in \{W, B\}$ for White and Black groups. Each city consists of neighborhoods indexed by $i = 1, \ldots, S$ and contains fixed population levels $\mathbb{L}_{gr}$ by education and race. Under the closed city assumption, no migration occurs across cities.

Workers – Individuals choose where to live (i) and work (j) depending on residential amenities, housing prices, wages, and commute costs after receiving separate idiosyncratic shocks for residences and workplaces. Worker o has Cobb-Douglas preferences over consumption of a numeraire good $c_{ij}(o)$ and residential floorspace $l_i(o)$ with β_{gr} share of income spent on the numeraire good.

Concretely, individual utility is represented as

$$\max_{c_{ij}(o), l_i(o)} \quad \frac{z_i(o)\epsilon_j(o)B_{igr}}{d_{ijgr}} \left(\frac{c_{ij}(o)}{\beta_{gr}}\right)^{\beta_{gr}} \left(\frac{l_i(o)}{1-\beta_{gr}}\right)^{1-\beta_{gr}}$$
$$\text{s.t.} \quad c_{ij}(o) + Q_i l_i(o) = w_{jgr}$$

and after utility maximization, indirect utility is expressed following

$$u_{ijgr}(o) = \frac{z_i(o)\epsilon_j(o)B_{igr}Q_i^{\beta_{gr}-1}w_{jgr}}{d_{ijgr}}$$

Several factors impact residential choice that correspond to the empirical responses observed previously. Differential sensitivity to prices Q_i is captured through non-homothetic preferences where β_{gr} differs by education and race, as each group has varying income levels. This formulation is a tractable approach taken by the literature to study sorting across space (Davis and Ortalo-Magné, 2011; Balboni et al., 2020; Diamond and Gaubert, 2021).[18] Variation in housing prices Q_i combined with group-specific consumption shares β_{gr} are then one factor leading to different residential choices

[18] Cobb-Douglas with varying shares β_{gr} allows for price changes to generate sorting but does not accommodate income changes leading to sorting compared to Stone-Geary. In Appendix 4.1, I derive an extension with Stone-Geary preferences.

across race and education. Living in location i also provides group-dependent amenities B_{igr} which further contribute to heterogeneous choices across residences.

On the workplace side, workers who choose location j receive wage w_{jgr} for their group, and these wages are determined by firms under a structure that I specify later on. Traveling from residence i to workplace j entails commute costs d_{ijgr} that reduce utility following the functional form $d_{ijgr} = (t_{ijgr})^{\kappa_{gr}}$ adopted from Heblich et al. (2020), and commute times t_{ijgr} are group-specific where the parameter κ_{gr} dictates how times translate into costs.

Beyond group-level factors, workers additionally have idiosyncratic preferences for residences $z_i(o)$ and for workplaces $\epsilon_j(o)$ that determine their choices. Residential idiosyncratic shocks $z_i(o)$ are drawn from a Frechet distribution $F(z_i(o)) = \exp(-z_i(o)^{-\theta_r})$ where θ_r is a shape parameter for the dispersion of shocks and how responsive individual choices are to changes in the attractiveness of residences, i.e. a substitution elasticity for mobility across residences. Following the reduced form evidence that Black and White households have differing responsiveness to CMA improvements, θ_r is heterogeneous by race.[19] Idiosyncratic workplace shocks $\epsilon_j(o)$ are likewise distributed Frechet from $F(\epsilon_j(o)) = \exp(-T_{jgr}\epsilon_j(o)^{-\phi})$ where ϕ is a workplace elasticity that governs the responsiveness of location choices to workplace changes. T_{jgr} is a group-specific scale parameter for the desirability of a workplace, e.g. through amenities beyond wages.

Departing from the canonical model of Ahlfeldt et al. (2015), I allow for separate residence and workplace shocks in the above specification.[20] The earlier estimated reduced form residential elasticities are substantially smaller in magnitude compared to estimates of ϕ found in the literature (Monte et al., 2018; Ahlfeldt et al., 2015; Severen, 2021). Further, the urban literature has also obtained lower residential elasticities (Mayo, 1981). Previous estimates leverage variation on the workplace side, so in this model, ϕ represents the substitution elasticity across workplaces.

With the model features for worker choice defined, I derive population levels across locations. Since $\epsilon_j(o)$ is distributed Frechet, conditional on living in i, the probability a worker works in j is

$$\pi_{j|igr} = \frac{T_{jgr}(w_{jgr}/d_{ijgr})^{\phi}}{\sum_s T_{sgr}(w_{sgr}/d_{isgr})^{\phi}} = \frac{T_{jgr}(w_{jgr}/d_{ijgr})^{\phi}}{\Phi_{igr}} \tag{2}$$

In this expression, when wages and workplace amenities relative to commute costs are higher, the probability a worker chooses to work in that location is greater. Because commute costs are scaled by the elasticity ϕ, I define the commuting elasticity ν_{gr} as combining the workplace elasticity ϕ with the commute cost parameter κ_{gr} such that $(d_{ijgr})^{\phi} = (t_{ijgr})^{\kappa_{gr}\phi} = (t_{ijgr})^{\nu_{gr}}$. The denominator Φ_{igr} is a transformation of the commuter market access (CMA) measure introduced earlier following

[19] In Appendix 4.1, I include an extension with a Nested Frechet structure to explicitly incorporate spatial frictions across types of neighborhoods for Black households. To allow for a more parsimonious framework, this feature is not included in the main model.

[20] The model implies residential choice follows the equation $L_{igr} = (B_{igr}\Phi_{igr}^{\frac{1}{\phi}}Q_i^{\beta_{gr}-1})^{\theta_r} / \sum_t (B_{tgr}\Phi_{tgr}^{\frac{1}{\phi}}Q_t^{\beta_{gr}-1})^{\theta_r}\mathbb{L}_{gr}$ where $\Phi_{igr} = \sum_j T_{jgr}(w_{jgr}/d_{ijgr})^{\phi}$. Under the null hypothesis that the residential elasticity and workplace elasticity are equivalent $\phi = \theta_r$, the coefficient λ_r in the estimating equation $\log L_{igr} = \lambda_r \log \Phi_{igr} + \gamma_{gr}$ should be approximately 1. Note that in Φ_{igr}, no assumptions are made on the value of ϕ because in the commuting gravity equation, $\nu_{gr} = \kappa_{gr}\phi$ is estimated directly from the data. See Appendix 5.2 for details on the values in Φ_{igr}. In Appendix Table A.8, I test for whether the elasticities to residential and workplace shocks should be the same value. I find the coefficient on Φ_{igr} is significantly less than one, suggesting the residential elasticity is in fact lower than the labor supply elasticity to workplaces.

$CMA_{igr} = \Phi_{igr}^{1/\phi}$ where for location i, higher wages w_{jgr} (with the scale parameter T_{jgr}) and lower commute costs d_{ijgr} increase CMA. Labor supply L_{Fjgr} is determined by aggregating over all the residential locations and the probability that each residence sends workers to workplace j following

$$L_{Fjgr} = \sum_i \pi_{j|igr} L_{igr} \tag{3}$$

where L_{igr} is the population of group gr workers at residence i.[21] The probability a worker lives in i follows a similar form using the properties of the Frechet distribution for residential shocks.

$$\pi_{igr} = \frac{\left(B_{igr} CMA_{igr} Q_i^{\beta_{gr}-1}\right)^{\theta_r}}{\sum_t \left(B_{tgr} CMA_{tgr} Q_t^{\beta_{gr}-1}\right)^{\theta_r}} \tag{4}$$

Neighborhoods with greater group-specific amenities, higher CMA, and lower housing prices attract a higher share of each group. The residential population in i combines the probability above with the total population of a group in a city $\mathbb{L}_{gr}$. Importantly, this expression for residential population is one of the key equations for determining segregation.

$$L_{igr} = \pi_{igr} \mathbb{L}_{gr} \tag{5}$$

Sources of Segregation – With the above factors characterizing residential choice, sorting by race and education can arise from (1) group-specific commuter market access, (2) housing prices which, while not group-specific, are valued differentially by race and education,[22] and (3) group-specific amenities. How amenities in particular vary across space will be a central determinant of segregation in levels and in changes, especially by race.

Racial Preferences in Endogenous Amenities – Amenities contain both a fundamental component b_{igr} and an endogenous component through racial composition L_{iW}/L_i that evolves with population flows.

$$B_{igr} = b_{igr} \underbrace{(L_{iW}/L_i)}_{\text{Pct White}}^{\rho_r} \tag{6}$$

The importance of racial composition in amenities is governed by ρ_r and differs by race, allowing White and Black households to have heterogeneous preferences from several possible explanations. Composition may affect the variety of amenities or public goods (e.g. school quality), with each group desiring distinct types (Diamond, 2016; Almagro and Dominguez-Iino, 2020). Prejudice or biased beliefs, often considered to be homophilic, may create taste-based reasons to prefer living with a particular racial group (Becker, 1971). The social environment can influence these preferences; reported White attitudes on race have evolved after the Civil Rights movement, and Black attitudes may be shaped by animosity from White households (Bobo et al., 2012). This paper does not measure which

[21] Expected income at location i can be computed by weighting wages with the probability of commuting to workplace j. $\bar{w}_{igr} = E[w_{jgr}|i] = \sum_j \pi_{j|igr} w_{jgr} = \sum_j \frac{T_{jgr}(w_{jgr}/d_{ijgr})^{\phi}}{\sum_s T_{sgr}(w_{sgr}/d_{isgr})^{\phi}} w_{jgr}$.

[22] Balboni et al. (2020) have a segmented housing construction sector where prices are group-specific. In this setting, prices are not sufficiently different by race after accounting for quality controls and neighborhood fixed effects to merit segmented housing. See Table A.1.

reasons rationalize the endogenous amenity parameters, but they are important for understanding equilibrium choices as any direct changes in racial composition are reinforced by preferences.

Institutional Exclusion in Fundamental Amenities – Exclusionary practices segregated black households to certain neighborhoods. I model these institutional factors in the fundamental component of amenities where they do not adjust over time, and consequently, they are also invariant to highway policy.[23]

There is limited evidence exclusion occurred along class divisions within race or that White households experienced obstacles, so I consider only that institutions prohibited Black households from the most desirable neighborhoods. I set constraints following the redlining maps: Black households face barriers outside of redlined neighborhoods, which I refer to as the set $\mathbf{R}$, through a wedge in fundamental amenities E_i that relatively lowers the desirability of the exclusive neighborhood. As Black households often experienced violent threats when living in predominantly White areas, this institutional wedge has normative implications for welfare.

As amenities are identified up to scale within each group, the wedge should be interpreted as a relative difference across groups and across locations. For simplicity, assume there are two representative neighborhoods $n \in \mathbf{R}, m \notin \mathbf{R}$ with similar exogenous amenities ignoring institutional factors. There exists a constant wedge $\bar{E}$ within the exclusive, non-redlined neighborhood and no wedge inside the redlined neighborhood. Accordingly, the fundamental amenity terms follow

$$\frac{b_{mB}}{b_{mW}} = \bar{E} = E_m < 1, \quad \frac{b_{mB}}{b_{nB}} = \bar{E} = E_m < 1, \quad \frac{b_{mB}}{b_{mW}} = 1, \quad \frac{b_{mW}}{b_{nW}} = 1, \quad \text{if } n \in \mathbf{R}, m \notin \mathbf{R}$$

In practice, within the two types of neighborhoods, there may be additional heterogeneity in the magnitude of the wedge. Locations that contain zero Black population would have a multiplier E_i equal to zero as barriers are so strong as to prevent any Black households from entering. More generally, the fundamental amenities of Black households are relative to White households following

$$b_{igB} = b_{igW} \times E_i \tag{7}$$

The race and location specific wedge has the same empirical implications as in Cutler, Glaeser, and Vigdor (1999) where institutions are difficult to distinguish from racial preferences. Any desire to self-segregate by Black households would appear as higher amenity values for redlined neighborhoods, which is isomorphic to the amenity drop from institutional barriers. A central aim of the estimation will then be disentangling the endogenous and fundamental components of amenities.

Firms – As workers alter their labor supply to workplaces in response to reductions in commute costs, wages are determined in equilibrium by firms. While adjustments at firms are not a central theme of the empirical evidence or the question of the paper, I include this feature to close the model and allow for a comprehensive assessment of the impacts of Interstate highways. In the counterfactual exercises,

[23] Exclusion may be a function of the proportion of the neighborhood that is White i.e. through endogenous institutions, but other institutions are codified into law and persistently invariant to the racial composition of the neighborhood. Aaronson et al. (2021) find a border discontinuity in racial composition at redlining borders even as non-redlined areas became more racially diverse over time. Zoning is an endogenous exclusionary barrier that arises as neighborhood racial composition changes, as studied in Lee (2022), Song (2022), and Krimmel (2022).

I probe its importance for welfare by shutting down firm adjustments in wages and housing.

Across workplaces, there are representative firms with constant returns to scale production so that demand by firms translates into demand at each workplace. Perfectly competitive firms produce varieties with Cobb-Douglas technology over labor and commercial floorspace following $Y_j = A_j N_j^{\alpha} H_{Fj}^{1-\alpha}$, where α is the share of labor and A_j is a Hicks-neutral productivity shock. Combining heterogeneous workers, labor N_j is a CES aggregate over education where workers of different education levels are imperfect substitutes (Katz and Murphy, 1992; Card, 2009). N_{jg} is further a CES aggregate of different racial groups.

$$N_j = \left(\sum_g \alpha_{jg} N_{jg}^{\frac{\sigma^g - 1}{\sigma^g}} \right)^{\frac{\sigma^g}{\sigma^g - 1}} \quad \text{with} \quad N_{jg} = \left(\sum_r \alpha_{jgr} L_{Fjgr}^{\frac{\sigma^r - 1}{\sigma^r}} \right)^{\frac{\sigma^r}{\sigma^r - 1}}$$

This nested-CES structure accommodates imperfect substitutability across race following evidence from Boustan (2009) that Black workers are closer substitutes to each other than to White workers. Imperfect substitutability can arise from occupational segregation that prevents workers of one race from switching into an occupation that is predominantly of another race or from unobserved skill gaps, even conditional on education (Higgs, 1977).[24]

Within education, locations employ workers from each race at varying intensities α_{jgr}. This spatial heterogeneity incorporates how firms in the central city may have different demands for Black workers compared to firms in the suburbs e.g. as a result of differences in discrimination across space (Holzer and Reaser, 2000; Miller, 2018). Moreover, it generalizes the labor aggregate structure of Tsivanidis (2022) which restricts group-specific wage differentials to result only from industry mix varying across locations due to the lack of group by workplace wage data.

Firm profit maximization then leads to the following equations for labor and commercial floorspace demand with the corresponding wage indices.

$$L_{Fjgr} = \left(\frac{w_{jgr}}{\alpha_{jgr} \omega_{jg}} \right)^{-\sigma^r} \left(\frac{\omega_{jg}}{\alpha_{jg} W_j} \right)^{-\sigma^g} N_j \quad \text{s.t.} \quad W_j = \left(\sum_g \alpha_{jg}^{\sigma^g} \omega_{jg}^{1-\sigma^g} \right)^{\frac{1}{1-\sigma^g}} \quad \omega_{jg} = \left(\sum_r \alpha_{jgr}^{\sigma^r} w_{jgr}^{1-\sigma^r} \right)^{\frac{1}{1-\sigma^r}} \tag{8}$$

$$H_{Fj} = \left(\frac{1-\alpha}{Q_j} A_j \right)^{1/\alpha} N_j \tag{9}$$

The zero-profit condition from perfect competition combined with profit maximization leads to the subsequent condition for commercial rental prices, which rise when productivity is high and wages are low. Firms thus aim to locate in more productive, cheaper, and lower-wage areas.

$$Q_j = (1-\alpha) \left(\frac{\alpha}{W_j} \right)^{\frac{\alpha}{1-\alpha}} A_j^{\frac{1}{1-\alpha}} \tag{10}$$

Agglomeration – Within productivity of locations, the term A_j contains a fundamental component a_j that does not vary with equilibrium outcomes and an endogenous component representing agglom-

[24]The average Black worker at this time attended lower quality schools, especially in the segregated South, compared to the average White worker which would lead equivalent years of schooling to translate into different skill levels (Margo, 2007).

eration economies in density (L_{Fj}/K_j). In the definition of density, L_{Fj} is total employment, K_j is the land area, and γ^A is the strength of agglomeration.

$$A_j = a_j \left(L_{Fj}/K_j\right)^{\gamma^A} \tag{11}$$

Many theories underlie why density affects productivity e.g. via knowledge spillovers, labor market pooling, or input-output linkages (Rosenthal and Strange, 2004; Ellison et al., 2010). Transportation infrastructure in conjunction with agglomeration can reallocate economic activity as in Faber (2014), Heblich et al. (2020), and Baum-Snow (2020) with racially disparate effects as studied by Miller (2018). Thus, the model allows for this feature as another channel of highway impacts.

Housing – Given that empirically, housing prices adjusted with highway construction, I allow for a housing construction sector that responds elastically to changes in demand from both residences and workplaces. In each location, there is H_i amount of floorspace that is allocated endogenously across residential versus commercial uses where θ_i is the share for residential use. Residential floorspace demand aggregates across the housing expenditures of each group Exp_{igr} in residential location i so that H_{Ri} is determined following

$$H_{Ri} = \theta_i H_i = \sum_{g,r} \frac{Exp_{igr}}{Q_i} \quad \text{where } Exp_{igr} = (1-\beta_{gr})\overline{w}_{igr} L_{igr} \tag{12}$$

Commercial floorspace demand comes from firm optimization in Equation (9), and with the two expressions for residential and commercial floorspace demand, the allocation across uses θ_i and total floorspace demand $H_i = H_{Ri} + H_{Fi}$ are then determined for land market clearing.

To parameterize how housing is supplied elasticity, I follow the literature where the housing production function is $H_i = K_i^{\mu} M_i^{1-\mu}$ with M_i as capital at universal price p and K_i as land at price r_i (Epple et al., 2010; Combes et al., 2021). The implied supply curve is then

$$H_i = \left(\frac{1-\mu}{\mu}\right)^{\frac{1-\mu}{\mu}} K_i Q_i^{\frac{1-\mu}{\mu}} \tag{13}$$

Welfare – Finally, welfare defined as U_{gr} aggregates over all of the residential locations accounting for amenities, commuter access, and rental prices.

$$U_{gr} = \left(\sum_i \left(B_{igr} CMA_{igr} Q_i^{\beta_{gr}-1}\right)^{\theta_r}\right)^{1/\theta_r} = \left(\sum_i \left(B_{igr} \underbrace{\left(\sum_j T_{jgr}(w_{jgr}/d_{ijgr})^{\phi}\right)^{\frac{1}{\phi}}}_{CMA_{igr}} Q_i^{\beta_{gr}-1}\right)^{\theta_r}\right)^{1/\theta_r} \tag{14}$$

4.2 Impacts of the Interstate Highway System

Commuting benefits of highways lead to declines in bilateral times t_{ijgr} in the commute cost function $d_{ijgr} = t_{ijgr}^{\kappa_{gr}}$. These reductions improve commuter access differentially across locations depending on which bilateral pairs are connected by highways and the scaled wages that are paid at workplaces.

Localized costs of highways appear as a scaling factor in fundamental amenities b_{igr} and decay over distance $DistHW_i$ at the rate of parameter η (Brinkman and Lin, 2022). At $DistHW_i = 0$, fundamental amenities are discounted by $1 - b^{HW}$, and remaining amenity shifters are contained in $\bar{b}_{igr}$.

$$b_{igr} = \bar{b}_{igr}(1 - b^{HW}\exp(-\eta DistHW_i)) \tag{15}$$

Residential Choice Expression – In summary, the Interstate system generates changes in the

- *Fundamentals* of (1) commute times between places and (2) amenities by highways

which through the general equilibrium system of equations lead to adjustments in the

- *Equilibrium objects* of endogenous amenities in racial composition, housing prices, and wages

The impact of these changes on population is characterized by the residential shares expression in Equation (25), which I lay out for the Black population to display where institutional barriers E_i appear.

$$\begin{aligned} L_{igB} &= \Big(\underbrace{B_{igB}}_{\text{Amenities}} \times \underbrace{CMA_{igB}}_{\text{Comm Access}} \times \underbrace{Q_i}_{\text{Prices}}{}^{\beta_{gB}-1}\Big)^{\theta_B} \mathbb{L}_{gB} U_{gB}^{-\theta_B} \\ L_{igB} &= \Bigg(\bar{b}_{igW} \underbrace{E_i}_{\text{Institutions}} \underbrace{(1 - b^{HW}\exp(-\eta DistHW_i))}_{\text{Local Costs}} \underbrace{(L_{iW}/L_i)}_{\text{Pct White}}{}^{\rho_B} \\ &\quad \times \Big(\sum_j T_{jgB}(\underbrace{w_{jgB}}_{\text{Wages}})^{\phi} * (\underbrace{t_{ijgB}}_{\text{Commute Times}})^{-\kappa_{gr}\phi}\Big)^{\frac{1}{\phi}} \times \underbrace{Q_i}_{\text{Prices}}{}^{\beta_{gB}-1}\Bigg)^{\theta_B} \mathbb{L}_{gB} U_{gB}^{-\theta_B} \end{aligned} \tag{16}$$

Residential choice across neighborhoods evolves with the benefits and costs of Interstate highways as well as with equilibrium adjustments across residences and workplaces. In parameter estimation, this expression reappears to guide the specification of the structural equation and appropriate controls.

Note that institutional barriers E_i are contained in fundamental amenities where they do not adjust with highway impacts. If much of the cross-sectional variation in where Black households live appears in fundamentals, only large shocks can alter the degree of segregation. Although Interstate highways impacted cities immensely, they may not be sufficiently large to affect Black residential locations, in line with the observed low Black mobility.

Impacts to Welfare in Equilibrium – Welfare changes are tightly tied to the population responses and can be expressed in exact-hat algebra form $\hat{x} = x'/x$ to show the dependence on initial allocations. The change in welfare $\hat{U}_{gr} = U'_{gr}/U_{gr}$ follows

$$\hat{U}_{gr} = \Bigg(\sum_i \pi_{igr}\Big(\underbrace{\hat{b}_{igr}}_{\text{Fund Amen}} \times \underbrace{(\hat{L}_{iW}/\hat{L}_i)}_{\text{Pct White}}{}^{\rho_r}\Big(\sum_j \pi_{j|igr}(\underbrace{\hat{w}_{jgr}}_{\text{Wages}})^{\phi} \times (\underbrace{\hat{t}_{ijgr}}_{\text{Commute Times}})^{-\kappa_{gr}\phi}\Big)^{\frac{1}{\phi}} \underbrace{\hat{Q}_i}_{\text{Prices}}{}^{\beta_{gr}-1}\Big)^{\theta_r}\Bigg)^{1/\theta_r} \tag{17}$$

In this expression, welfare impacts of highways are determined by the (1) initial spatial distribution of groups across locations in π_{igr} and $\pi_{j|igr}$, (2) changes in *fundamentals* of amenities and commute times and changes in *equilibrium outcomes* of racial composition, prices, and wages, and (3) elasticities to residential and workplace shocks. Given $\theta_N < \theta_W$, Black households respond less to any residential changes with subsequent impacts on welfare. Furthermore, the lower residential elasticities of the Black population imply that their initial residential locations, which may be determined heavily by fundamentals, are especially important for the incidence of the impacts of highway infrastructure.

4.3 General Equilibrium

Definition 1. Given the model's parameters $\{\beta_{gr}, \theta_r, \kappa_{gr}, \phi, \alpha, \alpha_{jg}, \alpha_{jgr}, \sigma^g, \sigma^r, \mu, \rho_r, \gamma^A\}$, city populations by education and race $\{\mathbb{L}_{gr}\}$, and location characteristics $\{T_{jgr}, t_{ijgr}, b_{igr}, a_j, K_i\}$, the general equilibrium is represented by the vector of endogenous objects $\{L_{igr}, L_{Fjgr}, Q_i, \theta_i, w_{jgr}, B_{igr}, A_j, U_{gr}\}$ determined by the following equations:

1. *Residential populations* in each neighborhood (25)
2. *Labor supply* at each workplace (3)
3. *Housing demand* from residences and firms (12) + (9)
4. *Housing supply* from the construction sector (27)
5. *Zero profit and profit maximization* by firms (8)
6. *Endogenous amenities* from racial composition (6)
7. *Endogenous productivity* from agglomeration (11)
8. *Closed City* where $\sum_i L_{igr} = \mathbb{L}_{gr}$

The equilibrium defined above has many sources of spillovers. The most immediate are through endogenous amenities and productivity from racial composition and agglomeration. Additional spillovers emerge through inelastic land generating a congestion force in housing supply and the idiosyncratic preferences of individuals creating dispersion forces. As the wages of each group depend on the labor supply of other workers, there are productivity spillovers across groups at workplaces. I next present sufficient conditions for the uniqueness of equilibria in the presence of spillovers.

Proposition 1. *There exists a unique equilibrium when the parameters $\{\beta_{gr}, \theta_r, \phi, \alpha, \sigma^g, \sigma^r, \mu, \rho_r, \gamma^A\}$ satisfy the condition $\rho(A) < 1$ where A is a matrix of elasticity bounds on the economic interactions across endogenous equilibrium outcomes and $\rho(A)$ is the spectral radius.*

Proof. See Appendix 6.4 □

I follow Allen et al. (2022) where I rewrite the equilibrium conditions as a set of H types of economic interactions conducted by the set of N heterogeneous agents. I then construct the $H \times H$ matrix of the uniform bounds of the elasticities on the strength of economic interactions. The equilibrium system falls under a constant elasticity form that is commonly used in spatial economics. Building on Tsivanidis (2022), I reformulate the CMA measures as solutions to a system of equations in residential and workplace populations and commute costs. With these conditions on model parameters, I derive theory-consistent equations to estimate parameter values in the next section.

5 Parameter Estimation and Model Inversion

The steps for estimation and inversion are intertwined, so I summarize the overarching goals before presenting the estimating equations.

5.1 Estimation and Inversion Overview

Parameter Estimation – The parameters can be categorized into the two strands of the paper: (1) the direct benefits and costs of Interstate highways through increases in commuting connectivity and local harms near routes, and (2) the sources of segregation. Although additional parameters are needed to close the model, the focus of estimation is on the two main strands.

To measure the benefits, a key initial step is estimating the "gravity" equation for how commute flows relate to commute times by race and education. I obtain commuting elasticities $\nu_{gr} = \kappa_{gr}\phi$, which combine the commute cost parameter κ_{gr} with the workplace substitution elasticity ϕ, and ν_{gr} enters into the CMA measures to determine residential location decisions. Within the commuting elasticity, the workplace elasticity is assigned from the literature to $\phi = 3$ following studies with settings similar to this paper.[25] The commute cost parameter is then $\kappa_{gr} = \nu_{gr}/\phi$.

Commuting elasticities on hand, I construct several instruments to estimate some of the sources of segregation, residential elasticity θ_r and racial preferences ρ_r, by exploiting quasi-random variation from the highway shock. Building on the reduced form empirical equations, CMA improvements directly affect residential attractiveness, providing variation for θ_r. They indirectly alter racial composition as the population responds in race-specific ways to CMA, providing variation for ρ_r. Lastly, housing price sensitivity β_{gr} is calibrated using the Consumer Expenditure Surveys (CEX).

With this set of parameters, I invert the model to retrieve fundamental amenities as the residual component of population choices unexplained by characteristics such as rental prices, racial composition, or commuting access. I then project the inverted fundamental amenities over distance from Interstate highways to measure the local costs in non-parametric bins.

In Section 7, I return to the fundamental amenity term to investigate the role of institutional barriers. As fundamental amenities remove the sources of segregation that arise from rents or racial preferences, they represent other factors that drive location choice. Importantly, only sources of segregation that evolve in equilibrium (prices, racial composition) are necessary for evaluating the impacts of highways. Institutions in fundamentals as time-invariant terms are not prerequisites, but they are explored as a relevant mechanism later on.

Model Inversion – In tandem with parameter estimation described above, model inversion occurs in the background to acquire components that enter estimation. Inversion uses the set of parameters (partially estimated, partially from the literature as described in later sections) to map observed data on residential and workplace populations, commute times, housing prices, and wages to productivity

[25] The elasticity ϕ has been estimated in various contexts and ranges from 6.8 in Ahlfeldt et al. (2015) during the division of Berlin, 1.9 in Morten and Oliveira (2018) with highway expansion in Brazil, 3.3 in Monte et al. (2018) with commuting data in the U.S, and 2.18 in Severen (2021) with development of the Los Angeles Metro Rail.

and residential amenities $\{A_i, B_{igr}\}$. During this process, several location characteristics $\{T_{jgr}, K_i, \theta_i\}$ are also inferred. Some such as T_{jgr} are important inputs for parameter estimation.

Using the commuting equation for labor supply in Equation (3) and following the iterative procedure of Allen and Arkolakis (2014), I invert for workplace factors $\omega_{jgr} = T_{jgr}(w_{jgr})^{\phi}$ which combines wages with the scale parameter T_{jgr}.[26] Aggregating the workplace factors into the CMA measure and combining CMA with rents, racial composition and the estimated parameters θ_r and β_{gr}, I infer amenities up to a scaling factor with the mapping provided by the residential share in Equation (24).

Not central to the paper but necessary for the general equilibrium system, I calibrate α_{jg}, α_{jgr} in the production function using the CES structure for labor demand, wages by race and education, and the elasticities of substitution by race and education σ_g, σ_r. From the zero profit condition in Equation (10), productivity is determined by housing prices and W_j, the CES wage index over education and race. As the workplace data is at the POW Zone unit, I assume that the distribution of economic activity across tracts is uniform within the POW Zone and invert for tract-level productivity with tract-level housing prices. Finally, housing supply H_i (and relatedly land used in housing production K_i) and the allocation across residential and commercial uses θ_i are recovered from the conditions on residential and commercial demand in Equations (9) and (12). Details are provided in Appendix 5.2.

Summary – The key parameters estimated in the subsequent sections pertain to the

Direct benefits and costs of highways

- Commuting elasticities: ν_{gr}, local costs: b^{HW} in amenities and η for rate of decay over distance

Sources of segregation

- Residential elasticity: θ_r, racial preferences: ρ_r, and price sensitivity: β_{gr}

Secondary parameters for the equilibrium framework are calibrated or taken from the literature.

5.2 Gravity Equation for Commuting Elasticity

With the commute shares expression from Equation (23) and the functional form of commute costs as $d_{ijgr} = t_{ijgr}^{\kappa_{gr}}$, I arrive at the following gravity equation to estimate the commuting elasticities.

$$\log \pi_{j|igr,t} = \underbrace{\gamma_{jgr,t}}_{\log \omega_{jgr}} + \underbrace{\gamma_{igr,t}}_{\log \Phi_{igr}} - \underbrace{\nu_{gr}}_{\kappa_{gr}\phi} \log t_{ijgr,t} + \epsilon_{ijgr,t}$$

Location by year fixed effects $\gamma_{jgr,t}$ and $\gamma_{igr,t}$ account for factors that are workplace-specific (scaled wages ω_{jgr}) and residence-specific (transformed commuter access Φ_{igr}) in each year. The error term $\epsilon_{ijgr,t}$ captures remaining factors outside of the model or mismeasurement in commute times. For estimation, commute flows from the 1960 and 1970 Censuses are pooled together, and commute times are

[26] This process is isomorphic to taking the workplace fixed effect from the "gravity" equation estimated later. Unlike Ahlfeldt et al. (2015) which lacks wage data, making inversion a necessity to infer determinants of workplace choice, I observe wages by race and education at employment locations. Confirming evidence from Severen (2021) and Kreindler and Miyauchi (2022), wages do not fully determine workplace location decisions. The scale parameter T_{jgr} is another determinant and captures variation in the size of the POW Zone units. T_{jgr} can also be affected by workplace amenities that are differential by race and education across locations e.g. through discrimination beyond in wages.

generated by overlaying segments of Interstate routes constructed by each decade on the historical urban road network. Bilateral variation in commute times then identifies the elasticity.

Splitting the data by race and education leads some bilateral pairs to have zero flows, which happens often for the Black population (11 percent of the sample). To reduce sparsity, I aggregate the residential tracts up to the Place of Work Zone, leading estimation to be at the POW Zone by POW Zone by year level with standard errors clustered at the POW Zone by POW Zone level.

In Table 5 Panel A, I find that less-educated groups, both White and Black, tend to have higher elasticities compared to the higher-educated, in line with findings from Tsivanidis (2022) in the context of Bogota. Parameter values for Black workers are lower than values for White workers, suggesting that Black households consider commute times less in their commuting decisions. These values are similar to those in Heblich et al. (2020) of -4.90, estimated with commuting data from 19th-century London. Instrumenting times with the planned routes and rays for 1970 in Panels B and C, respectively, does not greatly alter the results (first stages are reported in Appendix Table A.11), and the estimates from Panel C are the preferred values used for counterfactual analysis later on.

Additional Results – Even with the aggregation, some bilateral pairs continue to have zero counts for the Black population. In addition to estimating the log-log specification above, I conduct the robustness checks suggested by the trade literature in Head and Mayer (2014) and estimate the commuting elasticity with Poisson Pseudo Maximum Likelihood (PPML) following Silva and Tenreyro (2006) to address sparsity. For the Black population, the commuting elasticity rises slightly with PPML estimates in Panel D. For White workers, their elasticities are lowered with the PPML estimator although the observation count only increases a small amount.[27] Overall, the pattern remains quite similar.

As the historical road network may also have been endogenously placed, I instrument commute times with Euclidean distance in Appendix Table A.11 Panel C (with the first stage in Panel D). The ordering of the elasticities across groups remains the same, although the magnitude is higher. Additionally, I report the coefficient for the reduced form regression of commuting share on Euclidean distance in Panel E and find that Black elasticities are about 20% lower because they commute via slower modes of transport. Lastly, In Appendix Table A.12, I instrument the PPML estimates via a control function approach following Wooldridge (2015) and bootstrap standard errors. These values concur with the previous PPML estimates, so instrumenting is not crucial.

5.3 Residential Elasticity and Racial Preferences as Endogenous Amenities

I now estimate the residential elasticity θ_r and racial preferences ρ_r following the structural equation below which relates population flows to changes in CMA and changes in racial composition.

$$\Delta \log L_{igr} = \theta_r \Delta \log CMA_{igr} + \underbrace{\tilde{\rho}_r}_{\theta_r \rho_r} \Delta \log \underbrace{(L_{iW}/L_i)}_{\text{Pct White}} + \mathbf{X}_i \beta_{gr} + \underbrace{\gamma_{m(i)gr}}_{\Delta \log(\mathbf{L}_{gr}/U_{gr}^{\theta_r})} + \alpha_{red(i)} + \underbrace{\epsilon_{igr}}_{\theta_r \Delta \log b_{igr}} \tag{18}$$

[27] Accounting for zeros in bilateral pairs still leaves the observation count of the less-educated Black population at 22000 below that of the higher-educated White population as there are cases of residential and workplace units without any Black workers, and PPML only adjusts for *bilateral* counts of zero.

The first difference is between 1960 and 1970, and estimation uses the Census microdata. Similarly to the reduced form elasticities, this equation measures population responses to CMA improvements where CMA contains the inverted values for scaled wages ω_{jgr} and the commuting elasticities ν_{gr}. Included in the specification above are several controls $\mathbf{X}_i$, all interacted with race and education, city by group fixed effects, and redlining fixed effects.

This estimating equation is theory-consistent when the appropriate set of controls is specified. Combining the residential share expression (24), endogenous amenities in (6), and localized highway costs in (15) yields the above specification where the set of controls accordingly contains terms for changes in rental prices and bins in distance from routes for the localized costs (all interacted with race and education). Rents and highway costs must be controlled for as they are correlated with changing residential choice, commuter access improvements, and changing demographics.[28]

Unlike past studies that focus on one city, this paper analyzes multiple cities pooled together. I include city by group fixed effects to capture factors such as average welfare U_{gr} and aggregate population $\mathbf{L}_{gr}$, absorbing migration across cities and leading the variation to come from within cities. Importantly, this equation also contains redlining fixed effects to compare within neighborhood types (redlined vs. non-redlined) as in the previous empirical evidence, I found that Black households faced spatial frictions across types. The coefficient on racial composition thus more closely represents racial preferences since within redlined areas, institutional frictions are less relevant.

To obtain cleaner variation in commuter access changes, I augment the model-informed controls with geographic controls for distance from the central business district, large roads, railroads, canals, rivers, lakes, shores, and ports as well as the Borusyak and Hull (2023)-proposed control for CMA. I additionally incorporate socioeconomic status controls for average income, home values, percentage of residents who are high school graduates, bottom income quintile, and top income quintile. Consequently, the racial preferences parameter omits preferences for socioeconomic status.

Main Results – Displayed in Table 6 Column 1 are OLS estimates with the geographic controls and the base set of model-informed controls on rents and local costs. Standard errors are clustered at the tract-level with Conley (1999) standard errors in brackets. I measure that the residential elasticity is 0.802 (0.183) for White households and 0.119 (0.172) for Black households, so in line with the descriptive evidence, Black households are far less responsive to highway commuting impacts.

White households have strong preferences for living in neighborhoods that are more White with an estimated value of $\tilde{\rho}_W = 1.049$ (0.024) while Black households have weaker preferences against living in more White neighborhoods with $\tilde{\rho}_W = -0.364$ (0.055). Adding the demographic controls in Column 2 lowers how much Black households care about racial composition, implying that some of the earlier estimate was driven by the correlation between race and socioeconomic characteristics. However, the preference elasticity estimate for White households is unchanged, so their preferences are very strongly related to race rather than other status variables.

[28] As commuter access increases closer to highways, it is correlated with the localized costs of highways and it may not be immediately clear there are enough sources of variation for identification. By controlling for the distance bins, identification of the effects of commuter access comes from comparing neighborhoods by highways that experience minimal commuter access changes, e.g. closer to the central city, to neighborhoods by highways that experience large commuter access changes, e.g. in the suburbs.

Instruments for Estimation – Bias in the above estimates arises from the error term corresponding to structural residuals of fundamental amenities, i.e. $\epsilon_{igr} = \theta_r \Delta \log b_{igr}$. Group-specific amenities, for example through practices targeted to particular populations, affect residential locations and are correlated with changing racial composition. Historically, speculative realtors steered Black families into transitioning areas, and groups such as the Southtown Planning Association in Chicago barred Black households from prospering White neighborhoods (Hirsch, 1983). These events create mechanical correlations between changing composition and Black population flows, and by not neatly fitting into the model framework, they appear in the residual to produce an endogeneity problem.

To exogenously shift racial composition (percent White), I construct three sets of instruments. The first uses Hausman (1996) or Berry et al. (1995) style instruments from the industrial organization literature where changes in other markets (i.e. other neighborhoods) shift local demand. CMA improvements and rental price changes in neighborhoods 3-5, 5-10, and 10-15 miles away serve as such shifters. As price changes may not be fully exogenous, the second set of instruments only uses variation from the highway shock following the 3-step approach of Davis et al. (2019).

In an initial step, I estimate a simpler version of Equation (18) that represents the model without endogenous amenities. I then solve the pared-down model with the estimated residential elasticities, which I find are heterogeneous by race, to predict shifts in racial composition from the highway shock. Central areas that White families migrate from are those with the largest predicted changes in percent White, and in the final step, I use the predictions for racial composition as instruments (while keeping the Hausman instruments from CMA improvements). Lastly, following the same intuition of race-specific responses to CMA changes, the group-specific CMA measures serve as a third set.

To address the non-exogenous placement of Interstate routes, changes in commute times assume the planned maps or Euclidean rays are built rather than the Interstate network. Concretely, $\mathbb{Z}_{igr}^{Plans}$ is the planned CMA instrument previously presented in Section 3.2, and $\mathbb{Z}_{igr}^{Rays}$ is the corresponding measure for the Euclidean rays. More details on all instruments are provided in Appendix 5.3.

The two variables of interest are $\{\Delta \log CMA_{igr}, \Delta \log(L_{iW}/L_i)\}$ where the first variable of CMA changes uses $\mathbb{Z}_{igr}^{PLans}, \mathbb{Z}_{igr}^{Rays}$ as instruments. The second variable of racial composition changes uses the above three sets of instruments. Consistent estimation of parameters relies on the following orthogonality condition, where $\mathbf{Z}_{igr}$ contains all of the instruments.

$$\mathbf{E}[\epsilon_{igr} \times \mathbf{W}_{igr}] = \mathbf{E}[f(\theta, \rho_r) \times \mathbf{W}_{igr}] = 0$$

The matrix $\mathbf{W}_{igr}$ includes the city by group fixed effects, redlining fixed effects, controls $\mathbf{X}_i$, and excluded instruments $\mathbf{Z}_{igr}$.

In Table 6 Columns 3–8, I report IV estimates across the three sets of instruments. Residential elasticities for White households are in the range of 0.420 (0.185) to 0.918 (0.161) and are higher than those for Black households, except when using the Davis et al. (2019) instruments. Black residential elasticities are challenging to estimate precisely because of large standard errors and the lower point estimate. The most stable estimate across all specifications is racial preferences for White households, which ranges from 1.016 (0.154) to 1.239 (0.066) and is highly statistically significant. Interestingly, I now find that Black racial preferences are fairly weak with point estimates in the range of -0.0418

(0.137) to -0.0973 (0.048), many not statistically significant. These results suggest the previous findings on preferences for Black households stem from correlations with changing unobserved amenities.

Lastly, at the bottom of Table 6, Sanderson-Windmeijer multivariate F statistics for weak instruments with multiple endogenous regressors are reported and are consistently above 10 for parameters estimated for White households. For Black households, the F statistics can reach lower values, leading IV to be more biased towards OLS.

In summary, across OLS and IV, spatial mobility for Black households is substantially lower than for White households, who hold strong racial preferences. Black racial preferences are more minimal across all specifications.

Comparison to the Literature – There are few direct comparisons for the residential elasticity θ_r as previous studies commonly assume a joint Frechet distribution for both workplace choices and residential choices. The shape parameter then tends to be much higher, for example at 6.8 in Ahlfeldt et al. (2015). Similar to this paper, residential population elasticities to CMA in Tsivanidis (2022) are in the range of 0.318 to 0.746, around the estimates for White households. Lower Black migration responses to local labor demand shocks have been previously documented in Bound and Holzer (2000), and Diamond (2016) also finds smaller but noisy responses by Black households to cross-city changes. However, I am not aware of elasticities by race at the finer spatial scale of this paper.

For racial preferences, my findings are consistent with Bayer et al. (2004) indicating self-segregating preferences which are larger for Whites, Cutler et al. (1999) documenting that Black migrants to the urban North chose to live in Black enclaves, and Currarini et al. (2009) finding homophilic segregation of friendships.

5.4 Localized Costs

With the estimated residential elasticities and preferences parameters, I invert the model to recover fundamental amenities b_{igr}. I estimate how the Interstate highway system affected nearby neighborhoods through localized costs in amenities following $b_{igr} = 1 - b^{HW}\exp(-\eta DistHW_i)$. The exponential decay is approximated with non-parametric bins in the empirical specification below.

$$\Delta \log b_{igr} = \sum_{k=1}^{5} \beta_k \mathbf{1}\{DistHW_i = k\} + \mathbf{X_i}\eta_{gr} + \gamma_{m(i)gr} + \alpha_{red(i)} + \epsilon_{igr}$$

Bins are a mile-wide up to 5 miles from Interstate segments built between 1960 and 1970. The equation controls for distance from the CBD and geographic features in $\mathbf{X_i}$ interacted with group, city by group fixed effects, and redlining fixed effects. Standard errors are clustered at the tract level.

In Column 1 of Table 7 with no controls, I find there is a large drop in fundamental amenities where at 1 mile from the constructed network, $\Delta \log B_{igr} = -0.453$ (0.0501). However, much of the decline by highways is due to selection in route placement as including geographic controls in Column 2 reduces the estimate to -0.119 (0.0516). To gain precision, I further report results for 0.5 mile-wide bins in Column 3 where $\Delta \log b_{igr} = -0.191$ (0.0581) in the first 0.5 mile. These estimates are comparable in size to findings in Brinkman and Lin (2022) using cross-sectional variation from Chicago. To assign parameter values for b^{HW} and η, I match the functional form of $b_{igr} = 1 - b^{HW}\exp(-\eta DistHW_i)$ to two of the estimated bins in Column 3 at $k = 0.5, 1.5$.[29]

Additional Results – The population response is due to not just the direct negative consequences of highways but also the indirect changes in racial composition. I therefore set the composite amenity term $B_{igr} = b_{igr}(L_{iW}/L_i)^{\rho_r}$ as another outcome in Column 4 to include the indirect amenity impacts. I find endogenous amenities explain a small portion of the population drop by highways since the estimated value is only slightly more negative at $\Delta \log B_{igr} = -0.124$ (0.0517). Instrumented results are shown in Table A.13 and are too noisy to measure values precisely. In Table Appendix A.14, I project modern-day measures of environmental pollution over distance from Interstate roads and find a 2% increase within the first mile, which is a strong lower bound on pollution during Interstate construction as from the 1960s to today, car pollutants have been reduced by 99%.[30]

In a falsification test, I measure whether there were fundamental amenity changes near other historical large roads. Roads may have universally become more congested or polluted, and declines near Interstate highways may not be a distinctive feature that should be counted fully for welfare impacts. I replace the distance bins from Interstate roads with distance bins from historical control roads that were never re-built $\{\mathbf{1}\{DistLARGE_i = k\}_{k=1,\dots,5}\}$. Falsification results in Column 4 indicate no change in amenities near large roads with the estimate at 1 mile being very close to zero at -0.0057 (0.137). The negative consequences are thus a unique aspect of Interstate routes where their massive size and elevated ramps were particularly unpleasant for neighboring areas (Rose and Mohl, 2012).

5.5 Parameters from External Sources

To obtain the parameters for the economic forces of segregation, price sensitivity i.e. the consumption share of housing is predicted using the Consumer Expenditure Surveys (CEX) microdata in 1980, the earliest year available. I estimate a linear expenditure function over income in Appendix Table A.15, which matches the data well as shown in Figure 5.1. With average income levels by race and education, I predict housing consumption shares $1 - \beta_{gr}$, displayed in Appendix Table A.16.

Finally, I close the model with a few additional parameters. In the production function, the labor share is set to 0.7 following findings in Greenwood et al. (1997). The elasticity of substitution by education σ^g in the CES labor aggregate comes from Card (2009) which uses the education categories of high school versus college educated. Estimates range from 1.4 to 3 and are corroborated by several other sources, so I set $\sigma^g = 2$ (Borjas, 2003; Ottaviano and Peri, 2012). The elasticity of substitution by race is taken from Boustan (2009) to be $\sigma^r = 8$. Housing supply elasticity values are obtained from Baum-Snow and Han (2021) and set to differ in the central city (within 5 miles of the CBD) where $\mu_{cbd} = 0.35$ versus the suburbs (all other neighborhoods) where $\mu_{sub} = 0.25$. Lastly, the agglomeration parameter is set to 0.07 within the range of Rosenthal and Strange (2004) and Kline and Moretti (2014).

[29] Parameters are set to solve two equations: $1 - b^{HW}\exp(-\eta 0.5) = \exp(-0.191), 1 - b^{HW}\exp(-\eta 1.5) = \exp(-0.0994)$.

[30] The Clean Air Act of 1970 was the first of many federal legislative efforts to reduce air pollution.

5.6 Validation Exercises

With the estimated parameters, I conduct several tests to validate that the model predictions match the empirical moments. Using predictions from the commute shares equation in (23), I display in Figure B.9 Panel A the linear fit and binned scatter plot for log predicted commute flows and log observed commute flows in 1960 and 1970.[31] In Panel B, I plot the CDF of commute flows over commute time for predicted and observed counts. Across all the tests, the predictions tightly fit the data, and the R-squared of the weighted linear fit is around 0.9 in Table A.17.

I also examine how the model-recovered values for productivity are related to changes from the Interstate highway system and to characteristics across workplaces. In Table A.18, the change in log productivity is uncorrelated with distance from Interstate segments constructed between 1960 and 1970, so highway construction does not appear to contribute to agglomeration or reallocation of economic activity near Interstate routes. This result also rules out highway impacts on trade costs as an important channel for the location of firms because the productivity term contains any effects unrelated to the commuting channel of labor supply to workplaces. In the cross-section, firm productivity does not appear to be related to redlining in either 1960 or 1970 as shown in Table A.19. Interestingly, productivity is not substantially higher in the central city compared to the suburbs despite the evidence that dense cities tend to be more productive as in Combes et al. (2012) or that agglomeration operates at sub-metro scales as in Baum-Snow (2020). However, given the larger spatial scale of the workplace geographic units, it may be that finer correlations between workplace characteristics and productivity in changes or levels are present but challenging to detect.

6 Counterfactuals and Welfare

Having estimated the set of key model parameters, I undertake several counterfactual analyses to investigate the impacts of Interstate highways on inequality. In Section 7, I return to the role of institutions as a primary mechanism for racial disparities. However for the moment, because institutions are assumed to be within fundamentals and invariant to highway policy, I take the constraints on Black residential locations as given and proceed with the welfare assessment of the Interstate system.

6.1 The Impacts of the Interstates

What are the essential forces that drive inequality in highway policy? Returning to the expression in Equation (17) for welfare impacts in general equilibrium, the Interstate highway system affects the *fundamentals* of local costs by highways and commute times between bilateral pairs, which then lead to adjustments in *equilibrium outcomes* of prices, racial composition, and wages. Both changes in fundamentals and equilibrium outcomes are weighted by *initial shares* in the cross-sectional spatial distribution and are aggregated using the *substitution elasticities* across locations.

[31] The gravity equation was estimated from POW Zone by POW Zone bilateral counts in the observed data, so I aggregate predicted commute flows for residential tracts to the POW Zone level.

As the residential elasticities of the Black population are extremely low, their initial shares, which are heavily concentrated in the central city, are especially important for their incidence to highway impacts. This pattern strongly suggests that shocks to the urban core come close to fully determining highway impacts for Black welfare. To illustrate this hypothesis more definitively, in the next sections, I focus on specific channels in the quantitative framework and explore how welfare changes when different parts of the general equilibrium system are allowed to adjust.

Solving for Equilibrium Counterfactual Outcomes – For each city, I study the equilibrium changes in a counterfactual world where highways enter into the 1960 observed equilibrium with the reduction in commute times and the addition of localized costs along Interstate routes built between 1960 to 1970.[32] Across the 25 cities in the analysis, I calculate a weighted average with city-level population weights and report these averages as the main counterfactual numbers.

The general equilibrium framework through the system of equations provided in Appendix 6.3 governs how reallocation affects housing prices, endogenous amenities, and wages. Solving for equilibria follows the iterative procedure described by Allen and Arkolakis (2014), and model parameters are listed in Table 8. Residential elasticities are obtained from Table 6 where the elasticity for Black households is set to $\theta_N = 0.35$, lower than the elasticity for White households of $\theta_W = 0.8$. Preference parameter $\rho_B = 0$ because estimates of racial preferences were often insignificant for Black households, and for White households, $\rho_W = 1$ at the low range of the confidence intervals.[33]

To obtain the new equilibrium, I take the "covariates based approach" characterized by Dingel and Tintelnot (2023). Rather than the "exact-hat algebra" approach of predicting counterfactual changes from initial observed flows as in Dekle et al. (2008), I infer counterfactual changes with predicted flows generated using the estimated commuting elasticities. This approach avoids overfitting to the considerable sparsity of the data, especially for Black households, while largely conveying patterns of commuting behavior. The predicted flows are also used to recover fundamentals in levels.

General Equilibrium Impacts – In Table 9 Panel A, I present the GE welfare impacts separately for the four groups and pooled by race and by education to evaluate which demographic dimension exhibits larger disparities. Welfare changes are the lowest for less-educated Black households at -1.45% and the highest for higher-educated White households at 3.01%. Pooled by race, there continue to be losses for Black households of -1.04% and sizable gains for White households of 2.86%. Disparities by education are minimal with welfare gains of 2.07% and 2.79% for the less-educated and higher-educated, respectively. Consequently, I focus mainly on racial disparities.

[32] To conduct the simulation, I set the commuting time matrix to t_{ijgr}^{HW} where the Interstate highways are overlayed on the historical urban road network with the mode of transport weights for each race and education group set at their 1960 weights. The counterfactual therefore does not allow for changes in the mode of transport as a margin of adjustment. It is unlikely that accounting for this margin would change the ordering of the welfare impacts because Black households continue to commute with private automobiles at a lower rate than White households, even in the modern day (Bunten et al., 2022). I modify the exogenous amenity parameters $b_{igr,1960}$ to include the localized costs from the highway such that $b_{igr}^{HW} = b_{igr,1960}(1 - b^{HW}\exp(-\eta DistHW_i))$ with full decay at 5 miles.

[33] Larger values tend to create convergence issues. With these parameters, solving for counterfactuals with the iterative procedure leads to uniform convergence. Given the magnitude of these parameters, the sufficient conditions for uniqueness are no longer satisfied. See Appendix 6.4. However, the conditions are not necessary for uniqueness, and I do not encounter multiple equilibria with the smaller preference parameters.

Channels Behind Impacts – To show how general equilibrium effects alter the welfare results, I break down the impacts further with additional counterfactual exercises.

Direct Impacts – Returning to the welfare equation of (28), with a derivation provided in Appendix 6.1, the direct change in welfare ignoring residential or workplace responses is a transparent weighted average that depends on the initial shares in each residence and workplace and the changes in commute times and fundamental amenities from the Interstate system.

$$d \log U_{gr} = -\kappa_{gr} \sum_{i,j} \pi_{igr} \pi_{j|igr} \underbrace{\Delta t_{ijgr} / t_{ijgr}}_{\text{Commute Times}} - \sum_{i} \pi_{igr} \underbrace{b^{HW} \exp(-\eta DistHW_i)}_{\text{Local Costs}}$$

These direct impacts are visualized in Figure 3 for the city of Boston where the spatial distribution of the Black population is also provided for comparison. Given the disparate placement of highways across neighborhoods, it is unsurprising that in Table 9 Panel A, the direct changes in local costs are unequal; losses are -8.03% and -6.18% for Black and White populations, respectively. As Black car usage is lower and commute time reductions are muted in the central city, where the Black population lives, gains from commute benefits for Black households are 6.47% compared to 7.9% for White households. Both effects lean in the direction of increasing racial inequality so that on net, the direct impacts are -1.56% for the Black population and 1.72% for the White population.

Reallocation Only – To understand how spatial mobility impacts welfare, I then allow for household reallocation across locations, but no equilibrium outcome adjustments. The only forces at play are the changes in fundamentals of commute times and localized costs as well as the elasticities for residential and workplace choice. This exercise is operationalized by setting the hat $\hat{x} = x'/x$ of equilibrium outcomes to one in Equation (17). I find for the Black population that welfare losses are -0.98% and for the White population that welfare gains are 2.96%.

Compared to the direct impacts, the reallocation-only impacts are much more positive for White households. As they migrate both towards positive and away from negative aspects of the highway shock, they enlarge their gains relative to the Black population and widen racial inequality in welfare. Compared to the general equilibrium impacts, the reallocation-only values for welfare are quite similar, suggesting that equilibrium outcome adjustments play a small role in welfare. The minimal difference can be a result of offsetting equilibrium adjustments. For example, White households who reallocate to suburbs pay higher housing prices (lowering utility), but they also live in more segregated neighborhoods (raising utility), canceling out overall.

Partial Equilibrium and No Spillovers – As most of the empirical evidence was related to residential changes, I shut down adjustments on the firm side in a partial equilibrium counterfactual where any changes in labor supply do not affect wages paid to workers or housing demand from firms. To solve for this equilibrium, the system of equations is provided in Appendix 6.2. I find that the welfare results look broadly similar to the general equilibrium results in Table 9.

In a last counterfactual, I shut down spillovers from endogenous amenities and agglomeration, and I find again that welfare impacts are about the same. These findings highlight how changes in equilibrium outcomes play a smaller role while the mobility of groups is of greater importance.

Changes in Equilibrium Objects – The differential welfare effects are apparent through the changes in equilibrium outcomes, which I calculate for the four demographic groups and the two dimensions of race and education pooled in Appendix Table A.20. In the general equilibrium counterfactual, Black households experience large drops in amenities, which are central contributors to their welfare losses, although mobility responses somewhat reduce their incidence to highway costs. Their wages are slightly lower in general equilibrium, and they move marginally further into the periphery of cities and outside of redlined areas. With the commuting benefits, Black workers reduce their commute times overall even though they also substantially increase their commute distance in response.

White households experience much smaller drops in amenities, and they move substantially farther from the center of the city and from redlined areas. They respond more to the commute benefits by increasing their commute distances more than Black households. All of these equilibrium adjustments, differential by race, suggest how disparities emerge.

Additional Counterfactual Results – I probe the sensitivity of the welfare calculations to slight modifications in key model parameters with results in Appendix Table A.21. Researchers may hesitate to allow racial composition to affect welfare directly, so I set spillovers from racial preferences to zero for the White population and solve for the general equilibrium counterfactual. The results do not change substantially relative to the baseline with small differences in welfare for White households.

Alternatively, I allow for positive racial preferences for Black households by setting $\rho_N = -0.2$,[34] and with this change, Black households experience smaller welfare losses from the Interstate highway system of -0.62%. Because White households migrated out of the central city, the neighborhoods that Black households lived in experienced large shifts in racial composition. The model predicts that homophilic preferences contribute to some Black welfare gains, but they do not fully compensate for the overall losses from Interstate highways. Given the detrimental consequences of urban decline from the suburbanization of advantaged families, I report the results with positive racial spillovers for Black households as only a robustness check rather than as a main finding.[35]

Policy Implications – Lastly, I conduct a few counterfactual exercises related to policy directions for transportation infrastructure. Comparing the Interstate network to the planned routes or Euclidean rays in Appendix Table A.21, I find that welfare changes with planned routes are similar to the Interstate network, while gains are larger with the ray network, which removes beltways that contribute less to commute time reductions but substantially increase costs. All routes lead to disparities by race because Interstate highways were required to intersect central cities, where most Black households resided, and because of differences in car usage by race. Mitigating costs increases welfare, especially for Black households, shown in the bottom of Panel C in Table 9, which suggests that one effective policy would be reducing the harms of highways. Indeed, construction costs have greatly risen over time, partially in order to limit negative highway consequences (Brooks and Liscow, 2020).

[34] This value comes from the elasticity estimate of -0.07 from Table 6 divided by the residential elasticity of $\theta_N = 0.35$.

[35] Research across a variety of disciplines finds greater spatial separation by race leads to worse economic outcomes (Jackson, 1985; Massey and Denton, 1993; Bullard, 1993).

An alternative path to reducing racial disparities is increasing the mobility of Black households, e.g. by lowering information frictions, as studied in Ferreira and Wong (2020). In an additional exercise, setting the Black residential elasticity to the greater value of White households raises Black welfare, and setting both Black and White households' elasticities to three times the original value of White households leads the Black population to experience positive changes from the Interstate highway system. Interestingly, raising Black mobility leads to smaller gains for White households, who may then face more competition in response. These results highlight how spatial frictions (from institutions or other factors) that affect the mobility of Black households also limit their welfare gains.

7 The Role of Institutional Segregation in Welfare Impacts

In this section, I explore the factors that generated the spatial concentration of Black households in central areas and how segregation interacts with the Interstate highway system to produce unequal policy impacts. The focus is on how fundamental amenities shape Black residential choice, given that racial preferences are measured to be minimal and that economic differences do not appear to determine much of Black residential isolation (recall the summary statistics from Section 2.1). Fundamentals that affect the cross-sectional spatial distribution then translate into inequality in highway impacts since as documented above, initial shares are a key determinant of Black welfare.

I begin with differences in racial composition around the borders of redlining maps where an identification strategy permits clean tests of the presence of institutional barriers in fundamental amenities. I then discuss the takeaways from this strategy and consequently examine barriers away from the border to study institutional segregation more broadly.

7.1 Institutional Segregation in Border Discontinuity

Past research has measured how racial composition sharply shifts across the grades of HOLC maps (Hillier, 2003; Faber, 2014; Aaronson et al., 2021). As depicted in Figure 1, percent White drops by a sizable 18 percentage points upon crossing into redlined neighborhoods. Most similarly to Aaronson et al. (2021), I employ a border discontinuity design to analyze segregation associated with the HOLC maps. Yet instead of solely studying empirical changes in racial composition, I take a revealed preference approach and infer the *sources* of segregation. In equilibrium, rental prices respond and endogenous amenities reinforce sorting, so the differences Aaronson et al. (2021) find are not due entirely to institutions. Lower rents in redlined areas and racial preferences can be additional forces.

The empirical specification is a border discontinuity design estimated separately by race

$$Y_{igr} = \alpha_{gr} + \psi_r D_i^{red} + \mathbf{F}_r(DistRED_i) + D_i^{red} \times \mathbf{G}_r(DistRED_i) + \lambda_{ilr} + \xi_{igr}$$

where α_{gr} are education by race fixed effects, D_i^{red} is an indicator for i being a redlined neighborhood, and λ_{ilr} are fixed effects for if the nearest border is border l. $DistRED_i$ is the distance to the nearest border separating redlined neighborhoods from other neighborhoods, and a positive value represents being in a redlined neighborhood. $\mathbf{F}_r$ and $\mathbf{G}_r$ are non-linear functions of distance.

Y_{igr} corresponds to several outcomes for the factors behind residential choice. Redlined neighborhoods may be more racially diverse either because Black households live relatively more or because White households live relatively less in those areas. I start by setting the outcome as log population $\log L_{igr}$ to understand how population is distributed differentially by race. Then, with the estimated parameters and the model-implied relationship in Equation (16) for the cross-section, I decompose how residential choices are related to prices, commuter access, and amenities (endogenous and fundamental). Fundamental amenities, capturing discriminatory constraints outside of conventional determinants of choice, are the central outcome of interest where the identification assumption is that fundamentals should not be changing along the border except through institutional factors.[36]

In estimation, I follow the local polynomial approach of Calonico et al. (2014) and use the corresponding optimal bandwidth for each outcome. When inverting for amenities, I take the highest estimates for racial preferences for both White and Black households to obtain the most conservative value for the institutional component as well as the highest residential elasticity for White households and assign it to Black households. Parameters used are listed at the bottom of Table 10. Additionally, to avoid confounding the effect of social institutions with physical barriers or changes in school districts, I remove neighborhoods immediately by railroads, large roads, highways, and school district borders (which come from the National Center for Education Statistics).[37]

Main Results – The results are estimated using the 1960 microdata and are shown in Table 10. In Column 1, I find that the combination of Black households living more and White households living less in redlined neighborhoods leads to the drop in percentage White. In Panel A for Black households, I estimate a striking 1.425 (0.226) increase in log population, in line with historical evidence that Black households were heavily concentrated in redlined areas. For White households in Panel B, there is a sizable -0.546 (0.122) decline in log population entering into redlined neighborhoods.

When amenities are used as the outcome and the role of prices (and CMA, though it is continuous along the border) is removed in Column 2, I find the discontinuity is only slightly reduced for Black households to 1.370 (0.238). For White households, the estimate of -0.603 (0.112) is more negative as they also prefer lower prices. Accordingly, cheaper rents cannot explain why White households live less in redlined areas nor much of the change in Black and White populations over the border.

I finally set fundamental amenities as the outcome and remove racial preferences in Column 3. I find that the discontinuity remains large for Black households at 1.266 (0.209) and disappears for White households to 0.0031 (0.097). Racial preferences fully account for why the White population does not live in redlined areas, indicating institutions play *no role in White residential locations* and may benefit White households by preventing racial integration. Yet, preferences only explain a portion of the increase in the Black population, leaving a large residual to be attributed to institutions.

[36] This identification assumption is distinct from that of Aaronson et al. (2021) who search for borders where there were no pre-existing racial divisions before the maps were drawn in 1932, as they aim to measure the treatment effects of the HOLC maps. Consequently, the discontinuity estimates of this paper capture institutions prior to the HOLC while those of Aaronson et al. (2021) do not.

[37] The sample is limited to tracts at least 0.1 miles away from historical large urban roads, constructed highways in 1960, or historical railroads and also at least 0.1 miles away from a school district boundary where school districts come from the 1989-1990 school year, the earliest year with district maps from NCES.

Adding additional socioeconomic controls in Column 4 does not change the results for White households and lowers the discontinuity for Black households to a still sizable value of 0.914 (0.181). This estimate is the preferred value for the counterfactual simulations, and I set parameter $\bar{E}$ following that $\psi_B = 0.914 = \theta_r[\log b_{mB} - \log b_{nB}] = \theta_r \log \bar{E}$ for $m \in \mathbf{R}, n \notin \mathbf{R}$. As the discontinuity estimate in fundamental amenities is essentially zero for White households, neighborhoods at the border do not appear to be substantially different in their characteristics, and importantly, the identification assumption is satisfied. For Black households, 65% of the massive population rise entering redlined areas is in residual fundamental amenities and thus a result of institutional barriers.

Robustness for Identification – To further test whether the identification assumption is satisfied, in Appendix Table A.22, I assess if natural amenities of land cover types and tree cover change discontinuously along the border. Data on land types such as open water, wetlands, and deciduous forests come from the National Land Cover Database (NLCD), and tree cover canopy comes from the U.S. Forest Service. I find amenities that are less manipulable, i.e. open water and wetlands, are smooth across the border, which bolsters the identification assumption. Features such as tree cover and deciduous forest, which may be considered *endogenous* amenities, do however differ.

Additional Results – In Appendix Table A.23, I take a reduced form approach to decomposing the discontinuity, reducing the reliance on exact parameter magnitudes. Instead of using the parameters to invert for amenities, I residualize population on prices and commuter access, racial composition, and demographic controls successively with redlining fixed effects, using the cross-sectional variation within neighborhood types, and project the residuals onto the border. This cross-sectional variation likely contains bias compared to the estimated parameters, but I report the results as a robustness check. The estimates are broadly the same where in Column 4, the discontinuity is still large for Black households at 0.489 (0.204) and continues to be zero for White households at 0.021 (0.079).

Lastly, I present additional results in Appendix Table A.22. In Panel A, I probe the sensitivity of the drop in percent White across the redlined border by: (1) adding and removing the border fixed effects, (2) forming balanced samples where the number of tracts on the redlined and non-redlined sides is the same, and (3) altering the restrictiveness of how many neighborhoods are dropped away from physical barriers and school district borders. These adjustments do not greatly alter the findings. I further display in Panel B the border discontinuity estimates for control variables used to residualize the fundamental amenities. Lastly, I show how segregation along the border has changed over time from 1950 to 1990. The discontinuity in racial composition was largest in 1960 and 1970 and declined dramatically in 1980 after a decade of fair housing initiatives post Civil Rights legislation.

Welfare Impacts – To understand how institutions influence the impact of the Interstate highway system, I conduct a counterfactual exercise where (1) institutional barriers from the border discontinuity are removed and then (2) Interstate highways are constructed in this new environment. To do so, I adjust the fundamental amenity term for Black households in non-redlined neighborhoods by the factor $\bar{E}$ so that amenities are relatively higher, thereby expanding access to non-redlined areas.

I re-compute the general equilibrium counterfactual for Interstate impacts to measure the interaction between barriers by the border and infrastructure policy. Welfare results are displayed in Table 9 Panel C, and I find that losses for Black households are reduced from -1.04% to -0.64%. Notably, White households regardless of education experience essentially the same welfare effects. Reducing institutional barriers limits inequality in highway impacts because absent exclusionary barriers, the Black population is able to live farther from the central business district by 11% (see Table A.24). As noted earlier, initial locations are especially important for Black households because their residential elasticities are low. When they are less spatially concentrated in the central city, Black households bear less of the costs of Interstate highways (as shown through a smaller drop in amenities in Table A.20) and are able to gain more from commute benefits rising in suburban areas.

In the intermediate step for the welfare calculation when institutions are relaxed, higher-educated White households experience a -0.5% drop in welfare as shown in Table 9 Panel C; this result motivates why exclusionary barriers were upheld by White groups who fought tenaciously to preserve the status quo (Massey and Denton, 1993). Because of homophilic preferences, racial integration lowers the welfare of White households, and in Table A.24, I find that amenities for higher-educated White households are 0.8% lower as they face more competition in their choice of neighborhoods. Reducing barriers by the border allows Black households to live 27.5% less in redlined areas and in neighborhoods that are 8.4% more White.

7.2 Institutional Segregation Writ Large

Nonetheless, removing institutions by the border only somewhat improves Black welfare gains and does not greatly close the welfare gap by race, which decreases by a modest 10%. The border discontinuity estimate of institutions is likely an underestimate of the extent to which Black households are excluded from broad sections of cities as the change in racial composition at the border pales in comparison to the stark segregation during this era. In Figure B.10, I plot the spatial distribution of the Black and White population over the racial composition of census tracts in 1960. Visible in this figure is how most White families lived in racially homogeneous neighborhoods—70% of census tracts in this sample are more than 99% White. Near the border, non-redlined areas tend to be more racially integrated than the average White neighborhood while redlined areas tend to be less racially diverse than the average Black neighborhood. The discontinuity estimate is then a local average treatment effect that overlooks heterogeneity in institutional factors further away.

These findings illustrate a perennial tradeoff in economics between proper identification and the potential scope of the question. While the border design allows for a testable identification assumption, it limits the paper to examining a narrow set of neighborhoods. However, useful conceptual lessons are gleaned from the border design. Specifically, institutional barriers appear to be a large determinant of Black residential choices and the relative Black-White difference in fundamentals.

Welfare Results – With this insight, I examine how institutional segregation *writ large* impacts inequality from Interstate highways. I make the stronger assumption that fundamental amenities, generally representing natural amenities such as ocean views or green hills, should not be valued differentially by race *across all neighborhoods*. For example, forested suburbs should receive a high valuation from Black families but do not because of discrimination. I subsequently set the fundamentals of Black households equal to those of White households, who have free rein in choosing where to live, and open up residential access for the Black population. This scenario represents an upper bound on the extent to which erasing discrimination can reduce segregation as there may be true racial differences in fundamentals, e.g. from social networks or information frictions.

In this new environment, I display the changes to welfare from Interstate development in Panel C of Table 9. The racial gap in the general equilibrium impacts from highways is greatly diminished by 54%, and thus institutional barriers determine the majority of inequality from the Interstate system. Black households now receive welfare gains from highways 1.04%, versus previously they were facing losses of -1%. This result follows from the major reduction in the spatial concentration of Black families, who now live 93% farther from the CBD. White households experience similar gains to before of 2.8% so relaxing residential discrimination does not greatly alter their benefits from Interstate development. The remaining gap by race can be attributed to differences in commuting, mode of transport, and general equilibrium outcome adjustments, although this latter channel is likely less important given previous evidence. Segregation, specifically through forces that cannot be accounted for by economic or social factors, plays a crucial role in determining inequality in highway impacts.

Lastly, in the intermediate step when fundamental amenities are assigned to be equivalent across race, there are substantial welfare losses for White households of -1.8%. With further racial integration and increased competition in the housing market, White families fare far worse. Amenities of higher-educated White households are -5.64% lower, and their neighborhoods become 6% less White, providing further justification for why barriers were instituted.

8 Conclusion

In this paper, I develop a theoretical framework and build several rich historical datasets to measure the impacts of Interstate highways on inequality. To comprehend why there are profoundly disparate effects by race, I find that institutions are a primary determinant of the spatial concentration of Black families and interact with highway policy to produce vast disparities. The geographic separation of demographic groups and the selective placement of the Interstate network lead the benefits and costs to be shared unequally, with the most disadvantaged bearing more of the costs while garnering fewer of the benefits. Institutional segregation further generates spatial frictions that limit the spatial mobility of Black households and how much they are able to gain from the Interstate system.

While it may seem that Interstate highways and institutional segregation are things of the past, road infrastructure expansions in the modern day encounter the same equity concerns as they have historically.[38] The persistence of segregation along racial and economic lines and the political disempowerment of groups of color leads the harms of critical infrastructure, such as industrial facilities, to be borne by the most marginalized populations (Currie et al., 2022). Discrimination in housing continues, thereby restricting residential choice for Black families and how much they respond to

[38] A $9 billion highway widening project in Houston, Texas was paused by the Federal Highway Administration in 2021 after social justice groups opposed the expansion. The re-routing of parts of I-45 would displace predominantly Black and Latino neighborhoods as well as the original Chinatown of downtown Houston.

any placed-based shocks (Bayer et al., 2021). Moreover, the radical and permanent transformation of cities brought about by the Interstate highway system can continually endure through intergenerational consequences, as studied in a companion paper. Future research can aim to understand which neighborhood interventions improve the spatial mobility of the Black population and increase access to economic opportunity for the most disadvantaged families.

Chapter 3 – Opportunity in Motion: Equilibrium Effects of Highway Construction on Economic Mobility

9 Introduction

The U.S. exhibits vast disparities in economic opportunity across its cities and persistent racial gaps in the long-run outcomes of children, even conditional on parental income (Chetty et al., 2020). Place is commonly considered a leading determinant of intergenerational inequality as employment access, educational quality, and peer networks contrast starkly between the neighborhoods that Black versus White children live in (Wilson, 1987; Reardon and Bischoff, 2011). A natural question to ask is: can policies that target places alter them for the better and influence these gaps? Essential to the answer is a deeper understanding of the neighborhood characteristics that contribute to positive outcomes and to what extent place-based policies enhance or diminish these characteristics to transform economic opportunity across places. A tension arises, however, if promoting one location comes at the expense of others through the shuffling of peer quality between locations, given that high-status peers are in fixed supply.

In this paper, I exploit the construction of the interstate highway system, one of the most prominent place-based policies in the U.S., to investigate the importance of place for intergenerational mobility through *two channels*. First, I find highways dramatically increased access to workplaces for suburban neighborhoods, and these improvements in access raised incomes to markedly boost economic opportunity. Second, given the massive scale of the policy, households responded by migrating toward areas with greater commuting cost reductions. I document this response is heightened for more-educated, higher-occupational status, and White families. Places left behind—in this case, neighborhoods in the central city—became populated by fewer advantaged peers and subsequently faced a loss in peer externalities. The migration responses, which in turn affect peer composition, are "general equilibrium" indirect impacts of policies that serve as another channel through which highways influence the level of economic opportunity.

As is apparent, while some impacts are beneficial, secondary consequences may not be. Place-based policies such as infrastructure often generate spillovers that create local gains but losses elsewhere (Glaeser and Gottlieb, 2008; Kline and Moretti, 2014). Rather than studying spillovers through agglomeration economies, the focus of previous work in Duranton and Puga (2004) and Greenstone et al. (2010), this paper advances an alternative source of local externalities that likely play a larger role at the neighborhood level: spatial sorting or segregation, which has been documented to be a key determinant of inequality in productivity, human capital, and long-run outcomes (Massey and Denton, 1993; Sharkey, 2008; Diamond, 2016; Chetty and Hendren, 2018a; Fajgelbaum and Gaubert, 2020). Derenoncourt (2022) finds Black migration from the South re-shaped economic mobility in Northern urban areas. In this paper, I take a more structured approach to quantify the magnitude of the migration channel.

To capture the indirect effects, I develop a flexible theoretical framework to examine how equilibrium sorting of large-scale policies affects long-run consequences. The innovation of this paper is to measure the economic mobility of children, rather than the productivity or welfare consequences of policy interventions, as the central outcome of recent quantitative economic geography models (Redding and Rossi-Hansberg, 2017). The framework thus integrates two strands of literature on spatial economics and intergenerational mobility, and notably, it is sufficiently general for investigating how myriad place-based policies impact long-run outcomes.

In the framework, spillovers are a force for amplifying inequality across places. Targeted neighborhoods experience both economic improvements and increases in peer externalities, while those not targeted face declines in peer externalities. Reduced form comparisons across commuting access improvements combine positive treatment effects for targeted areas as well as negative treatment effects for non-targeted ones. Taking the structure of the quantitative framework seriously, I use the model to decompose these reduced form comparisons into direct and indirect equilibrium impacts. Since the policy creates winners and losers, the framework enables assessing whether in *aggregate*, children's outcomes were improved—or rather, if gains in some locations were offset by losses in others—and whether the gains were shared equally across race.

Quantifying the net impact of place-based policies on children's long-run outcomes involves an array of parameters that are each challenging to measure. To start, it requires (1) empirical estimates of the degree to which characteristics of neighborhoods change in response to a place-based policy. In the case of the interstate highway system, I focus on the characteristics of average income and peer composition. The former is affected by access to employment and the latter is determined by reallocation in response to commuting access improvements. Importantly, reallocation not only influences the characteristics of neighborhoods, it also shapes which children are exposed to the characteristics. These statistics are *specific to the policy* of interest.

To translate these place characteristics into children's outcomes, I then require (2) parameters on the treatment effects of exposure to higher average neighborhood income and higher status peers for different groups of children. While research by Chetty et al. (2020) provides correlational evidence on how segregation and average income relate to outcomes by race, they do so often at a coarser geography (commuting zone or county versus neighborhood). In later sections, I emphasize how estimates at the finer spatial scale of the neighborhood level (census tract) differ substantially from estimates at larger geographies. I then describe how to exploit empirical variation from the interstate highway system to come closer to causality for these characteristics. The treatment effects of place characteristics are *not specific to a policy* and can be applied broadly.

To measure the long-run outcomes of children for the period of interstate construction, I employ novel parent-child linkages for the near universe of the 57 million children born between 1964–1979, constructed at the Census Bureau using historical IRS tax data. To build the linkages, described in more detail in Stinson and Weiwu (2023), we apply name-matching techniques that incorporate machine learning methods and restricted names from the Social Security Administration. We attain an exceptionally high match rate of 67% for the whole population.

These newly linked cohorts born between 1964–1979 fill a gap for large-scale measures of intergenerational mobility, and a particularly crucial one. The period of their childhood spans several pivotal moments in American history: the Civil Rights movement, the War on Poverty, and the creation of Medicaid—some of the most ambitious domestic programs to promote economic and social justice. Modern-day measures as in Chetty et al. (2014) begin with cohorts in the 1980s, and earlier measures with full-count Censuses end with cohorts in the 1910s (Abramitzky et al., 2014).

Using these linkages, I find upward mobility for Black children was strikingly low during this time. Black children from the top quintile of the parental income distribution were more likely to fall to the bottom quintile than remain in the top quintile, and intergenerational gaps by race are considerable. Conditional on family income, Black children reach adult income ranks that are around 17 ranks lower than White children across the parental distribution. Place appears to be strongly associated with children's outcomes as I document substantial cross-county and cross-tract variation in incomes for Black and White children, suggesting a possible role for place-based policy to shape the course of children's lives. Furthermore, given these large intergenerational gaps, a key question to ask is: how much did the interstate highway system contribute to them? In later sections, I go beyond associations to measure causal impacts of neighborhoods, which then allows for understanding the impacts of policies.

I then turn to developing the theoretical framework to map economic policy into long-run incomes for children, and this framework clarifies the sources of statistics necessary for assessing aggregate consequences on intergenerational mobility. Some statistics can be derived directly from the quantitative model. This quantitative spatial framework builds on intra-city models of neighborhoods from Ahlfeldt et al. (2015) and Tsivanidis (2022) and captures the factors behind the residential choices of households such as housing prices, amenities, and commuting access. Moreover, it characterizes the reallocation response after a policy shock, which impacts the peer composition of places, and predicts adjustments in average income across locations.

However, rather than relying solely on the model, I also provide empirical evidence on the impacts of interstate highways on neighborhood characteristics with quasi-random placement from an instrumental variables strategy. With planned maps that I digitized for 100 cities, I instrument highway location with the location of the planned routes, similarly to Baum-Snow (2007). To show that economic conditions are affected, I find that average income rises significantly with increases in commuting access. I also document strong reallocation and sorting responses to the interstate highway system as areas with greater increases in connectivity, which tend to be suburban neighborhoods, experienced inflows of White, higher-education, and higher occupational status households. These inflows necessarily imply outflows from central neighborhoods. The mobility responses then lead to changes in peer composition in both suburban and central neighborhoods. Importantly, this differential response by each group to commuting access is used to discipline the structural elasticities in the quantitative framework, so the predictions of the model align tightly with what is observed in the data.

Given these observed changes in neighborhood characteristics in response to the interstate highway system, I then examine how they translate into children's outcomes by estimating the treatment effects of characteristics. Here, the model serves the purpose of clarifying why there might be challenges in obtaining quasi-random variation in exposure to neighborhoods. In the model, residential demand is governed by both observable characteristics of neighborhoods and idiosyncratic factors.

The former suggests there is likely selection in the choice of neighborhoods if more advantaged families search for areas that benefit their children's outcomes. However, if idiosyncratic factors also lead to varying exposure to neighborhoods, harnessing that idiosyncratic variation enables estimating causal impacts of places and their characteristics.

I begin with descriptive correlations between average income, racial composition, educational composition and adult income ranks of White and Black children at the tract-level. I find strong relationships between all these characteristics and adult income for both White and Black children, with larger magnitudes for White children. However, these estimated values may not be due to causal treatment effects because of selection and omitted variables bias. Selection arises when more advantaged families, whose children fare better on average, are more likely to choose neighborhoods that are higher income and with higher status peers. Accordingly, the association between the characteristics and children's outcomes is partially driven by systematic differences in the types of families that live in better neighborhoods. Omitted variables bias may be another concern as neighborhoods with higher income or more White peers may be unobservably different along other dimensions correlated with children's outcomes.

To address selection, I implement an extension of an empirical design originally developed in Chetty and Hendren (2018a) which employs moves of children at different ages to generate quasi-random variation in exposure, hence its title of the "movers design." The intuition behind this strategy is that children who move at earlier ages receive a greater dosage of the neighborhood they move to compared to children who move later. With this design, Chetty and Hendren (2018a) compute the causal impact of counties and commuting zones across the United States, which they correlate with observable features of places.

I build on this design by measuring moves *along each neighborhood characteristic* and calculate how children's incomes in adulthood vary depending on the length of time spent in tracts where average income is higher or peers are more White, higher-educated, and higher-occupational status. This strategy is in contrast to the approach commonly taken in the past of estimating place effects for each location and then projecting these place effects on neighborhood characteristics (Alesina et al., 2021; Heath Milsom, 2023). At the tract-level for Black children, estimating place effects is infeasible given the limited number of observations and the large count of tracts.

Consequently, employing the extended design, I instead concentrate the variation along a single dimension of the tract characteristic and substantially reduce the dimensionality of the exercise. I obtain coefficients for the treatment effects of neighborhood characteristics and find similar magnitudes for both Black and White children. Higher income and better peers lead to improved outcomes later in life, and the treatment effect is strongest for being with higher-educated peers. For example, a one standard deviation increase in the percentage of the neighborhood that is higher-educated (high school graduate for this period) leads to an increase of one income rank for White children. This treatment effect is about one-third the size of the descriptive correlation between income and educational composition of peers, so two-thirds of the association stems from the selection of families across locations.

In contrast to prior evidence for counties or commuting zones, selection plays a larger role than treatment effects for the relationship between neighborhood quality and children's income. Selection is also more extensive for White children, which is consistent with the greater magnitude of the descriptive correlations for White children, despite the similar treatment effects by race. Given that the degree of selection is related to the ease with which families move to choose better neighborhoods, it is unsurprising that selection is higher for small geographies, where moves are more frequent, and for White families, who are more geographically mobile because they face fewer constraints that hinder mobility as found in Weiwu (2023).

In the final section of the paper, I combine the previously estimated statistics to assess the aggregate consequences of interstate highways on intergenerational mobility. This section is currently in progress. In future work, I solve for the full predicted migration response and changes in neighborhood characteristics post interstate development using the spatial equilibrium model and the set of empirical elasticities to commuting improvements. Combining the treatment effects of place characteristics with the model structure, I measure the impact of the interstate highway system on intergenerational mobility by race through its two channels. By shutting down the reallocation response, I isolate the role of the economic impacts through improvements in commuting access. The full general equilibrium counterfactual then informs how important the secondary migration responses and peer composition changes are for economic mobility.

Related Literature – This paper is connected to a vast literature on intergenerational mobility. An early body of work by Loury (1976), Becker and Tomes (1979), and Solon (1992) highlights how the dynamics of racial inequality depend on the persistence of income across generations. Approaches to compute intergenerational mobility are studied by Mazumder (2005), Dahl and DeLeire (2008) and Nybom and Stuhler (2016), and historical linkages that enable measuring its evolution over time are constructed in Abramitzky et al. (2012), Long and Ferrie (2013), Collins and Wanamaker (2014), Olivetti and Paserman (2015), Feigenbaum (2016), and Bailey et al. (2020). This paper provides the first large-scale measures of intergenerational mobility for the 1960s and 1970s, a turbulent but critical period for the United States. I find extensive disparities by race that are larger than for modern cohorts, and exploiting the near-universal scope of the data, I also uncover significant spatial variation in intergenerational inequality.

This paper is also related to a body of work on the long-run impacts of transportation infrastructure. Previous studies have measured how roads fueled the Great Migration of African Americans from the rural South as in Black et al. (2015) or affected local labor market opportunities (Adukia et al. 2020, Costas-Fernandez et al. 2023). Yet, few papers have detailed measures of job access capturing the richness of the transportation network, and most study changes over distance from roads (only Heath Milsom (2023) has market access terms, but for trade networks). However, whether being near a road is beneficial depends on what the connection leads to, thus requiring additional information. Further, few transit developments are sufficiently large enough to trigger detectable general equilibrium impacts, especially during a time period with indicators of intergenerational mobility at fine spatial scales. In this paper, the context of the interstate highway system plus the timing of the parent-child linkages enable quantifying the importance of general equilibrium impacts for economic mobility.

This paper is further tied to a rich literature on the geographic determinants of children's outcomes. Kain (1968), Wilson (1987), and Haltiwanger et al. (2020) analyze how spatial mismatch, i.e. disconnection between residences and employment opportunities, worsens the economic prospects of low-income, Black families. This paper directly shows that reducing spatial mismatch produces positive economic consequences. Research by Massey and Denton (1993); Sampson et al. (2002); Sharkey (2008); Andrews et al. (2017); Chyn (2018) measures how concentrated poverty and segregation are detrimental for long-run outcomes. Most closely related is recent work by Chetty and Hendren (2018a) and Chetty et al. (2020) using IRS administrative data to study the geography of opportunity across locations. In this paper, I provide evidence of how large-scale policies alter places and change segregation to show that economic opportunity is not fixed over time. The key implication is that instead of moving families to better neighborhoods, policy-makers can influence the levels of opportunity across places (while being cognizant of their general equilibrium impacts).

Finally, the framework of this paper builds on a rich literature in quantitative spatial economics (Allen and Arkolakis, 2014; Ahlfeldt et al., 2015; Tsivanidis, 2022). Busso et al. (2013) and Diamond and McQuade (2018) measure how neighborhood interventions interact with population movements and the housing market to produce add-on effects. Gaubert et al. (2021) assesses how to optimally design place-based policies given subsequent mobility and sorting impacts. I also highlight the importance of population mobility for equilibrium outcomes but I modify the objective function to study intergenerational mobility. Most closely related to this paper is recent work by Chyn and Daruich (2022) which develops an overlapping generations model to measure the impacts of housing vouchers and location-specific subsidies on children's outcomes. However, their model features only two locations and is calibrated to estimates from the literature. This paper constructs original empirical estimates directly linked to the policy of interest with rich spatial variation for the whole country.

Summary – The remainder of the paper is structured as follows. Section 2 describes the novel parent-child linkages and administrative data. Section 3 produces estimates of intergenerational mobility for the mid-20th century. Section 4 provides historical background on the interstate highway system. Section 5 characterizes the theoretical framework. Section 6 presents the empirical evidence on highway impacts. Section 7 measures how place characteristics affect economic mobility. Section 8 conducts counterfactual analyses. Section 9 concludes with policy implications.

10 Novel Historical Data on Intergenerational Income Mobility by Race

To measure intergenerational mobility and Black-White income gaps for the mid-20th century, I create a new panel dataset of children born in the years of 1964 to 1979 with novel parent-child linkages from Stinson and Weiwu (2023). In this dataset, economic outcomes and detailed locations are observed over the entire span of the children's lives into the modern day.

Name-Matching Children to Parent Tax Filers for the 1964-1979 Cohorts – For the technical details behind constructing the parent-child linkages, a report is available in Stinson and Weiwu (2023). In this paper, I provide a brief overview.

We begin with the universe of children in the cohorts of 1964 to 1979 from the Numident, a Social Security Administration (SSA) database of individuals with Social Security numbers (SSNs). In the Census version of the Numident, SSNs are replaced by unique personal identifiers called Protected Identification Keys (PIKs) that allow for linking to other Census surveys. These children are matched to parents who filed IRS 1040 tax forms in 1974 and 1979, the earliest years the Census and IRS retained complete income tax data. We follow an iterative matching approach similar to Abramitzky et al. (2012) and successively relax the comparison criteria to obtain a larger number of children-parent linkages. Each round of matching is detailed in Appendix 7.1.

The matching variables we assign for the children are: (1) names of both parents provided by the SSA in a restricted Numident file and (2) state of birth. These two variables are respectively matched to (1) names of the primary and secondary tax filers on the 1040 forms and (2) state of tax filing. Only native-born children are included in the sample because state of birth is unavailable for the foreign-born, who would not match on the variable for state of tax filing. As names are listed imprecisely, we modify and apply the fuzzy matching techniques of Cuffe and Goldschlag (2018) created for business record linkage to this setting for child-parent name matching. The linkage algorithm integrates multiple string comparison functions from natural language processing into a machine learning (random forest) model to flexibly distinguish matches.

To calibrate the algorithm, training data is constructed using true children-parent matches from IRS 1040 tax forms in 1994, the first year that tax filings included dependent identifiers. With the trained algorithm, completing the full set of matches for the universe of 1964-1979 cohorts is computationally intensive as n-squared pairwise comparisons are required.[39] We parallelize the algorithm of Cuffe and Goldschlag (2018), which was designed for smaller samples, and conduct the matching on Amazon Web Services through a pilot project with the Center for Optimization and Data Science at the Census Bureau.

Match Rates – With these linkages, I calculate match rates listed by year of birth in Table A.25 with an average rate across the years of 67%. In total, 38 million children are matched to parents in either the 1974 or 1979 tax filings. These rates are substantially higher than those found in other historical linking studies such as 6% in Ferrie (1996), 7-20% in Abramitzky et al. (2012, 2014), 21% in Collins and Wanamaker (2014), 45% in Bailey et al. (2020), 56%-60% in Feigenbaum (2015, 2016) who also employs a machine learning approach, and 5-30% in Abramitzky et al. (2020) who uses the Expectation-Maximization (EM) algorithm.

Several factors contribute to the high linkage rates of this paper. Notably, all names inputted into the matching procedure come from comprehensive administrative sources that cover the entire population and are less error-prone than survey responses. Additionally, rather than relying on manual matches such as in Feigenbaum (2015), the machine learning model is trained on true matches from corresponding modern administrative data for parent tax-filers and listed dependents from the 1994 IRS 1040 form. The flexibility of the random forest model with its multiple string comparison functions further captures additional matches compared to traditional Jaro-Winkler comparison based

[39] Pairwise comparisons occur within each block where blocking variables are formed from state of birth and the first and last initials of parent names. See Appendix 7.1 for a detailed description.

methods. Lastly, name matching such as in Abramitzky et al. (2012) uses the first and last names of children (and often only sons as the last names of daughters change after marriage) leading to many non-unique names that cannot be disambiguated. We link on both parents' names, and the combination of two names eliminates a substantial amount of non-uniqueness in comparisons. Any preliminary matches where either the Numident observation or tax filing observation is matched to more than one counterpart are dropped from the sample, so all final matches are unique.

Match rates by gender and race are displayed at the bottom of Table A.25. Rates are essentially the same across men and women because matching on parent names addresses the complication of name changes upon marriage for women. As in other studies, it is challenging to attain match rates for the Black population that are as high as that for the White population due to their lower coverage in survey and administrative sources. While the match rate for the White population is exceptionally high at 72%, the match rate for the Black population of 60% is still notable, reaching the highest match rates in other datasets for the White population.

Parental Income Measures – Parental income is obtained from IRS 1040 forms available in 5 year intervals from 1974 to 1994, in 1995, and annually from 1998 to 2018. As measurement error and volatility in reported income can introduce bias into calculations of intergenerational mobility, I compute average income with the four years of tax data available between 1974 and 1989 during the youth of the selected cohorts (Solon, 1999; Mazumder, 2005). For birth cohorts born in years up to 1974, the 1974 form is the first available. For those born after 1974, the first is the 1979 form.

Child Income Measures – For children, I measure household income in adulthood from IRS 1040 forms in the years between 1999 and 2018 when the cohort is between the ages of 35 to 39. Income is averaged over these 5 years for a stable measure of household income at mid-life to avoid the previously mentioned issues of measurement error and volatility. Calculating income during this age range also addresses some of the life-cycle biases of Chetty and Hendren (2018a) as noted in Haider and Solon (2006) and Nybom and Stuhler (2016).

When studying intergenerational dynamics, researchers often use household income to represent the economic resources available to children through their parents. However, large differences in marital status by race mechanically create Black-White gaps in household income as Black households more frequently are comprised of single-earners (Chetty et al., 2020). To isolate the role of marriage, I measure individual income using W-2 earnings records from the IRS. For the 1970 to 1979 cohorts, I calculate average individual income over the age range of 35 to 39, the same range as for household income. For the 1964 to 1969 cohorts, I instead measure average individual income over the age range of 41 to 45 since W-2 earnings files are available starting in 2005 when the 1964 cohort is aged 41.

Race – Both parents and children are linked to the 2000 and 2010 complete-count Decennial Censuses and ACS surveys from 2001-2020 to retrieve race. In Panel C of Table 11, I display counts for each race group. A small percentage (8%) of the children are unable to be located in either the 2000 or 2010 census or ACS and have no race specified. Hispanic is separated out from White and Black throughout.

Geographic Variables – Moves are observable in the 1040 forms at detailed geographies through the address of filing variable. As filings are available infrequently in the earlier tax data, I approximate the year of the move as the midpoint of the 5 year interval (or 3 year interval for 1995 to 1998). For example, if I observe that the county has changed between 1974 and 1979, I assign the household location as the origin county from 1974 to 1976 and as the destination county from 1977 to 1979. I count moves over the span of the individual's childhood starting with the first available tax year in 1974 until age 23, following Chetty and Hendren (2018a).

Geographic variables are available at the county-level for almost all children and at the tract-level for the large majority, as shown in Table 12. As I use a movers design later on, I verify that this smaller sample is representative of most children in the U.S. One-time movers are strikingly similar along many economic characteristics to those who never move or who more than once at both the county and tract-level. While White households are more likely to move across counties, Black households are more likely to move across tracts within counties. This pattern suggests Black families may face more residential instability without greatly transitioning across types of neighborhoods.

Parental Background and Later Life Outcomes – The long form version of the Decennial Census in 2000 and the American Community Surveys from 2005 to 2020 contain additional individual-level variables such as education, occupation, marital status, incarceration which are linked to both parents and children. With this sample of earlier cohorts of children, many more outcomes in adulthood are observable in the 2000 Census and ACS surveys compared to previous studies where the sample of children may not have realized outcomes by this date (Chetty and Hendren, 2018a,b).

Representativeness – I examine how representative the matched children are of the overall population of children in Table A.26. Comparing the unmatched Numident children from the 1964-1979 cohorts to the children matched into the IRS parent tax filers, I find that matched children tend to fare better later in life in both educational attainment and adult income. Matched children are in households where adjusted gross income (AGI) is $92,000 in 2018 dollar terms while the AGI of unmatched children is $82,000. As a result of the differing match rates across birth years and across race, some of the differences are driven by the differing cohort and racial composition of unmatched vs. matched children. In Column (4) of Table A.26, I test whether group means are statistically different while including birth year and race fixed effects. While the difference between matched and unmatched children is reduced by around 40% across most outcomes, a statistically significant difference remains. For AGI, there is continues to be $5,600 difference between the two groups with year and race fixed effects. Consequently, positive selection into the sample should be considered when evaluating the results later on.

Summary Statistics – In Panel A of Table 11, I display summary statistics related to income, rank, and education for White and Black children. Large racial disparities are present across all variables. On average, AGI for Black children in adulthood is $49,000 which is less than half the average of AGI for White children in adulthood of $102,000. Children also begin in economic backgrounds that are vastly disparate by race; the average parental household income rank for White children is 55.5 while parental income rank for Black children is 34.4 in the national distribution of their cohort.

As much of the analysis on causal impacts of locations employs a movers design, I examine if summary statistics for the sample where geographic variables are available in the tax data appear to be different from those for the overall matched sample. Comparing Table A.26 Column (2) to Table 12 Column (1) for county one-time movers and Table 12 Column (4) for tract one-time movers, I find that children whose county is observed tend to be of higher economic status and even more so for children whose tract is observed. Positive selection is therefore a larger factor for estimates from the movers design than for those from the descriptive analysis.

11 Intergenerational Mobility in the Mid-20th Century

With these novel linkages, I present descriptive statistics measuring intergenerational elasticities and rates of upward mobility and downward mobility by race to illustrate striking differences in intergenerational mobility between Black and White children.

Intergenerational Elasticities – To measure the relationship between income inequality of one generation and the income distribution of the next, I start by estimating a simple linear equation between the log of parental income w_{0i} and the log of child income w_{1i} for the intergenerational elasticity of group g (i.e. defined by race or race by gender) as in Solon (1992). Idiosyncratic factors appear in error term ϵ_{1i} with expectation $[\epsilon_{1i}] = 0$.

$$\log(w_{1i}) = \gamma_g + \rho_g \log(w_{0i}) + \epsilon_{1i}$$

In Table A.27, I display the coefficient estimates for the whole population, separately for White and Black individuals and then separately for Black and White individuals by gender. The IGE estimate pooled across all racial groups is 0.435, indicating moderate levels of intergenerational persistence in income. This magnitude is comparable to the existing literature where the IGE has been estimated to be 0.45 for cohorts born in the early 1960s with the National Longitudinal Surveys (NLS) in Davis and Mazumder (2022). However, the estimate is lower than the remarkably high IGE of 0.8 for the 1960s and 1970s cohorts using predicted family income in Jacome et al. (2023) and lower than the IGE of 0.6 measured with the Panel Survey of Income Dynamics (PSID) in Mazumder (2015). It is also similar to the estimate of 0.45 in Chetty et al. (2014) for the cohorts of 1980-1982, although in their study, adult income of children is measured at age 30, likely leading to negative bias as suggested by Nybom and Stuhler (2016).

In Columns (2)-(7), I provide estimates by race and gender. Within race, IGE estimates tend to be lower, especially so for the Black population whose estimated IGE is 0.24. Both White women and Black women tend to have slightly higher persistence of family income, perhaps due to assortative mating, with the gender difference being larger for White individuals.

Rank-Rank Relationships – An alternative representation of intergenerational mobility is the rank-rank correlation, which was originally estimated in Dahl and DeLeire (2008) and is the approach taken by Chetty et al. (2014). Compared to the IGE, rank-rank relationships do not depend on the marginal distribution of income, where changes in income inequality by race also affect the IGE coefficient. Nybom and Stuhler (2016) further find that rank-rank correlations are less subject to the life-cycle biases of income measurement. Throughout the rest of the paper, income rank will be the main measure of adult income.

Let y_{1i} be the child's income rank in the national distribution of their cohort, and analogously let y_{0i} be the parents' income rank in the national distribution of parents in the cohort of their child. The empirical specification is as follows.

$$y_{1i} = \alpha_g + \beta_g y_{0i} + v_{1i} \tag{19}$$

In the model above, parental income y_{0i} is associated with child income y_{1i} linearly according to parameter β_g which represents relative mobility, following the terminology of Chetty et al. (2014), for persistence across generations. Conditional on parental income, the constant α_g represents absolute mobility and measures the average income rank of children when parental income rank is the lowest in the national distribution. If we assumed that these two parameters are constant over time, in the long run for groups $g \in \{White, Black\}$, β_g governs the rate of convergence to the steady state and in combination with α_g, determines the steady state gap in Black-White incomes (Becker and Tomes, 1979; Chetty et al., 2020). It is straightforward to show that if the linear rank-rank relationship held across generations, the average income rank for group g of generation t defined following $\bar{y}_{t,g} = \alpha_g + \beta_g \bar{y}_{t-1,g}$ would approach as $t \to \infty$ the steady-state value of

$$\bar{y}_{t,g} = \bar{y}_g^{SS} = \frac{\alpha_g}{1 - \beta_g}$$

The steady-state gap in average Black-White income ranks is then defined as

$$\Delta \bar{y}_{BW}^{SS} = \frac{\alpha_W}{1 - \beta_W} - \frac{\alpha_B}{1 - \beta_B}$$

with subscript B for *Black* and W for *White*. In Figure 4, I display the rank-rank correlations with the estimated coefficients and intercepts separately by race (with the estimated values and additional statistics in Table 13).

As with the intergenerational elasticities, there is less intergenerational persistence in the rank-rank correlations for Black families with $\beta_B = .217$ compared to White families with $\beta_W = .295$. Yet, this translates into upper-income Black children falling back into the lower parts of the income distribution as absolute mobility is far lower for Black children with $\alpha_B = 26.8$ and $\alpha_W = 39.8$. The Black-White gap is wider at higher parental income ranks with a 15.2 difference and 18.2 difference for children from the 25th and 75th percentiles of the parental income distribution, respectively. Assuming these values hold in the long-run, the steady-state gap can be found by intersecting the rank-rank plots with the 45-degree line and tracing the gap between the two points of intersection. In Appendix Figure B.11, I follow these steps and arrive at a gap of 22.2 ranks between Black and White income in the long-run.

While these linear relationships provide simple statistics on average income ranks, they mask heterogeneity in whether children end up lower or higher relative to their parents in the distribution of ranks. With transition matrixes between quintiles of children and parents in Panel B of Table 11, I compute the probability children from either the top or bottom quintiles of the parental income distribution end up in the top or bottom quintiles of their own adult income distribution. These matrices paint a bleak picture for economic mobility of Black children—those with parents from the top quintile are more likely to fall back into the bottom quintile (20.7%) than remain in the top quintile (17.3%). Moreover, only 3.8% of bottom quintile Black children reach the top quintile, and 44.6% remain in the bottom quintile. For White children, a contrasting pattern emerges; 40.9% of top quintile White children remain in the top quintile, while a small portion (7.2%) fall into the bottom quintile. The low rates of upward mobility for Black children combined with the concentration of parental income in the lower end of the national income distribution results in the next generation experiencing far lower income compared to White children.

Although other racial groups are not a core component of the analysis, I calculate rank-rank correlations for Hispanic, Asian, and Native American (Indigenous) children in Columns (4)-(6) of Table 13. In line with the findings from Chetty et al. (2020), Native American children experience low rates of upward mobility ($\alpha_{Indig} = 28.0$, $\beta_{Indig} = 0.26$) of similar magnitude as Black children although the rank gap widens at higher levels of parental income because of low intergenerational persistence for Black children. Absolute mobility of Hispanic children ($\alpha_{Hispanic} = 37.0$) is not substantially different from that of White children, but greater intergenerational persistence leads White children to reach higher income ranks than Hispanic children at the right tail of the parental income distribution where $[y_{1i,Hispanic}|y_{0i,Hispanic} = 100] = 61.1$ while $[y_{1i,White}|y_{0i,White} = 100] = 69.3$. Asian children (including children of immigrants) experience higher rates of absolute mobility compared to all racial groups with $\alpha_{Asian} = 49.1$ and $\beta_{Asian} = 0.238$.

Differences in household income in adulthood between Black and White children can be due to the lower levels of marriage and cohabitation with spouses among the Black population. In Figure 5, I plot rank-rank relationships by race and gender for individual income measured with W-2 earnings files rather than household income.[40] While the Black-White gap does indeed shrink (e.g. at the 75th percentile of the parental income distribution, the gap is now 14.5 ranks rather than 18.2 ranks for household income), a sizeable gap still prevails. Thus, many additional factors can be at play in determining the vast disparities in economic mobility by race.

11.1 Place-Level Variation in Black-White Outcomes

For place-based policies to be meaningful contributors to long-run outcomes, there must be a concomitant role of place for economic opportunity. In order to begin to understand the importance of place, I characterize differences in adulthood income ranks of children by race across fine geographic levels of counties and tracts in the U.S. In this section, I solely present descriptive statistics and do not claim that they represent causal effects of place. In later sections, I aim to attain estimates of the effects of place and place-based characteristics that address selection of families across locations and therefore come closer to causality.

[40] In Appendix Table A.28, I provide the rank-rank estimates by race and gender of both individual income and household income for side-by-side comparison.

Variation Across Counties – Children often move across counties and tracts during their childhood, as was summarized previously in Table 12. I therefore assign exposure weights for each child based on the number of years they reside in each county or tract. I consider childhood to be the first 23 years of life since 23 is the age when Chetty and Hendren (2018a) find that treatment effects of childhood location attenuate.

With these exposure weights, I predict income ranks of children from the 25th percentile of the parental income distribution by estimating Equation (19) for each county denoted c and pooling all cohorts born between 1964 and 1979.

$$y_{1i} = \alpha_g^c + \beta_g^c y_{0i} + v_{1i} \tag{20}$$

The mean county predicted income rank for children of race group g is then obtained for those with parents at the 25th percentile $\bar{y}_{cg,25}$ to focus on the outcomes of lower-income children. With these predicted income ranks, I present several summary statistics on variation across counties in children's outcomes weighted by population in Panel A of Table 14.[41] Across counties, there are large differences in upward mobility between the 10th and 90th percentiles of county mean predicted household income ranks. The difference is around 10 income ranks for low-income White children and around 8 income ranks for low-income Black children. When individual income rank is used as the outcome variable, the 10th to 90th percentile cross-county gap remains of similar magnitudes for both race groups. Therefore, while disparities across places are large, they cannot fully explain racial disparities as the Black-White gap is wider than place gaps between the worst (10th percentile) and best (90th percentile) counties. Indeed, in counties with the best outcomes for Black children, they still reach lower adult incomes than White children in counties with the worst outcomes in the White county distribution.

County mean predicted ranks are correlated across race, so the same locations that benefit White children also do so for Black children. Weighting counties by population and trimming the bottom and top 1% of county mean predicted ranks to eliminate outliers, I find the correlation across White and Black county mean ranks is 0.535 for household income and 0.54 for individual income. The similar magnitudes of the correlation coefficients for household income and individual income suggest that place-based factors affect both measures of income in related ways.

Variation Across Tracts – As counties often contain many tracts and racial segregation occurs across neighborhoods within counties, tracts may more closely capture the local environments that children face. I estimate Equation (20) at the tract-level to obtain neighborhood mean income ranks for low-income children with parents at the 25th percentile $\bar{y}_{ng,25}$. In Panel B of Table 14, the difference between the 10th and 90th percentiles of tract mean predicted income ranks is around 13.5 ranks for White children and 11 ranks for Black children for both household income and individual income. The correlation between mean predicted ranks across race is lower for tracts than for counties with correlation coefficients around 0.18 which can arise from true race-specific heterogeneity across

[41] Population weights also account for number of years of residence e.g. counties where children live for their entire childhood are weighted more in the county distribution than counties with the same number of children who only live there for a part of their childhood.

neighborhoods or additional noise from fewer observations at this higher resolution geographic unit.

Note that the cross-county and cross-tract variation in mean predicted ranks is larger for White children than for Black children, and this is unlikely to be due to noise as sampling variation would be greater for Black children given their smaller population count. The greater inequality across locations for White children can come from greater sorting within White households in family status or larger differences in treatment effects relative to Black children and will be a key point of interest in the estimation of causal impacts of place.

12 Historical Background on the Interstate Highway System

During this time of low upward mobility for Black children and substantial income gaps by race, several developments were occurring within cities. The central focus on this paper is on the impacts of the interstate highway system, one of the most influential place-based policies and the largest infrastructure project in the United States. Its construction aligns with the period of early childhood for the cohorts born between 1964 and 1979. In this section, I provide background on the changes associated with the interstate system that may have affected intergenerational mobility. I highlight the aspects that are within the scope of this project and others that will be left for consideration in future work.

Brief History – When the construction of the interstate network began, suburbanization into peripheral neighborhoods was already well underway. The expansion of the existing road network with high-speed limited access freeways further precipitated migration away from central areas (Jackson, 1985). With the Federal-Aid Highway Act of 1956, President Dwight D. Eisenhower authorized funding to build what would eventually become the 47,000 mile long network that exists today (Rose and Mohl, 2012). Originally, the Bureau of Public Roads estimated that $27.2 billion would be required over 10 years. By 1996, federal spending on interstate construction had reached $114 billion (approximately $500 billion in 2020 dollars). With continued expansions, such as through the Infrastructure Investment and Jobs Act of 2021, interstate development never concluded.

Transportation engineers and congressional lawmakers directed interstate roads to traverse through central business districts as congestion rose within cities. Routes that serviced the largest number of motorists were selected. The economic benefits for neighborhoods connected through interstate roads, an impact of transportation that has been studied extensively (e.g. in Faber (2014) and Duranton and Turner (2012)), motivated highway building. Consequently, suburban neighborhoods grew rapidly across the country. In search of opportunity from the sudden increase in access to employment made possible by the interstate system, predominately White households migrated outwards. A clear racial divide emerged as African American families faced discriminatory housing markets that prevented them from leaving central areas, a topic explored in the companion paper to this one (Weiwu, 2023). Neighborhoods in the center of city were thus left behind in the wake of progress in suburban areas.

In contrast to the benefits, interstate routes displaced hundreds of thousands of families and polluted the nearby environment, often in a racialized manner. Local politicians directed building through minority neighborhoods in urban revitalization programs that replaced pre-existing property with commercial development. While the displacement caused by interstate highways may have affected intergenerational mobility, the panel dataset of this paper starts in 1974, past when most segments of the interstate system were built. Future datasets that extend individual migration history to the 1960s will allow for studying the long-run consequences of urban renewal.

The Federal-Aid Highway Act of 1973 was passed to limit the negative auxiliary effects of highways. This legislation increased the role of local decison-making to modify highways in response to political activism or environmental opposition. It also increased funding for alternative modes of transport such as mass transit systems.

Addressing Selection in Placement – Taking into consideration the non-exogenous placement of interstate routes, I follow several approaches to obtain cleaner variation in highway impacts. First, I account for factors that influenced where highways were eventually located. To address traffic and minimize costs of construction, federal engineers recommended that interstate development occur through "the improvement of a limited mileage of the most heavily traveled highways" in the report *Interregional Highways*. I thus digitize historical urban roads for 71 cities from Shell Atlases from 1951-1956 as possible candidates for highway routes and control for their location in the empirical specification. Other geographic factors that affected highway placement such as the location of historical railroad networks, canals, and steam-boat navigable rivers for the late 19th century are retrieved from Atack (2015, 2016, 2017) and bodies of water, shores, and ports from Lee and Lin (2017).

Second, I construct two sets of instruments for highway location. I digitize interregional routes in a 1947 plan of the interstate system from Baum-Snow (2007) at a finer spatial scale. As the geographic unit of this project is more granular than in that study, I obtain maps of within-city plans from the 1955 *General Location of National System of Interstate Highways* (also referred to as the "Yellow Book"), which were previously used in Brinkman and Lin (2022). I digitize these intra-city maps for 100 cities and combine them with the 1947 plan into a single network of planned routes. Interstates were designed to intersect the central business districts of major cities. I therefore further construct an Euclidean ray spanning network to connects cities in the planned maps, a strategy that is similar to the "inconsequential units" approach of Chandra and Thompson (2000).

An example of the various networks for the Baltimore Area is depicted in Figure 6. As is visible in these maps of Baltimore, planned routes and Euclidean rays trace the general direction of interstate highways, and highway routes replaced existing large roads in many cases.

13 Mapping Policy into Economic Mobility: A Theoretical Framework

With this background on the channels through which interstate highways impacted neighborhoods, I now provide a general framework for mapping place-specific changes into individual outcomes in the aggregate. As suggested in the historical literature, the neighborhoods affected were not simply the ones directly targeted by transportation infrastructure. Reallocation in response may have led some areas (in this case, suburban ones) to improve in economic status at the expense of others (in this case, central areas). This reallocation further changes which individuals are exposed to the neighborhood-level impacts of infrastructure policy. Altogether, it is unclear if in aggregate, individual outcomes were improved.

Moreover, there were likely highly heterogeneous impacts given the initial spatial distribution and differential mobility of various groups of children. The spatial sorting in response can be an additional indirect effect that translates into long-run outcomes if local peer composition matters for economic mobility. In this setting, the migration of White households out of the central city increased the isolation of Black households, and racial segregation has been previously measured to be an important determinant of children's outcomes (Wilson, 1987; Massey and Denton, 1993; Ananat, 2011; Chyn, 2018; Chyn and Daruich, 2022). The strength of these forces is an empirical question that requires estimates of treatment effects as well as a quantitative question that combines these estimates with additional structure for a general equilibrium assessment.

I lay out a simple spatial framework that captures neighborhood-level changes from interstate development and delineates the set of parameters needed for an aggregate quantification of intergenerational impacts. With its general structure, it can further be applied to other place-based policies that include peer sorting as a local spillover for long-run outcomes.

13.1 Aggregate Consequences of Place-Based Policies on Intergenerational Mobility

In this study, children's adult incomes are the ultimate outcome of interest. In other studies where the central objective is measuring the aggregate impact of policy interventions on output or consumer welfare, the model has immediate normative implications. With the modified objective of this paper, the model instead serves a different purpose of predicting how parents choose residential locations. These choices then have subsequent consequences for children's long run outcomes.

I consider the impacts across heterogeneous groups of children. Children of different demographic types are denoted with the subscript g for race and economic status i.e. parental income percentile. Children's adult incomes y_i are a group-specific function $f_g(\cdot)$ of the characteristics of the neighborhood they reside in n, individual covariates X_i, and idiosyncratic factors ϵ_i. Let S_g be the set of children in group g so that $|S_g|$ is the size of the set. The average income of children in group g is defined as

$$\bar{y}_g = \frac{1}{|S_g|} \sum_{i \in S_g} y_i = \frac{1}{|S_g|} \sum_{i \in S_g} f_g(\mathbf{x}_{n(i)}, X_i, \epsilon_i)$$

The vector of neighborhood characteristics $\mathbf{x}_n$ in this setting includes average income of residents, which can be impacted by the interstate system connecting workers to different locations of employment. It also includes peer composition, such as the percentage White of the population, which can change if there is differential sorting in response to policy shocks.

While children's outcomes are determined at the individual level, I aggregate to the neighborhood level to clarify how place-specific shocks affect outcomes. To do so, I specify a linear function for $f_g(\cdot)$, as is typically done in the literature, where income is the following

$$y_i = f_g(\mathbf{x}_{n(i)}, X_i, \epsilon_i) = \alpha_g + \mathbf{x}_{n(i)}\beta_g + X_i\gamma_g + \epsilon_i$$

I partition the set of children in S_g into the neighborhoods they live in for $n = 1, \ldots, N$ such that $S_g = \{S_{g1}, \ldots, S_{gN}\}$. Average income can then be re-formulated as an aggregator of neighborhood characteristics with neighborhood shares as weights.

$$\begin{aligned}
\bar{y}_g = \frac{1}{|S_g|}\sum_{i \in S_g} f_g(\mathbf{x}_{n(i)}, X_i, \epsilon_i) &= \frac{1}{|S_g|}\sum_{i \in S_g}\left(\alpha_g + \mathbf{x}_{n(i)}\beta_g + X_i\gamma_g + \epsilon_i\right) \\
&= \sum_{n=1}^{N} \frac{|S_{gn}|}{|S_g|} \cdot \frac{1}{|S_{gn}|}\sum_{i \in S_{gn}} \left(\alpha_g + \mathbf{x}_n\beta_g + X_i\gamma_g + \epsilon_i\right) \\
&= \sum_{n=1}^{N} \pi_{ng}\left(\alpha_g + \mathbf{x}_n\beta_g + [X_i | i \in S_{gn}]\gamma_g + [\epsilon_i | i \in S_{gn}]\right) \\
&= \sum_{n=1}^{N} \pi_{ng}(\mathbf{x}_n\beta_g) + \alpha_g + [X_i]\gamma_g \qquad \text{with } [\epsilon_i] = 0
\end{aligned}$$

In the notation above, π_{ng} is the share of children from group g living in n. With this expression for average income, it should be clear that the only relevant factors in assessing the impact of a place-based policy is how it changes where children live across neighborhoods (π_n) and how it changes neighborhood characteristics ($\mathbf{x}_n$). These characteristics can further be a function of where children of different groups live (i.e. racial composition). Suppose that $\mathbf{x}_n$ is of length K so there are K neighborhood characteristics. A general shock represented by δ transmits into average child income with the following approximation

$$\frac{d\bar{y}_g}{d\delta} = \sum_{n=1}^{N} \underbrace{\frac{d\pi_{ng}}{d\delta}}_{(1)} \underbrace{(\mathbf{x}_n \beta_g)}_{(3)} + \pi_{ng} \sum_{k=1}^{K} \underbrace{\frac{d\mathbf{x}_{n,k}}{d\delta}}_{(2)} \tag{21}$$

Sufficient statistics that are not immediately observable from data are thus (1) the change in residential shares from the shock, (2) the change in characteristics from the shock, and (3) the causal impact of neighborhood characteristics on children. In the next section, I provide model structure to compute (1) and (2) across neighborhoods, and these predictions are specific to each place-based shock. The parameters β_g for (3) the treatment effects of neighborhood characteristics on children are not specific to interstate highways and are of general interest in labor economics and to policy-makers. To preview later results, in Section 14, I provide empirical evidence on (1) and (2) in response to the interstate development. In Section 15, I estimate (3) using a movers design for families that move across origins and destinations with different characteristics.

13.2 A Spatial Model of Neighborhoods

In this framework, I proceed by building on standard quantitative spatial models with commuting networks as described in Allen and Arkolakis (2014); Ahlfeldt et al. (2015); Tsivanidis (2022). Individuals in the model are the parents of children differentiated by group g. Neighborhoods are indexed by $n = 1, \ldots, N$, and each city contains fixed population levels of each group $\mathbb{L}_g$. Parents choose which residential neighborhood to live n and which workplace to work at m depending on residential amenities (B_{ng}), housing prices (Q_n), wages (w_{mg}), and commute costs ($d_{nmg} = t_{nmg}^{\kappa_g}$) after receiving an idiosyncratic shock for residential locations and an idiosyncratic shock for workplaces. An elastic housing construction sector responds to changing housing demand across neighborhoods. In equilibrium, housing markets clear to determine residential populations, housing prices, and welfare for all workers.[42]

Individual i's utility is represented as

$$\max_{c_{nm}(i), l_n(i)} \quad \frac{z_n(i)\epsilon_m(i)B_{ig}}{d_{nmg}}\left(\frac{c_{nm}(i)}{\beta_g}\right)^{\beta_g}\left(\frac{l_n(i)}{1-\beta_g}\right)^{1-\beta_g}$$
$$\text{s.t.} \quad c_{nm}(i) + Q_n l_n(i) = w_{mg}$$

and after utility maximization, indirect utility is expressed following

$$u_{nmg}(i) = \frac{z_n(i)\epsilon_m(i)B_{ng}Q_n^{\beta_g-1}w_{mg}}{d_{nmg}} \tag{22}$$

Beyond group-level factors across spatial units, workers have idiosyncratic preferences for residences $z_n(i)$ and idiosyncratic preferences for workplaces $\epsilon_m(i)$ that affect their location choices. Residential idiosyncratic shocks $z_n(i)$ are drawn from a Frechet distribution $F(z_n(i)) = \exp(-z_n(i)^{-\theta_g})$ where θ_g is a shape parameter that captures the dispersion of shocks and how responsive individual choices are to changes in the attractiveness of each residential location. θ_g can be heterogeneous by group. Idiosyncratic workplace shocks $\epsilon_m(i)$ are also distributed Frechet from $F(\epsilon_m(i)) = \exp(-T_{mg}\epsilon_m(i)^{-\phi})$ where ϕ similarly determines the dispersion of shocks and the responsiveness of workplace choices to employment location changes. Lastly, T_{mg} is a scale parameter that affects the attractiveness of a workplace, for example through amenities, beyond wages paid to workers.

This expression for indirect utility highlights how residential choice is determined by observable place characteristics and idiosyncratic household factors. In the empirical section, I will return to this expression and note how structural features of the model translate into empirical features in the identification strategy.

Following that $\epsilon_m(i)$ is distributed Frechet, conditional on living in n, the probability a worker works in m is

$$\pi_{mg|n} = \frac{T_{mg}(w_{mg}/d_{nmg})^{\phi}}{\sum_l T_{lg}(w_{lg}/d_{nlg})^{\phi}} = \frac{T_{mg}(w_{mg}/d_{nmg})^{\phi}}{\Phi_{ng}} \tag{23}$$

The denominator Φ_{ng} is a transformation of commuting market access (CMA) following $CMA_{ng} = \Phi_{ng}^{1/\phi}$ where for location n, higher wages w_{mg} (with the scale parameter T_{mg}) and lower commute costs

[42] In this set-up, firms are in a separate commercial housing market that does not interact with the residential housing market. Therefore, labor supply changes across workplaces do not impact residential housing prices or the allocation of housing supply between residential and commercial uses. Wages across locations are also fixed and do not respond to labor supply. This last assumption implies the model environment is only in partial equilibrium.

d_{nmg} from m increase CMA.

The probability a worker of group g lives in n follows a similar form using the properties of the Frechet distribution for residential shocks.

$$\pi_{ng} = \frac{(B_{ng} CMA_{ng} Q_n^{\beta_g - 1})^{\theta_g}}{\sum_t (B_{tg} CMA_{tg} Q_t^{\beta_g - 1})^{\theta_g}} \tag{24}$$

Neighborhoods with greater group-specific amenities, higher CMA, and lower housing prices are the locations the population of a group will more likely reside in. The residential population in n combines the probability above with the total population of a group in a city $\mathbf{L}_g$.

$$L_{ng} = \pi_{ng} \mathbb{L}_g \tag{25}$$

Housing – To close the model, residential housing markets must clear. The housing supply curve as a function of housing prices is of Cobb-Douglas form. For housing supply to meet demand, expenditures by families on housing should equal the amount of housing available for purchase. These two statements imply the following equations.

$$H_n = \left(\frac{1-\mu}{\mu}\right)^{\frac{1-\mu}{\mu}} K_n Q_n^{\frac{1-\mu}{\mu}} \tag{26}$$

$$Q_n H_n = \sum_g (1 - \beta_g) \overline{w}_{ng} L_{ng} \tag{27}$$

Welfare – Welfare of group g in a city aggregates over all neighborhoods and accounts for each location's amenities, commuter market access, and rental prices. It is defined as U_g.

$$U_g = \left[\sum_n \left(B_{ng} CMA_{ng} Q_n^{\beta_g - 1}\right)^{\theta_g}\right]^{1/\theta_g} = \left[\sum_n \left(B_{ng}\left(\sum_m T_{mg}(w_{mg}/d_{nmg})^{\phi}\right)^{1/\phi} Q_n^{\beta_g - 1}\right)^{\theta_g}\right]^{1/\theta_g} \tag{28}$$

Equilibrium – Given the model's parameters $\{\beta_g, \theta_g, \kappa_g, \phi, \mu\}$, city populations by group $\{\mathbb{L}_g\}$, and location characteristics $\{T_{mg}, t_{nmg}, B_{ng}, K_n\}$, the equilibrium is represented by the vector of endogenous objects $\{L_{ng}, Q_n, U_g\}$ determined by the following equations:

1. *Residential populations* in each neighborhood (25)
2. *Housing supply and demand* (27)
3. *Closed City* where $\sum_i L_{ng} = \mathbb{L}_g$

13.3 Model Predictions for Neighborhood Characteristics

Residential Shares and Sorting – Improvements in CMA can lead heterogeneous migration responses if the residential preference elasticity θ_g differs by group as $\frac{d \log L_{ng}}{d \log CMA_{ng}} = \theta_g$. With the expression for residential shares in Equation (24), the model provides a means for obtaining (1) the change in residential shares in response to changes in commuting access from the interstate highway system. It also characterizes changes in peer composition across neighborhoods if peer composition is defined as the share of the population that is of a particular group. For example, racial composition represented as percentage White can be easily calculated as

$$pctWhite = \frac{LnW}{L_n}$$

where $L_n = \sum_g L_{ng}$. Given that segregation across neighborhoods is a characteristic that influences children's outcomes, the structure of the model has thus determined one component of (3) the change in place characteristics from the interstate highway system.

Expected Income – Expected income is another characteristic of each neighborhood that corresponds to a prediction of the model. Using the equation for conditional commuting shares in Equation (23), neighborhood average income aggregates across workplaces and the wages received in those locations.

$$\bar{w}_{ng} = E[w_{mg}|n] = \sum_m \pi_{mg|n} w_{mg} = \sum_m \frac{T_{mg}(w_{mg}/d_{nmg})^{\phi}}{\sum_s T_{sg}(w_{sg}/d_{nsg})^{\phi}} w_{mg} \tag{29}$$

Note that this expression is closely related to CMA, which is also an aggregator over workplaces. However, within the aggregation, there is an additional weight from wages w_{mg} divided by CMA.

14 Empirical Evidence of Highway Impacts on Neighborhoods

14.1 Decennial Census Data

To measure job access, I use microdata from the Decennial Censuses in 1960 and 1970 to create neighborhood and workplace level aggregates. Neighborhoods are represented by Census tracts which have populations of around 4,000 people, and for each tract I retrieve population by race. Since tracts are constantly being re-defined over time, I create consistent tract definitions with the Longitudinal Tract Database. The Decennial Censuses starting in 1960 reported place of work for the county and city, which I use to create a workplace geographic unit called a Place of Work Zone from the intersection of the two geographies. Wages and employment for workplaces are then measured by race. Job access requires data on commuting across neighborhoods and workplaces which I generate using digitized maps of the interstate highway system with dates of construction and historical urban roads. Commute time matrixes are calculated with ArcGIS Network Analyst for 25 of the largest U.S. cities for commuting in 1960 and 1970 where constructed segments of the interstate network are overlayed on the historical road network. I also collect various other geographic data on planned engineering maps of highways, natural features, and historical canals and railroads to use as controls and obtain quasi-random variation in highway placement.

Lastly, I create measures of segregation, employment, and educational intergenerational mobility with the full-count 1940 census to conduct placebo tests of the highway variation. In 1940, most children completed their education before leaving home which allows me to measure children's educational attainment conditional on their parents following the work of Card et al. (2018). In a single Census without linkages over time, intergenerational mobility can be calculated.

Summary Statistics of Neighborhood Characteristics - In Table A.29, I present summary statistics on characteristics for counties and tracts with the 1970 Decennial microdata. The long-form 15% sample of the Census is the main source for measuring neighborhood characteristics that impact children's outcomes. To provide a sense of whether White and Black children experience different levels of neighborhood characteristics on average, I weight the place-level characteristics with group-specific population levels. The weighted averages are provided in Columns (1) and (3). At the county-level, racial composition does not differ greatly between the White and Black averages. This muted difference can be a result of the greater degree of segregation across neighborhoods within counties. In Panel B, tract-level averages are presented, and the difference in racial composition for the tracts where White and Black households live is substantially larger. White households live in tracts that are approximately 96% White while Black households live in tracts that are 65% White. The Black population also lives in neighborhoods that have fewer high-occupational status individuals. On average, the percentage in the top quintile of occupation scores for their tracts is 8%. Versus for the White population, the corresponding number is 12%. Black households further live in tracts with lower average income ($42,000) compared to White households ($53,000).

14.2 Measurement of Job Access

Job access is characterized as a specific case of commuter access measures micro-founded off the quantitative urban model presented previously. Let n be the residential neighborhood at the tract level and m be the workplace location. Job access from a neighborhood n aggregates over all workplaces $m \in \{1, \ldots, M\}$ with the two connected by commute costs d_{nm}.

$$JMA_n = \sum_m \frac{w_m L_m}{d_{nm} \mathbf{L}}$$

In the summation above, wages w_m at workplace m are discounted by the commute costs d_{nm} which follow the functional form $d_{nm} = \exp(t_{nm})$ with t_{nm} being the commute time on the road network. It also include the share of employment at workplaces $\frac{L_m}{\mathbf{L}}$ so locations with more employment are given greater weight in the job access measure. To nest JMA under the definition of CMA from the model section, labor supply elasticity $\phi = 1$ and the weighted within the aggregator $T_m = \frac{L_m}{\mathbf{L}}$.

Exogeneity of Job Access Induced by Interstate Highways - To obtain a more exogenous form of job access, in the measure above, I set wages and employment for the workplace to 1960 levels and commute times to 1970 levels from the construction of the interstate highway system.

$$JMA_{n,HW} = \sum_m \frac{w_{m,1960} L_{m,1960}}{d^{HW}_{nm,1970} \mathbf{L_{1960}}}$$

As the highway shock is not completely random, I create two additional instruments where the change in commute costs comes from the construction of the planned network or the Euclidean rays. For the instruments JMA^{YB}_n and JMA^{Rays}_n respectively, I replace $d^{HW}_{nm,1970}$ with d^{YB}_{nm} and d^{Rays}_{nm} (YB is an abbrevation for the Yellow Book planned routes).

14.3 Estimating the Relationship between Place Characteristics and Job Access

I present results indicating that job access is correlated with employment at the tract-level in the cross-section of the 1970 Decennial Census. To measure the impacts of place on children's outcomes, I later employ cross-sectional variation in place characteristics. Therefore, I provide evidence that job access is correlated with place characteristics using cross-sectional differences in the interstate network to exogenously shift job access. For quasi-random variation in interstate placement, I also instrument highway locations with the planned route and Euclidean ray network.

The estimating equation correlates tract-level average income in 1970 with JMA while including several geographic controls in X_n.

$$\log(avgincome_{n,1970}) = \alpha + \beta \log JMA_{n,HW} + \mathbf{X}_n \zeta + \nu_n$$

The geographic controls in X_n account for the non-random placement of the interstate highway system and are distance to the central business district, large historical urban roads, rivers, lakes, shores, ports, historical railroads and canals. With fixed effects at the city-level, the empirical variation is only across tracts within metropolitan areas. I present OLS results in Table A.30 Column (1) where I find that average income is strongly correlated with JMA. I estimate the same specification for the additional economic variables of employment rate and labor force participation rate, and I find strong relationships with JMA for those characteristics as well. In Panel B, I instrument JMA in 1970 with a modified version where wages and employment come from the 1960 census. This modified JMA is more exogenous as it does not include the endogenous adjustment of wages and employment that may be correlated with neighborhood average income. I further include instrumented results where the interstate network is replaced with the planned and euclidean ray network in Panels C and D. Across all these specifications, the relationship between log average income and log JMA follows the same qualitative pattern. With the planned and ray network instruments, the coefficient increases to almost twice the magnitude of the OLS estimate.

To show the cross-sectional relationship is not entirely driven by the non-randomness of the commuting network, I estimate a similar equation over time in a long-difference from 1960 to 1970.

$$\Delta \log(avgincome_n) = \rho + \gamma \Delta \log JMA_{n,HW} + \mathbf{X}_n \omega + \mu_n$$

The OLS results are shown in Table A.31 and plotted in Figure 7 where I find that changes in JMA are correlated with increases in tract-level average income. In Panel B, I set wages and employment in JMA to 1960 levels and only allow for commute time changes between 1960 and 1970 from the interstate network to form the instrument. The coefficient remains the same magnitude, so wage and employment adjustments at workplaces are not driving the results. In Panels C and D, I conduct the same procedure but replace the changes in commute times with those from the planned and ray networks. The coefficient increases in magnitude to about twice the size, and this increase is likely due to negative selection on trends in interstate placement. In Columns (2)-(3), I present results for changes in the employment rate and labor force participation rate and find positive relationships with changes in JMA as well.

14.4 Estimating Sorting on Improvements in Job Access

As evidence for how the interstate system can alter local spillovers through peer externalities, I turn to measuring whether there is differential sorting by groups of varying demographic and economic status. With differential migration, there would mechanically be changes in peer composition in response to changes in commuting access from the interstate network.

Taking logs and first differences of Equations 24 and 25, I obtain the estimation equation below.

$$\Delta \log L_{ng} = \alpha_g + \theta_g \Delta \log JMA_{ng} + \mathbf{X}_n \eta + \epsilon_{ng} \tag{30}$$

Additionally, the specification above includes controls in $\mathbf{X}_n$ for the previously discussed geographic features as well as city fixed effects. The first difference is over 1960 and 1970 using Decennial restricted data as with the earlier estimating equation.

I split the microdata along many dimensions to look at who responds to improvements in JMA. Groups g are separated by race, by education, and by occupation quintile. Occupation quintiles are calculated using occupation income scores (the median of the national income distribution for each occupation). Since income can change as a result of JMA, changes in population responses by income can reflect changes in access to workplaces (i.e. not migration but the same individual increasing the income they receive). In this section, I would like to isolate changes from population movements. As occupation is less variable within an individual, population responses by occupational quintiles more closely reflects sorting.

I present the elasticities for each group in Figures 8. Within race, they are not large differences in population elasticities to JMA by education. Among White households, the more educated are slightly more mobile. Although the standard errors are large for the Black population, I find they are less mobile than the White population with small differences by education. Higher occupational status households are also more responsive to job access improvements. These results all point to more advantaged populations responding to a greater extent to access to employment.

Shroder (2002) finds that take-up in the Moving To Opportunity (MTO) project is low in areas with tight housing markets and is higher for those who are more educated. This finding is related to the differential sorting documented above as the response to JMA depends on how spatially mobile households are across neighborhoods. While general equilibrium effects from relocating families were a concern raised in the MTO literature, because of the small scale of the program, few neighborhoods experienced large enough inflows of disadvantaged families to meaningful alter their characteristics (Ludwig et al., 2013). Because the interstate highway system was a large shock that induced substantial migration, equilibrium effects are likely important for their impact on economic opportunity.

14.5 Estimating the Relationship between Parental Income and Job Access

If parental income changes, then assessing the impacts of a place-based policy becomes more complicated. In Equation (21), a place-based shock does not impact individual level characteristics such as parental income. Further, parental income can transmit into residential location choice, thereby being another second-order effect of the place-based shock. I test whether parental income changes in the following fixed effects regression that uses a panel of parent movers and looks at whether parental income changes as they move into areas with greater job access.

$$p_{i,t} = \alpha_i + \alpha_t + JMA_{n(i,t),HW} + \mathbf{X}_{n(i,t)}\eta + \xi_{i,t}$$

In this specification, $JMA_{n(i,t),HW}$ captures changes in job access of where parents live with time subscript t (JMA of locations is fixed to 1970 levels so the time variation comes entirely from moves). I control for other characteristics of the locations they are living in $\mathbf{X}_{n(i,t)}$ such as geographic controls previously mentioned. As with JMA, the time variation is only through changing location as the geographic variables are time-invariant.

15 Place Characteristics and Economic Mobility

An important set of parameters for measuring the impacts of place-based policies on economic mobility are coefficients for the relationship between characteristics affected by these policies and later life outcomes. Previously, I measured how interstate highways impact average income and the peer composition of neighborhoods. I now estimate how these characteristics are related to children's adult incomes.

15.1 Descriptive Correlations Between Place Characteristics and Economic Mobility

Previous research has documented that place effects vary greatly across locations with some locations causally leading to better outcomes. I study the *mechanisms behind why* some places contribute to improved outcomes following that Chetty and Hendren (2018a) find much of county level place effects can be explained by observable characteristics. I focus on the characteristics predicted by the model to change with the interstate highway system—average income and peer composition (racial composition, educational composition, and occupational composition).

I begin with descriptive relationships between these characteristics and children's adult outcomes. In Figure B.12, I display the correlations between one standard deviation change in each tract-level characteristic and the change in predicted family income rank. The outcomes of White and Black children are strongly associated with average income, educational and racial composition where the magnitudes of the coefficients are larger for White children. [43]

However, much of these associations can be driven by selection as sorting of households would lead to the same results. More advantaged families may choose higher income neighborhoods with higher status peers, and their children would fare better in adulthood absent any treatment effects from place. This selection can be a larger force for White households who, as measured in the migration response to the interstate highway system, respond more to differences across neighborhoods.

[43] Additional results at the county-level are also available in Appendix Figure B.13. In both of these figures, I present results for the economic characteristics of employment rate and labor force participation rate that are not directly predicted by the model. Previous work on spatial mismatch has suggested that employment connectivity affected the employment prospects of Black adults, which I study with the latter set of characteristics (Wilson, 1987).

Their higher mobility suggests they select to a greater extent into neighborhoods perceived as beneficial for their children, leading to a stronger association between place characteristics and children's adult outcomes.

Another source of bias arises from the correlation between neighborhood characteristics and other omitted variables. For example, neighborhoods with higher average income or a greater percentage of White families tend to vary along many dimensions such as crime levels, racial attitudes, and social networks that are harder to observe precisely by researchers. These other factors may be downstream of changes in income or racial composition and be considered auxiliary effects of these characteristics. However, if e.g. neighborhood income is not driving the differences in outcomes for children but rather racial attitudes correlated with income, the correlations would not be informative for the treatment effect of increasing neighborhood income.

In light of these identification challenges, in later sections I turn to more complex research designs to estimate the causal impacts of place characteristics on children. I implement a movers design to address selection in neighborhood choice where I estimate treatment effects for children who move along the dimension of the neighborhood characteristics of interest. To address bias from omitted variables, I employ the structure of the quantitative model to construct shifters for neighborhood characteristics. With tract-level variation from the interstate highway system, I predict changes in average income and migration of different types of households, which alters the peer composition of neighborhoods. I then exploit this empirical variation to assess how changes in income and peer composition impact children's outcomes.

15.2 Movers Exposure Design to Address Selection in Place Effects

In a movers exposure design that builds on previous work by Chetty et al. (2020), children vary in the amount of time exposed to characteristics of place depending on the age at which they move, assuming that age at move is quasi-random. The motivation for the movers design comes from the observation that families do not randomly choose the neighborhoods they live but there are many idiosyncratic factors that can push families to move. The non-random drivers of their choice leads to selection and therefore challenges in estimating treatment effects of place. By exploiting the idiosyncratic factors behind changes in location, it is then possible to estimate exposure effects for location. The spatial model has the same features of idiosyncratic shocks as well as the characteristics of places affecting neighborhood choice where the extent to which location choice is determined by idiosyncratic versus place-level features is determined by the distribution of the Frechet shocks.

I first present the basic mechanics behind the movers design before extending it to the particular application of this paper. Let i denote each child, p_i be their parental income rank, and r_i be their race. The sample focuses on the set of children who move once during their childhood until up the age of 28. Let m_i be age at move from origin neighborhood o to destination neighborhood d. In this specification, I examine moves across counties. Let $\bar{y}_{pcr}$ be the the exposure-weighted outcome of y_i (child household income rank) for children of race r who grew up in location c with parental household income rank p.[44] These county-level average predicted income ranks serve as a measure

[44]These predicted child income ranks do not include one-time movers to ensure that a child's own outcome does not

of neighborhood quality.

I measure how children incomes in adulthood vary depending on the length of time spent in counties where the average child of the same race group fares better in adulthood. Let $\Delta_{odpr} = \bar{y}_{pdr} - \bar{y}_{por}$ be the predicted difference in household income ranks in the destination versus origin county for children of race r and parental income rank p. I regress the income rank of children who move on the change in origin and destination quality interacted with age-at-move fixed effects separately for each race.

$$y_i = \sum_{s=1964}^{1979} I\{s_i = s\}(\lambda_s^1 + \lambda_s^2 \bar{y}_{por}) + \sum_{m=1}^{28} I\{m_i = m\}\phi_m + \sum_{m=1}^{28} b_m I\{m_i = m\}\Delta_{odpr} + \epsilon_{1i} \tag{31}$$

In this specification, I include age-at-move fixed effects in ϕ_m to capture disruption effects that can differ with age of the child. I also include cohort fixed effects and their interaction with the origin income rank in $(\lambda_s^1 + \lambda_s^2 \bar{y}_{por})$ to account for differing outcomes across cohorts and how families coming from higher income areas tend to have better outcomes (controlling for selection cross-sectionally at origin locations).

The key parameters of interest are the b_m coefficients, which capture how children's outcomes vary with the age at which they move to an area with higher or lower predicted earnings. To increase the power of the coefficient estimates, I make the parametric assumption of linearity before and after cutoff of age 23 and combine the estimated coefficients for the age bins before and after age 23. The specification is then the following

$$\begin{aligned} y_i = &\sum_{s=1964}^{1979} I\{s_i = s\}(\lambda_s^1 + \lambda_s^2 \bar{y}_{por}) + \sum_{m=1}^{28} I\{m_i = m\}\phi_m \\ &+ I\{m_i \leq 23\}(\rho + (23 - m_i)\gamma)\Delta_{odpr} + I\{m_i > 23\}(\delta^1 + (23 - m_i)\delta^2)\Delta_{odpr} + \epsilon_{2i} \end{aligned} \tag{32}$$

where as in the above specification, age at move fixed effects and cohort fixed effects interacted with origin predicted income rank are included.[45] The coefficient of interest is γ for the exposure effect for each year spent in the destination location up until age 23. Exposure to the treatment after age 23 δ^2 is presumed to be zero, and I test for this result in the estimation. The intercept term δ^1 is the correlation between the difference in quality of origin versus destination locations and children outcomes who move to the destination at age 23. Because it is assumed that treatment effects end at this point, any correlation would signal selection in the choice of destination neighborhood relative to the origin neighborhood.

I present the results in Figure 9 for White and Black boys where the age at move coefficients are presented in the scatter plot. In Figure 9, linear lines fit the estimated coefficients for the age at move bins before and after age 23 and indicate that the exposure effects per year differ by race with stronger exposure effects for Black boys. These results are consistent with previous work that finds there are substantial treatment effects from length of time spent in better places and that treatment

enter the definition of neighborhood quality. These exposure-weighted income ranks are estimated following Equation 20. The exposure weights that are used for predicting income rank are constructed from residential location up until age 23.

[45] I further include origin and destination fixed effects and family fixed effects to test additional selection in robustness checks.

effects are greater for Black children. Sorting across locations additionally differs by race with the intercept term at age 23 being larger for White children compared to Black children. These results are in line with the greater mobility of White households. Related to the descriptive evidence, the stronger sorting of White households can be a large contributor to the associations observed in the descriptive correlations. While the magnitude of the correlations are larger for White children, the estimated treatment effects of neighborhood quality are lower which is again consistent with greater selection by White families.

Estimates for the linear parametric specification are presented in Table 15 for both boys and girls by race. The findings for boys largely mirror those from the age bins (with weights at the individual level rather than across the age bins) with similar values for the coefficients for exposure effects. For girls, the estimated exposure effects tend to be smaller in size with insignificant differences by race. These results align with the findings in Chetty et al. (2020) which show that boys, especially Black boys, tend to be more influenced by their neighborhood environment.

15.3 Extending the Movers Design to Study Characteristics Behind Place Effects

The movers design presented above can be extended to study the impact of neighborhood characteristics on children and address selection at finer spatial scales. Previous studies often estimate place effects for each location and then project these place effects on neighborhood characteristics to understand how much of the variation can be explained by any one feature. At the tract-level, this approach becomes intractable given the small sample of movers. This challenge is particularly acute by race given the limited number of observations for Black children (829,000 as shown in Table 12) for the 40,000 tracts in the U.S.

The specification I estimate for the extension is similar in form to Equation 32. Instead of studying moves across locations with different average predicted income ranks, I study moves along each neighborhood characteristic. Let $\Delta^x_{od} = x_d - x_o$ where x is average income and peer composition at the tract-level. I follow the linear exposure estimating equation over age at move and suppress displaying the list of controls and fixed effects by placing them in the vector $\mathbf{X}_i = \sum_{s=1964}^{1979} I\{s_i = s\}(\lambda^1_s + \lambda^2_s x_o) + \sum_{m=1}^{23} I\{m_i = m\}\phi_m$.

The estimating equation is then

$$y_i = (\rho^x + (23 - m_i)\gamma^x)\Delta^x_{od} + \beta\mathbf{X}_i + \epsilon_{3i}$$

where the vector of controls $\mathbf{X}_i$ can include additional location-specific controls to remove omitted variables bias from factors correlated with neighborhood characteristics. These controls will become more important as the highway variation is employed to provide shifters for the neighborhood characteristics.

With the equation above, I estimate the coefficient ψ^x for a one standard deviation difference in the same set of neighborhood characteristics as studied in the descriptive correlations. I present the results in Figure 10 and find that there are significant treatment effects for the causal impacts of tracts (addressing selection with the movers design) and the characteristics of average income, racial composition, educational composition and occupation composition.[46] The treatment effects of length of exposure to each characteristic tend to be similar by race. The signs of the treatment effects tend to follow the descriptive correlations with positive effects for average income and the percentage of the neighborhood that is White and higher-educated. Children experience worse adult outcomes in neighborhoods with lower occupation status households.

In Figure 10, I additionally provide the intercept term for the degree of selection by families on these characteristics. I find that selection is stronger for White families, which is consistent with the descriptive correlations being larger in size for White families despite the treatment effects of tracts being the same by race. Selection is not limited to White children since I also find there is considerable selection for Black children.

16 Aggregate Effects of Highways on Children's Outcomes

In future work, I will solve for the full predicted migration responses and change in neighborhood characteristics from the equilibrium model in response to interstate development. Returning to the equation for aggregate consequences on children's outcomes

$$\frac{d\bar{y}_g}{d\delta} = \sum_{n=1}^{N} \underbrace{\frac{d\pi_{ng}}{d\delta}}_{(1)} \underbrace{(\mathbf{x}_n \beta_g)}_{(3)} + \pi_{ng} \sum_{k=1}^{K} \underbrace{\frac{d\mathbf{x}_{n,k}}{d\delta}}_{(2)} \tag{33}$$

each piece for the change in aggregate income is now defined. In Equation 24, the change in residential location is an equilibrium outcome of the model that corresponds to the term of (1) in the above expression. The differential change in migration across groups predicted by the model with the residential elasticities estimated in Equation 30 then translates into changes in peer composition in the term (2). Average income is predicted by the model in Equation 29 and is another characteristic of locations altered by the interstate highway system for (2). Lastly, the coefficients for the treatment effect of these characteristics in (3) are the ψ^x coefficients estimated in the movers design.

17 Conclusion

In this paper, I study the impacts of a far-reaching infrastructure project on economic mobility using rich historical linkages built through the data infrastructure at the Census Bureau (Stinson and Weiwu, 2023). I explore the mechanisms behind how this particular policy impacts the features of locations, both those directly targeted by highway infrastructure and those indirectly affected through general equilibrium effects. These changes at the neighborhood level then translate into lasting intergenerational consequences for children who are impacted by how the features of places are altered by infrastructure policy and their individual exposure to these features through migration responses.

[46] Additional results at the county-level are provided in Figure B.14.

Bibliography

Aaronson, D., Hartley, D., and Mazumder, B. The Effects of the 1930s HOLC ”Redlining” Maps. *American Economic Journal: Economic Policy*, 13(4):355–392, 2021. doi: 10.1257/pol.20190414. URL https://www.aeaweb.org/articles?id=10.1257/pol.20190414.

Abramitzky, R., Boustan, L. P., and Eriksson, K. Europe’s Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review*, 102(5):1832–1856, 2012. doi: 10.1257/aer.102.5.1832. URL https://www.aeaweb.org/articles?id=10.1257/aer.102.5.1832.

Abramitzky, R., Boustan, L. P., and Eriksson, K. A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration. *Journal of Political Economy*, 122(3):467–506, jun 2014. ISSN 0022-3808. doi: 10.1086/675805. URL https://doi.org/10.1086/675805.

Abramitzky, R., Mill, R., and Pérez, S. Linking individuals across historical sources: A fully automated approach*. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 53(2):94–111, apr 2020. ISSN 0161-5440. doi: 10.1080/01615440.2018.1543034. URL https://doi.org/10.1080/01615440.2018.1543034.

Adão, R., Arkolakis, C., and Esposito, F. General Equilibrium Effects in Space: Theory and Measurement. *National Bureau of Economic Research Working Paper Series*, No. 25544, 2019. doi: 10.3386/w25544. URL http://www.nber.org/papers/w25544http://www.nber.org/papers/w25544.pdf.

Adukia, A., Asher, S., and Novosad, P. Educational Investment Responses to Economic Opportunity: Evidence from Indian Road Construction. *American Economic Journal: Applied Economics*, 12(1):348–376, 2020. doi: 10.1257/app.20180036. URL https://www.aeaweb.org/articles?id=10.1257/app.20180036.

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. The Economics of Density: Evidence From the Berlin Wall. *Econometrica*, 83(6):2127–2189, nov 2015. ISSN 0012-9682. doi: 10.3982/ECTA10876. URL https://doi.org/10.3982/ECTA10876.

Alesina, A., Hohmann, S., Michalopoulos, S., and Papaioannou, E. Intergenerational Mobility in Africa. *Econometrica*, 89(1):1–35, jan 2021. ISSN 0012-9682. doi: https://doi.org/10.3982/ECTA17018. URL https://doi.org/10.3982/ECTA17018.

Aliprantis, D., Carroll, D. R., and Young, E. R. What explains neighborhood sorting by income and race? *Journal of Urban Economics*, page 103508, 2022. ISSN 0094-1190. doi: https://doi.org/10.1016/j.jue.2022.103508. URL https://www.sciencedirect.com/science/article/pii/S0094119022000845.

Allen, T. and Arkolakis, C. Trade and the Topography of the Spatial Economy *. *The Quarterly Journal of Economics*, 129(3):1085–1140, may 2014. ISSN 0033-5533. doi: 10.1093/qje/qju016. URL https://doi.org/10.1093/qje/qju016.

Allen, T., Arkolakis, C., and Li, X. On the Equilibrium Properties of Network Models with Heterogeneous Agents. *Working Paper*, 2022.

Almagro, M. and Dominguez-Iino, T. Location Sorting and Endogenous Amenities: Evidence from Amsterdam. *Working Paper*, 2020.

Ananat, E. O. The wrong side(s) of the tracks: The causal effects of racial segregation on urban poverty and inequality. *American Economic Journal: Applied Economics*, 3(2):34–66, 2011. ISSN 19457782. doi: 10.1257/app.3.2.34.

Andrews, R., Casey, M., Hardy, B. L., and Logan, T. D. Location matters: Historical racial segregation and intergenerational mobility. *Economics Letters*, 158:67–72, 2017. ISSN 0165-1765. doi: https://doi.org/10.1016/j.econlet.2017.06.018. URL https://www.sciencedirect.com/science/article/pii/S0165176517302458.

Atack, J. Historical Geographic Information Systems (GIS) database of Steamboat-Navigated Rivers During the Nineteenth Century in the United States, 2015.

Atack, J. Historical Geographic Information Systems (GIS) database of U.S. Railroads for 1826-1911, 2016.

Atack, J. Historical Geographic Information Systems (GIS) database of Nineteenth Century U.S. Canals, 2017.

Avila, E. The Folklore of the Freeway: Race and Revolt in the Modernist City, 2014. ISSN 0021-8723. URL http://site.ebrary.com/id/10875082.

Bagagli, S. The (Express)Way to Segregation: Evidence from Chicago. 2022.

Bailey, M. J., Cole, C., Henderson, M., and Massey, C. How Well Do Automated Linking Methods Perform? Lessons from US Historical Data. *Journal of Economic Literature*, 58(4):997–1044, 2020. doi: 10.1257/jel.20191526. URL https://www.aeaweb.org/articles?id=10.1257/jel.20191526.

Balboni, C., Bryan, G., Morten, M., and Siddiqi, B. Transportation, Gentrification, and Urban Mobility: The Inequality Effects of Place-Based Policies. 2020.

Banerjee, A., Duflo, E., and Qian, N. On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics*, 145:102442, 2020. ISSN 0304-3878. doi: https://doi.org/10.1016/j.jdeveco.2020.102442. URL https://www.sciencedirect.com/science/article/pii/S0304387820300171.

Baum-Snow, N. Did Highways Cause Suburbanization? *The Quarterly Journal of Economics*, 2007. ISSN 0033-5533, 1531-4650. doi: 10.1162/qjec.122.2.775.

Baum-Snow, N. Urban Transport Expansions and Changes in the Spatial Structure of US Cities: Implications for Productivity and Welfare. *Review of Economics and Statistics*, 2020.

Baum-Snow, N. and Han, L. The Microgeography of Housing Supply. *Working Paper*, 2021.

Baum-Snow, N. and Lutz, B. F. School desegregation, school choice, and changes in residential location patterns by race. *American Economic Review*, 2011. ISSN 00028282. doi: 10.1257/aer.101.7.3019.

Bayer, P., Ferreira, F., McMillan, R., Bajari, P., Berry, S., Black, S., Card, D., Chay, K., Epple, D., Jacob, B., Lee, D., Nechyba, T., Rothstein, J., Rueben, K., Sieg, H., and Staiger, D. Tiebout Sorting, Social Multipliers and the Demand for School Quality. *National Bureau of Economic Research Working Paper Series*, 10871(203), 2004. doi: 10.3386/w10871. URL http://www.nber.org/papers/w10871.

Bayer, P., Ferreira, F., and McMillan, R. A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy*, 115(4):588–638, aug 2007. ISSN 0022-3808. doi: 10.1086/522381. URL https://doi.org/10.1086/522381.

Bayer, P., Charles, K. K., and Park, J. Separate and Unequal: Race and the Geography of the American Housing Market. *Working Paper*, 2021.

Becker, G. S. *The economics of discrimination*. University of Chicago press, 1971. ISBN 0226041042.

Becker, G. S. and Tomes, N. An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Political Economy*, 87(6):1153–1189, mar 1979. ISSN 00223808, 1537534X. URL http://www.jstor.org/stable/1833328.

Berry, S., Levinsohn, J., and Pakes, A. Automobile Prices in Market Equilibrium. *Econometrica*, 63(4): 841–890, aug 1995. ISSN 00129682, 14680262. doi: 10.2307/2171802. URL http://www.jstor.org/stable/2171802.

Black, D. A., Sanders, S. G., Taylor, E. J., and Taylor, L. J. The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South. *American Economic Review*, 105(2): 477–503, 2015. doi: 10.1257/aer.20120642. URL https://www.aeaweb.org/articles?id=10.1257/aer.20120642.

Bobo, L. D., Charles, C. Z., Krysan, M., and Simmons, A. D. The Real Record on Racial Attitudes. In Marsden, P. V., editor, *Social Trends in American Life*, pages 38–83. Princeton University Press, 2012. doi: doi:10.1515/9781400845569-005. URL https://doi.org/10.1515/9781400845569-005.

Borjas, G. J. The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market. *The Quarterly Journal of Economics*, 118(4):1335–1374, jan 2003. ISSN 00335533, 15314650. URL http://www.jstor.org/stable/25053941.

Borusyak, K. and Hull, P. Non-Random Exposure to Exogenous Shocks. *Working Paper*, 2023.

Bound, J. and Holzer, H. J. Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s. *Journal of Labor Economics*, 18(1):20–54, aug 2000. ISSN 0734306X, 15375307. doi: 10.1086/209949. URL http://www.jstor.org/stable/10.1086/209949.

Boustan, L. P. Was postwar suburbanization "White flight"? Evidence from the black migration. *The Quarterly Journal of Economics*, 2010. ISSN 0033-5533. doi: 10.1162/qjec.2010.125.1.417.

Boustan, L. P. Competition in the Promised Land: Black Migration and Racial Wage Convergence in the North, 1940-1970. *The Journal of Economic History*, 69(3):755–782, aug 2009. ISSN 00220507, 14716372. URL http://www.jstor.org/stable/40263942.

Boustan, L. P. School Desegregation and Urban Change: Evidence from City Boundaries. *American Economic Journal: Applied Economics*, 4(1):85–108, jan 2012. ISSN 19457782, 19457790. URL http://www.jstor.org/stable/41419425.

Brinkman, J. and Lin, J. Freeway Revolts! The Quality of Life Effects of Highways. *The Review of Economics and Statistics*, pages 1–45, sep 2022. ISSN 0034-6535. doi: 10.1162/rest_a_01244. URL https://doi.org/10.1162/rest_a_01244.

Brooks, L. and Liscow, Z. D. Infrastructure Costs. 2020. doi: https://dx.doi.org/10.2139/ssrn.3428675.

Bullard, R. Race and Environmental Justice in the United States. *Yale Journal of International Law*, 18(1), jan 1993. ISSN 0889-7743. URL https://digitalcommons.law.yale.edu/yjil/vol18/iss1/12.

Bunten, D. M., Fu, E., Rolheiser, L., and Severen, C. The Problem Has Existed over Endless Years: Racialized Difference in Commuting, 1980–2019. *Working Paper*, 2022.

Busso, M., Gregory, J., and Kline, P. Assessing the Incidence and Efficiency of a Prominent Place Based Policy. *American Economic Review*, 103(2):897–947, 2013. doi: 10.1257/aer.103.2.897. URL https://www.aeaweb.org/articles?id=10.1257/aer.103.2.897.

Calonico, S., Cattaneo, M. D., and Titiunik, R. Robust Data-Driven Inference in the Regression-Discontinuity Design. *The Stata Journal*, 14(4):909–946, dec 2014. ISSN 1536-867X. doi: 10.1177/1536867X1401400413. URL https://doi.org/10.1177/1536867X1401400413.

Card, D. Immigration and Inequality. *American Economic Review*, 99(2):1–21, 2009. doi: 10.1257/aer.99.2.1. URL https://www.aeaweb.org/articles?id=10.1257/aer.99.2.1.

Card, D., Domnisoru, C., and Taylor, L. The Intergenerational Transmission of Human Capital: Evidence from the Golden Age of Upward Mobility. *National Bureau of Economic Research Working Paper Series*, No. 25000, 2018. doi: 10.3386/w25000. URL http://www.nber.org/papers/w25000http://www.nber.org/papers/w25000.pdf.

Caro, R. A. *The Power Broker: Robert Moses and the Fall of New York*. [First edition]. New York : Knopf, 1974., 1974. URL https://search.library.wisc.edu/catalog/999476860402121.

Census Bureau. 1960 Census of the Population: Supplementary Reports: Race of the Population of the United States, by States: 1960. Technical report, 1961.

Chandra, A. and Thompson, E. Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Regional Science and Urban Economics*, 30(4):457–490, 2000. ISSN 0166-0462. doi: https://doi.org/10.1016/S0166-0462(00)00040-5. URL https://www.sciencedirect.com/science/article/pii/S0166046200000405.

Chetty, R. and Hendren, N. The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates*. *The Quarterly Journal of Economics*, 133(3):1163–1228, aug 2018a. ISSN 0033-5533. doi: 10.1093/qje/qjy006. URL https://doi.org/10.1093/qje/qjy006.

Chetty, R. and Hendren, N. The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects*. *The Quarterly Journal of Economics*, 133(3):1107–1162, aug 2018b. ISSN 0033-5533. doi: 10.1093/qje/qjy007. URL https://doi.org/10.1093/qje/qjy007.

Chetty, R., Hendren, N., Kline, P., and Saez, E. Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *. *The Quarterly Journal of Economics*, 129(4):1553–1623, nov 2014. ISSN 0033-5533. doi: 10.1093/qje/qju022. URL https://doi.org/10.1093/qje/qju022.

Chetty, R., Hendren, N., Jones, M. R., and Porter, S. R. Race and Economic Opportunity in the United States: an Intergenerational Perspective*. *The Quarterly Journal of Economics*, 135(2):711–783, may 2020. ISSN 0033-5533. doi: 10.1093/qje/qjz042. URL https://doi.org/10.1093/qje/qjz042.

Christensen, P. and Timmins, C. The Damages and Distortions from Discrimination in the Rental Housing Market. *National Bureau of Economic Research Working Paper Series*, No. 29049, 2021. doi: 10.3386/w29049. URL http://www.nber.org/papers/w29049http://www.nber.org/papers/w29049.pdf.

Christensen, P. and Timmins, C. Sorting or Steering: The Effects of Housing Discrimination on Neighborhood Choice. *Journal of Political Economy*, 130(8):2110–2163, mar 2022. ISSN 0022-3808. doi: 10.1086/720140. URL https://doi.org/10.1086/720140.

Chyn, E. Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children. *American Economic Review*, 108(10):3028–3056, 2018. doi: 10.1257/aer.20161352. URL https://www.aeaweb.org/articles?id=10.1257/aer.20161352.

Chyn, E. and Daruich, D. An Equilibrium Analysis of the Effects of Neighborhood-based Interventions on Children. *National Bureau of Economic Research Working Paper Series*, No. 29927, 2022. doi: 10.3386/w29927. URL http://www.nber.org/papers/w29927http://www.nber.org/papers/w29927.pdf.

Collins, W. J. and Wanamaker, M. H. Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data. *American Economic Journal: Applied Economics*, 6(1):220–252, 2014. doi: 10.1257/app.6.1.220. URL https://www.aeaweb.org/articles?id=10.1257/app.6.1.220.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection. *Econometrica*, 80(6):2543–2594, nov 2012. ISSN 0012-9682. doi: https://doi.org/10.3982/ECTA8442. URL https://doi.org/10.3982/ECTA8442.

Combes, P.-P., Duranton, G., and Gobillon, L. The Production Function for Housing: Evidence from France. *Journal of Political Economy*, 129(10):2766–2816, apr 2021. ISSN 0022-3808. doi: 10.1086/715076. URL https://doi.org/10.1086/715076.

Conley, T. G. GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92(1):1–45, 1999. ISSN 0304-4076. doi: https://doi.org/10.1016/S0304-4076(98)00084-0. URL https://www.sciencedirect.com/science/article/pii/S0304407698000840.

Connolly, N. D. B. A World More Concrete: Real Estate and the Remaking of Jim Crow South Florida. page 376, 2014.

Costas-Fernandez, J., Guerra, J.-A., and Mohnen, M. Train to Opportunity: the Effect of Infrastructure on Intergenerational Mobility. *Working Paper*, 2023.

Cuffe, J. and Goldschlag, N. Squeezing More Out of Your Data: Business Record Linkage with Python. 2018.

Cullen, J. B. and Levitt, S. D. Crime, Urban Flight, and the Consequences for Cities. *The Review of Economics and Statistics*, 81(2):159–169, jan 1999. ISSN 00346535, 15309142. URL http://www.jstor.org/stable/2646853.

Currarini, S., Jackson, M. O., and Pin, P. An Economic Model of Friendship: Homophily, Minorities, and Segregation. *Econometrica*, 77(4):1003–1045, aug 2009. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/40263853.

Currie, J. and Walker, R. Traffic Congestion and Infant Health: Evidence from E-ZPass. *American Economic Journal: Applied Economics*, 3(1):65–90, 2011. doi: 10.1257/app.3.1.65. URL https://www.aeaweb.org/articles?id=10.1257/app.3.1.65.

Currie, J., Voorheis, J., and Walker, R. What Caused Racial Disparities in Particulate Exposure to Fall? New Evidence from the Clean Air Act and Satellite-Based Measures of Air Quality. *American Economic Review*, 2022.

Cutler, D. M., Glaeser, E. L., and Vigdor, J. L. The Rise and Decline of the American Ghetto. *Journal of Political Economy*, 107(3):455–506, 1999. doi: 10.1086/250069/0.

Dahl, M. W. and DeLeire, T. *The association between children's earnings and fathers' lifetime earnings: estimates using administrative data*. University of Wisconsin-Madison, Institute for Research on Poverty, 2008.

Davis, J. and Mazumder, B. The Decline in Intergenerational Mobility After 1980. 2022.

Davis, M. A. and Ortalo-Magné, F. Household expenditures, wages, rents. *Review of Economic Dynamics*, 14(2):248–261, 2011. ISSN 1094-2025. doi: https://doi.org/10.1016/j.red.2009.12.003. URL https://www.sciencedirect.com/science/article/pii/S1094202509000830.

Davis, M. A., Gregory, J., and Hartley, D. The Long-Run Effects of Low-Income Housing on Neighborhood Composition. 2019.

Dekle, R., Eaton, J., and Kortum, S. Global Rebalancing with Gravity: Measuring the Burden of Adjustment. *IMF Staff Papers*, 55(3):511–540, jan 2008. ISSN 10207635, 15645150. URL http://www.jstor.org/stable/40377786.

Department of Transportation, U. U.S. Secretary of Transportation Buttigieg Announces Nearly $1 Billion in Grant Awards for America's Infrastructure, 2021.

Derenoncourt, E. Can You Move to Opportunity? Evidence from the Great Migration. *American Economic Review*, 112(2):369–408, 2022. doi: 10.1257/aer.20200002. URL https://www.aeaweb.org/articles?id=10.1257/aer.20200002.

Diamond, R. The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000. *American Economic Review*, 106(3):479–524, 2016. doi: 10.1257/aer.20131706. URL https://www.aeaweb.org/articles?id=10.1257/aer.20131706.

Diamond, R. and Gaubert, C. Spatial Sorting and Inequality. *Working Paper*, 2021.

Diamond, R. and McQuade, T. Who Wants Affordable Housing in Their Backyard? An Equilibrium Analysis of Low-Income Property Development. *Journal of Political Economy*, 127(3):1063–1117, oct 2018. ISSN 0022-3808. doi: 10.1086/701354. URL https://doi.org/10.1086/701354.

Dingel, J. and Tintelnot, F. Spatial Economics for Granular Settings. *Working Paper*, 2023.

Donaldson, D. Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*, 108(4-5):899–934, 2018. doi: 10.1257/aer.20101199. URL https://www.aeaweb.org/articles?id=10.1257/aer.20101199.

Donaldson, D. and Hornbeck, R. Railroads and American Economic Growth: A "Market Access" Approach. *The Quarterly Journal of Economics*, 131(2):799–858, may 2016. ISSN 0033-5533. doi: 10.1093/qje/qjw002. URL https://doi.org/10.1093/qje/qjw002.

Duranton, G. and Puga, D. Chapter 48 - Micro-Foundations of Urban Agglomeration Economies. In Henderson, J. V., Thisse, J.-F. B. T. H. o. R., and Economics, U., editors, *Cities and Geography*, volume 4, pages 2063–2117. Elsevier, 2004. ISBN 1574-0080. doi: https://doi.org/

10.1016/S1574-0080(04)80005-1. URL https://www.sciencedirect.com/science/article/pii/S1574008004800051.

Duranton, G. and Turner, M. A. Urban Growth and Transportation. *Review of Economic Studies*, 79(4): 1407–1440, 2012. ISSN 00346527. doi: 10.1093/restud/rds010.

Duranton, G., Morrow, P. M., and Turner, M. A. Roads and Trade: Evidence from the US. *The Review of Economic Studies*, 81(2):681–724, apr 2014. ISSN 0034-6527. doi: 10.1093/restud/rdt039. URL https://doi.org/10.1093/restud/rdt039.

Ellison, G., Glaeser, E. L., and Kerr, W. R. What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review*, 100(3):1195–1213, 2010. doi: 10.1257/aer.100.3.1195. URL https://www.aeaweb.org/articles?id=10.1257/aer.100.3.1195.

Epple, D., Gordon, B., and Sieg, H. A New Approach to Estimating the Production Function for Housing. *American Economic Review*, 100(3):905–924, 2010. doi: 10.1257/aer.100.3.905. URL https://www.aeaweb.org/articles?id=10.1257/aer.100.3.905.

ESRI. USA (Interstate Highways, 1996), 1996.

ESRI. US Major Roads (2010), 2019. URL https://www.arcgis.com/home/item.html?id=871852b13b53426dabdf875f80c04261.

Faber, B. Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *The Review of Economic Studies*, 81(3):1046–1070, mar 2014. ISSN 0034-6527. doi: 10.1093/restud/rdu010. URL https://doi.org/10.1093/restud/rdu010.

Faber, J. W. We Built This: Consequences of New Deal Era Intervention in America's Racial Geography. *American Sociological Review*, 85(5):739–775, aug 2020. ISSN 0003-1224. doi: 10.1177/0003122420948464. URL https://doi.org/10.1177/0003122420948464.

Fajgelbaum, P. D. and Gaubert, C. Optimal Spatial Policies, Geography, and Sorting*. *The Quarterly Journal of Economics*, 135(2):959–1036, jan 2020. ISSN 0033-5533. doi: 10.1093/qje/qjaa001. URL https://doi.org/10.1093/qje/qjaa001.

Feigenbaum, J. Intergenerational Mobility during the Great Depression. 2015.

Feigenbaum, J. A Machine Learning Approach to Census Record Linking. 2016.

Ferreira, F. V. and Wong, M. Estimating Preferences for Neighborhood Amenities Under Imperfect Information. *National Bureau of Economic Research Working Paper Series*, No. 28165, 2020. doi: 10.3386/w28165. URL http://www.nber.org/papers/w28165http://www.nber.org/papers/w28165.pdf.

Ferrie, J. P. A New Sample of Males Linked from the Public Use Microdata Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 29(4):141–156, oct 1996. ISSN 0161-5440. doi: 10.1080/01615440.1996.10112735. URL https://doi.org/10.1080/01615440.1996.10112735.

Fishback, P. V., LaVoice, J., Shertzer, A., and Walsh, R. The HOLC Maps: How Race and Poverty Influenced Real Estate Professionals' Evaluation of Lending Risk in the 1930s. *National Bureau of Economic Research Working Paper Series*, No. 28146, 2020. doi: 10.3386/w28146. URL http://www.nber.org/papers/w28146http://www.nber.org/papers/w28146.pdf.

Fishback, P. V., Rose, J., Snowden, K., and Storrs, T. New Evidence on Redlining by Federal Housing Programs in the 1930s. *Working Paper*, 2022.

Gaubert, C., Kline, P. M., and Yagan, D. Place-Based Redistribution. *National Bureau of Economic Research Working Paper Series*, No. 28337, 2021. doi: 10.3386/w28337. URL http://www.nber.org/papers/w28337http://www.nber.org/papers/w28337.pdf.

Gibbons, J. W. and Proctor, A. Economic Costs of Traffic Congestion. *Highway Research Board Bulletin*, (86), 1954.

Glaeser, E. L. and Gottlieb, J. D. The Economics of Place-Making Policies. *Brookings Papers on Economic Activity*, 2008:155–239, jul 2008. ISSN 00072303, 15334465. URL http://www.jstor.org/stable/27561617.

Glaeser, E. L. and Ponzetto, G. A. M. The political economy of transportation investment. *Economics of Transportation*, 13:4–26, 2018. ISSN 2212-0122. doi: https://doi.org/10.1016/j.ecotra.2017.08.001. URL https://www.sciencedirect.com/science/article/pii/S2212012217300771.

Greenstone, M., Hornbeck, R., and Moretti, E. Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. *Journal of Political Economy*, 118(3):536–598, jun 2010. ISSN 0022-3808. doi: 10.1086/653714. URL https://doi.org/10.1086/653714.

Greenwood, J., Hercowitz, Z., and Krusell, P. Long-Run Implications of Investment-Specific Technological Change. *The American Economic Review*, 87(3):342–362, 1997. ISSN 00028282. URL http://www.jstor.org/stable/2951349.

Haider, S. and Solon, G. Life-Cycle Variation in the Association between Current and Lifetime Earnings. *American Economic Review*, 96(4):1308–1320, 2006. doi: 10.1257/aer.96.4.1308. URL https://www.aeaweb.org/articles?id=10.1257/aer.96.4.1308.

Haltiwanger, J. C., Kutzbach, M. J., Palloni, G. E., Pollakowski, H., Staiger, M., and Weinberg, D. The Children of HOPE VI Demolitions: National Evidence on Labor Market Outcomes. *National Bureau of Economic Research Working Paper Series*, No. 28157, 2020. doi: 10.3386/w28157. URL http://www.nber.org/papers/w28157http://www.nber.org/papers/w28157.pdf.

Harriss, C. L. *History and Policies of the Home Owners' Loan Corporation*. National Bureau of Economic Research, 1951.

Hausman, J. A. Valuation of new goods under perfect and imperfect competition. *NBER Chapters*, pages 207–248, 1996. URL https://ideas.repec.org/h/nbr/nberch/6068.htmlhttps://ideas.repec.org//h/nbr/nberch/6068.html.

Head, K. and Mayer, T. Chapter 3 - Gravity Equations: Workhorse,Toolkit, and Cookbook. In Gopinath, G., Helpman, E., and Rogoff, K. B. T. H. o. I. E., editors, *Handbook of International Economics*, volume 4, pages 131–195. Elsevier, 2014. ISBN 1573-4404. doi: https://doi.org/10.1016/B978-0-444-54314-1.00003-3. URL https://www.sciencedirect.com/science/article/pii/B9780444543141000033.

Heath Milsom, L. Moving Opportunity, Local Connectivity and Spatial Inequality. *Working Paper*, 2023.

Heblich, S., Redding, S. J., and Sturm, D. M. The Making of the Modern Metropolis: Evidence from London*. *The Quarterly Journal of Economics*, 135(4):2059–2133, nov 2020. ISSN 0033-5533. doi: 10.1093/qje/qjaa014. URL https://doi.org/10.1093/qje/qjaa014.

Higgs, R. Firm-Specific Evidence on Racial Wage Differentials and Workforce Segregation. *The American Economic Review*, 67(2):236–245, jan 1977. ISSN 00028282. URL http://www.jstor.org/stable/1807245.

Hillier, A. E. Redlining and the Home Owners' Loan Corporation. *Journal of Urban History*, 29(4):394–420, may 2003. ISSN 0096-1442. doi: 10.1177/0096144203029004002. URL https://doi.org/10.1177/0096144203029004002.

Hirsch, A. R. *Making the Second Ghetto: Race and Housing in Chicago, 1940-1960*. The University of Chicago Press, 1983.

Holian, M. J. and Kahn, M. E. Household carbon emissions from driving and center city quality of life. *Ecological Economics*, 116:362–368, 2015. ISSN 09218009. doi: 10.1016/j.ecolecon.2015.05.012. URL http://linkinghub.elsevier.com/retrieve/pii/S092180091500227X.

Holzer, H. J. and Reaser, J. Black Applicants, Black Employees, and Urban Labor Market Policy. *Journal of Urban Economics*, 48(3):365–387, 2000. ISSN 0094-1190. doi: https://doi.org/10.1006/juec.2000.2171. URL https://www.sciencedirect.com/science/article/pii/S0094119000921715.

Jackson, K. T. Race, Ethnicity, and Real Estate Appraisal: The Home Owners Loan Corporation and the Federal Housing Administration. *Journal of Urban History*, 6(4):419–452, aug 1980. ISSN 0096-1442. doi: 10.1177/009614428000600404. URL https://doi.org/10.1177/009614428000600404.

Jackson, K. T. *Crabgrass Frontier: The Suburbanization of the United States*. Oxford University Press, New York, 1985. ISBN 0195036107 9780195036107 9780195049831 0195049837.

Jacome, E., Kuziemko, I., and Naidu, S. Mobility for All: Representative Intergenerational Mobility Estimates over the 20th Century. 2023.

Jones-Correa, M. The Origins and Diffusion of Racial Restrictive Covenants. *Political Science Quarterly*, 115(4):541–568, aug 2000. ISSN 00323195. doi: 10.2307/2657609. URL http://www.jstor.org/stable/2657609.

Kain, J. F. Housing Segregation, Negro Employment, and Metropolitan Decentralization*. *The Quarterly Journal of Economics*, 82(2):175–197, may 1968. ISSN 0033-5533. doi: 10.2307/1885893. URL https://doi.org/10.2307/1885893.

Katz, L. F. and Murphy, K. M. Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics*, 107(1):35–78, jan 1992. ISSN 00335533, 15314650. doi: 10.2307/2118323. URL http://www.jstor.org/stable/2118323.

Kline, P. and Moretti, E. Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority *. *The Quarterly Journal of Economics*, 129(1):275–331, feb 2014. ISSN 0033-5533. doi: 10.1093/qje/qjt034. URL https://doi.org/10.1093/qje/qjt034.

Kreindler, G. and Miyauchi, Y. Measuring Commuting and Economic Activity Inside Cities with Cell Phone Records. *Review of Economics and Statistics*, 2022.

Krimmel, J. Reclaiming Local Control: School Finance Reforms and Housing Supply Restrictions. *Working Paper*, 2022.

Kuminoff, N. V., Smith, V. K., and Timmins, C. The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets. *Journal of Economic Literature*, 51(4):1007–1062, 2013. doi: 10.1257/jel.51.4.1007. URL https://www.aeaweb.org/articles?id=10.1257/jel.51.4.1007.

Lee, S. and Lin, J. Natural Amenities, Neighbourhood Dynamics, and Persistence in the Spatial Distribution of Income. *The Review of Economic Studies*, pages rdx018–rdx018, mar 2017. ISSN 0034-6527. URL http://dx.doi.org/10.1093/restud/rdx018.

Lee, S. K. When Cities Grow: Urban Planning and Segregation in the Prewar US. *Working Paper*, 2022.

Logan, J. R. Separate and Unequal: The Neighborhood Gap for Blacks and Hispanics in Metropolitan America. Technical report, 2011.

Logan, J. R., Xu, Z., and Stults, B. J. Interpolating U.S. Decennial Census Tract Data from as Early as 1970 to 2010: A Longitudinal Tract Database. *The Professional Geographer*, 66(3):412–420, jul 2014. ISSN 0033-0124. doi: 10.1080/00330124.2014.905156. URL https://doi.org/10.1080/00330124.2014.905156.

Long, J. and Ferrie, J. Intergenerational Occupational Mobility in Great Britain and the United States Since 1850. *The American Economic Review*, 103(4):1109–1137, apr 2013. ISSN 00028282. URL http://www.jstor.org/stable/23469614.

Loury, G. A Dynamic Theory of Racial Income Differences. 1976.

Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R., and Sanbonmatsu, L. Long-term neighborhood effects on low-income families: Evidence from moving to opportunity. In *American Economic Review*, volume 103, pages 226–231, 2013. ISBN 0002-8282, 0002-8282. doi: 10.1257/aer.103.3.226.

Mahajan, A. Highways and segregation. *Journal of Urban Economics*, page 103574, 2023. ISSN 0094-1190. doi: https://doi.org/10.1016/j.jue.2023.103574. URL https://www.sciencedirect.com/science/article/pii/S0094119023000438.

Manson, S. U. o. M., Schroeder, J., Van Riper, D., and Ruggles, S. IPUMS National Historical Geographic Information System: Version 12.0 [Database]. 2017. URL http://doi.org/10.18128/D050.V12.0.

Margo, R. A. *Race and Schooling in the South, 1880-1950: An Economic History*. University of Chicago Press, 2007. ISBN 9780226505015. doi: doi:10.7208/9780226505015. URL https://doi.org/10.7208/9780226505015.

Massey, D. S. and Denton, N. A. *American Apartheid: Segregation and the Making of the Underclass*. Harvard University Press, Cambridge, Mass, 1993. ISBN 0674018206 (acid-free paper).

Mayo, S. K. Theory and estimation in the economics of housing demand. *Journal of Urban Economics*, 10(1):95–116, 1981. ISSN 0094-1190. doi: https://doi.org/10.1016/0094-1190(81)90025-5. URL https://www.sciencedirect.com/science/article/pii/0094119081900255.

Mazumder, B. Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data. *The Review of Economics and Statistics*, 87(2):235–255, mar 2005. ISSN 00346535, 15309142. URL http://www.jstor.org/stable/40042900.

Mazumder, B. Estimating the Intergenerational Elasticity and Rank Association in the US: Overcoming the Current Limitations of Tax Data. 2015.

Michaels, G. The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System. *The Review of Economics and Statistics*, 90(4):683–701, oct 2008. ISSN 0034-6535. doi: 10.1162/rest.90.4.683. URL https://doi.org/10.1162/rest.90.4.683.

Miller, C. When Work Moves: Job Suburbanization and Black Employment. *NBER Working Paper No. 24728*, 2018. ISSN 1098-6596. doi: 10.1017/CBO9781107415324.004.

Mohl, R. A. Stop the Road: Freeway Revolts in American Cities. *Journal of Urban History*, 30(5): 674–706, 2004. ISSN 00961442. doi: 10.1177/0096144204265180.

Monte, F., Redding, S. J., and Rossi-Hansberg, E. Commuting, Migration, and Local Employment Elasticities. *American Economic Review*, 108(12):3855–3890, 2018. doi: 10.1257/aer.20151507. URL https://www.aeaweb.org/articles?id=10.1257/aer.20151507.

Morten, M. and Oliveira, J. The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City. *Working Paper*, 2018.

Nall, C. The Road to Inequality: How the Federal Highway Program Polarized America and Undermined Cities. 2018.

Nall, C. and O'Keeffe, Z. P. "What Did the Interstate Highway Program Do to Urban Neighborhoods?". 2019.

Nelson, R. K., Winling, L., Marciano, R., Connolly, N., and Al., E. Mapping Inequality. In Nelson, R. K. and Ayers, E. L., editors, *American Panorama*. 2020.

Nybom, M. and Stuhler, J. Heterogeneous Income Profiles and Lifecycle Bias in Intergenerational Mobility Estimation. *The Journal of Human Resources*, 51(1):239–268, jan 2016. ISSN 0022166X. URL http://www.jstor.org/stable/24736005.

Olivetti, C. and Paserman, M. D. In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940. *American Economic Review*, 105(8):2695–2724, 2015. doi: 10.1257/aer.20130821. URL https://www.aeaweb.org/articles?id=10.1257/aer.20130821.

Ottaviano, G. I. P. and Peri, G. Rethinking the Effect of Immigration on Wages. *Journal of the European Economic Association*, 10(1):152–197, feb 2012. ISSN 1542-4766. doi: 10.1111/j.1542-4774.2011.01052.x. URL https://doi.org/10.1111/j.1542-4774.2011.01052.x.

Reardon, S. F. and Bischoff, K. Income Inequality and Income Segregation. *American Journal of Sociology*, 116(4):1092–1153, 2011. ISSN 0002-9602. doi: 10.1086/657114. URL http://www.journals.uchicago.edu/doi/10.1086/657114.

Redding, S. J. and Rossi-Hansberg, E. Quantitative Spatial Economics. *Annual Review of Economics*, 9 (1):21–58, aug 2017. ISSN 1941-1383. doi: 10.1146/annurev-economics-063016-103713. URL https://doi.org/10.1146/annurev-economics-063016-103713.

Roback, J. Wages, Rents, and the Quality of Life. *Journal of Political Economy*, 90(6):1257–1278, 1982. ISSN 00223808, 1537534X. URL http://www.jstor.org/stable/1830947.

Rose, M. and Mohl, R. A. *Interstate: Highway Politics and Policy since 1939*. The University of Tennessee Press, 2012.

Rose, M. H. *Interstate: Express Highway Politics, 1939-1989*. University of Tennessee Press, 1990.

Rosen, S. Wage Based Indices of Urban Quality of Life. In Mieszkowski, P. and Straszheim, M., editors, *Current Issues in Urban Economics*. Johns Hopkins University Press, 1979.

Rosenthal, S. S. and Strange, W. C. Chapter 49 - Evidence on the Nature and Sources of Agglomeration Economies. In Henderson, J. V., Thisse, J.-F. B. T. H. o. R., and Economics, U., editors, *Cities and Geography*, volume 4, pages 2119–2171. Elsevier, 2004. ISBN 1574-0080. doi: https://doi.org/10.1016/S1574-0080(04)80006-3. URL https://www.sciencedirect.com/science/article/pii/S1574008004800063.

Rothstein, R. *The Color of Law*. 2017.

Rumsey, D. *David Rumsey Map Collection*. Cartography Associates, San Francisco, California, 2020.

Saiz, A. The Geographic Determinants of Housing Supply*. *The Quarterly Journal of Economics*, 125 (3):1253–1296, aug 2010. ISSN 0033-5533. doi: 10.1162/qjec.2010.125.3.1253. URL https://doi.org/10.1162/qjec.2010.125.3.1253.

Sampson, R. J., Morenoff, J. D., and Gannon-Rowley, T. Assessing "Neighborhood Effects": Social Processes and New Directions in Research. *Annual Review of Sociology*, 28(1):443–478, aug 2002. ISSN 0360-0572. doi: 10.1146/annurev.soc.28.110601.141114. URL https://doi.org/10.1146/annurev.soc.28.110601.141114.

Severen, C. Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification. *The Review of Economics and Statistics*, pages 1–99, sep 2021. ISSN 0034-6535. doi: 10.1162/rest_a_01100. URL https://doi.org/10.1162/rest_a_01100.

Sharkey, P. The Intergenerational Transmission of Context. *American Journal of Sociology*, 113(4): 931–969, feb 2008. ISSN 00029602, 15375390. doi: 10.1086/522804. URL http://www.jstor.org/stable/10.1086/522804.

Shroder, M. Locational Constraint, Housing Counseling, and Successful Lease-up in a Randomized Housing Voucher Experiment. *Journal of Urban Economics*, 51(2):315–338, 2002. ISSN 0094-1190. doi: https://doi.org/10.1006/juec.2001.2247. URL http://www.sciencedirect.com/science/article/pii/S0094119001922478.

Silva, J. M. C. S. and Tenreyro, S. The Log of Gravity. *The Review of Economics and Statistics*, 88(4): 641–658, nov 2006. ISSN 0034-6535. doi: 10.1162/rest.88.4.641. URL https://doi.org/10.1162/rest.88.4.641.

Solon, G. Intergenerational Income Mobility in the United States. *The American Economic Review*, 82 (3):393–408, feb 1992. ISSN 00028282. URL http://www.jstor.org/stable/2117312.

Solon, G. Chapter 29 - Intergenerational Mobility in the Labor Market. volume 3, pages 1761–1800. Elsevier, 1999. ISBN 1573-4463. doi: https://doi.org/10.1016/S1573-4463(99)03010-2. URL https://www.sciencedirect.com/science/article/pii/S1573446399030102.

Song, J. Exclusionary Zoning and Transportation. *Working Paper*, 2022.

Stinson, M. and Weiwu, L. Intergenerational Linkages between Historical IRS 1040 Data and the Numident: 1964-1979 Cohorts. 2023.

Taylor, K.-Y. *Race for Profit*. University of North Carolina Press, aug 2019. ISBN 9781469653662. URL http://www.jstor.org/stable/10.5149/9781469653686_taylor.

Trounstine, J. *Segregation by Design: Local Politics and Inequality in American Cities*. Cambridge University Press, Cambridge, 2018. ISBN 9781108429955. doi: DOI:10.1017/9781108555722. URL https://www.cambridge.org/core/books/segregation-by-design/9CEF629688C0C684EDC387407F5878F2.

Tsivanidis, N. Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio. *Working Paper*, 2022.

Walters, A. A. The Theory and Measurement of Private and Social Cost of Highway Congestion. *Econometrica*, 29(4):676–699, jan 1961. ISSN 00129682, 14680262. doi: 10.2307/1911814. URL http://www.jstor.org/stable/1911814.

Weiwu, L. Unequal Access: Racial Segregation and the Distributional Impacts of Interstate Highways in Citie. 2023.

Wilson, W. J. *The Truly Disadvantaged: the Inner City, the Underclass, and Public Policy*. Chicago : University of Chicago Press, 1987., 1987. URL https://search.library.wisc.edu/catalog/999585772202121.

Wooldridge, J. M. Control Function Methods in Applied Econometrics. *The Journal of Human Resources*, 50(2):420–445, aug 2015. ISSN 0022166X. URL http://www.jstor.org/stable/24735991.

18 Tables

Table 1: Summary Statistics by Race and Education in 1960

Variables	(1) Black <HS	(2) Black HS Grad	(3) White <HS	(4) White HS Grad
Economic Variables – Mean (SD)				
Weekly Wages (2010$)	402.3	495.2	569.4	726.7
	(115.6)	(131.6)	(97.6)	(159.1)
Rent (2010$)	382.9	464.6	444.9	607.7
	(124.6)	(136.0)	(140.8)	(194.8)
Home Value (2010$)	65900	92200	93700	130000
	(29850)	(35740)	(32050)	(40200)
Home Ownership Rate	0.334	0.379	0.599	0.626
	(0.238)	(0.273)	(0.266)	(0.280)
Neighborhood Variables – Mean (SD)				
Pct White	0.397	0.418	0.943	0.959
	(0.325)	(0.336)	(0.130)	(0.104)
Pct HS Grad	0.347	0.410	0.474	0.569
	(0.133)	(0.144)	(0.146)	(0.148)
Pct HOLC D	0.661	0.554	0.303	0.189
	(0.405)	(0.436)	(0.409)	(0.342)
Pct HOLC D (w/ missing)	0.509	0.442	0.187	0.110
	(0.451)	(0.448)	(0.353)	(0.278)
Dist Highway (mi)	1.779	1.643	2.578	2.565
	(4.072)	(3.753)	(4.891)	(4.781)
Dist CBD (mi)	6.187	6.337	9.743	9.892
	(7.703)	(6.888)	(9.485)	(9.086)
Commuting Variables – Mean (SD)				
Commute Time (min)	26.86	26.76	26.65	27.74
	(11.07)	(10.75)	(12.03)	(12.42)
Commute Dist (mi)	9.30	9.36	9.22	10.11
	(7.38)	(7.03)	(7.20)	(7.25)
Pct Auto	0.392	0.488	0.561	0.663
	(0.324)	(0.361)	(0.315)	(0.293)
Commute Time (min), Auto	28.56	28.27	28.02	28.31
	(14.89)	(13.97)	(14.66)	(13.93)
Commute Dist (mi), Auto	11.07	10.84	10.56	10.68
	(7.56)	(7.34)	(7.38)	(7.19)
Rounded Count	2,834,000	1,334,000	16,190,000	18,240,000

Notes: Data comes from the 1960 Census restricted microdata. Weekly wages are calculated for employed workers and CPI-adjusted to 2010 dollars from 1960 dollars. Rents are monthly and CPI-adjusted to 2010 dollars from 1960 dollars. Home values are CPI-adjusted to 2010 dollars from 1960 dollars and rounded to four significant digits to meet Census disclosure rules. Pct HOLC D is calculated on tracts where redlining maps exist while Pct HOLC D (w/ missing) includes tracts without redlining maps. Distance from highways is calculated using 1960 residential location and constructed highways. Percentage automobile is the percentage of employed workers whose main mode of transport is private automobile which includes truck and van drivers. Commute time and distance in the bottom rows are shown for workers whose mode of transport is private automobile. Counts of each race by education group are rounded to four significant digits to meet Census disclosure rules. Sample standard deviations are included in parentheses.

Table 2: Placement of Highway Routes for 1950 Variables

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Log Distance from Highway Routes							Plans	Rays
Pct White	0.754***				0.156	0.104	0.125	0.251	0.0645
	(0.0800)				(0.120)	(0.115)	(0.104)	(0.158)	(0.162)
Pct HS Grad		1.015***			0.538***	0.507***	0.428**	0.0343	-0.0327
		(0.181)			(0.204)	(0.193)	(0.171)	(0.284)	(0.343)
Log Median Income			0.0987***		0.0196	0.0247	0.0386*	0.0125	0.0231
			(0.0271)		(0.0221)	(0.0217)	(0.0197)	(0.0205)	(0.0216)
Pct HOLC D				-0.546***	-0.381***	-0.257***	-0.189***	-0.0646	-0.0918
				(0.0576)	(0.0730)	(0.0584)	(0.0533)	(0.0715)	(0.0694)
Pct Bottom Quintile					-0.0584	0.288	0.169	0.429	0.558
					(0.421)	(0.421)	(0.355)	(0.395)	(0.379)
Pct Top Quintile					0.0281	0.234	0.189	0.437	0.623
					(0.293)	(0.293)	(0.297)	(0.401)	(0.386)
Log Rent					-0.0341***	-0.0173*	-0.0147	0.00162	-0.0142
					(0.0113)	(0.0103)	(0.00962)	(0.0133)	(0.0136)
Log Home Value					0.0257*	0.00439	0.00104	0.0233*	0.0215
					(0.0143)	(0.0123)	(0.0111)	(0.0122)	(0.0142)
Log Dist CBD						0.225***	0.207***	0.287***	0.408***
						(0.0411)	(0.0376)	(0.0509)	(0.0660)
R-squared	0.050	0.064	0.040	0.073	0.085	0.106	0.165	0.141	0.170
CBSA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo Controls	No	No	No	No	No	No	Yes	Yes	Yes
Observations	14,235	14,235	14,235	14,235	14,235	14,235	14,235	14,235	14,235
No. Counties	165	165	165	165	165	165	165	165	165

Notes: Unit of observation is census tract. Data comes from 1950 tract-level aggregates retrieved from IPUMS NHGIS. Tracts are limited to those within 5 miles of the nearest highway route. Fixed effects are at the CBSA (Core-based statistical area) level. Standard errors are cluster-robust with clusters at the county level. The geographic controls are log distance from the central business district (included in all specifications), log distance from rivers, lakes, shores, ports, historical railroads, canals, and historical large urban roads. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Changes Over Distance from Central Business District and Distance from Highway (1950-1960, 1960-1970)

	(1)	(2)	(3)	(4)	(5)
	Δ Log Pop	Δ Log Rent	Δ Pct White	Δ Log White Pop	Δ Log Black Pop
Variables			*Panel A* – OLS		
Log Dist Highway	0.0724***	0.0360***	0.0169**	0.110***	0.00703
	(0.0155)	(0.0119)	(0.00674)	(0.0184)	(0.0357)
Log Dist CBD	0.622***	0.165***	0.0629***	0.691***	0.246***
	(0.0267)	(0.0169)	(0.00943)	(0.0303)	(0.0440)
Redlined	0.0102	-0.0408*	-0.159***	-0.154***	0.237***
	(0.0383)	(0.0245)	(0.0211)	(0.0480)	(0.0906)
R-squared	0.143	0.089	0.091	0.155	0.063
			Panel B – OLS + Geo Controls		
Log Dist Highway	0.0594***	0.0348***	0.0249***	0.108***	-0.0298
	(0.0164)	(0.0123)	(0.00742)	(0.0197)	(0.0370)
Log Dist CBD	0.596***	0.102***	0.0742***	0.692***	0.270***
	(0.0311)	(0.0159)	(0.00905)	(0.0346)	(0.0486)
Redlined	0.0626*	-0.00340	-0.165***	-0.110**	0.247***
	(0.0373)	(0.0255)	(0.0210)	(0.0460)	(0.0905)
R-squared	0.153	0.099	0.104	0.173	0.066
			Panel C – IV Log Dist Plans for Log Dist Highway [KP F-Stat = 613.2]		
Log Dist Highway	0.0504	0.0222	0.0266	0.0947**	-0.0160
	(0.0332)	(0.0273)	(0.0188)	(0.0398)	(0.0861)
Log Dist CBD	0.598***	0.105***	0.0738***	0.695***	0.267***
	(0.0318)	(0.0173)	(0.00929)	(0.0353)	(0.0516)
Redlined	0.0607	-0.00600	-0.165***	-0.113**	0.250***
	(0.0380)	(0.0263)	(0.0212)	(0.0468)	(0.0921)
R-squared	0.153	0.099	0.104	0.173	0.066
			Panel D – IV Log Dist Rays for Log Dist Highway [KP F-Stat = 466.7]		
Log Dist Highway	0.121**	0.116***	0.0630***	0.191***	-0.00400
	(0.0475)	(0.0350)	(0.0232)	(0.0539)	(0.116)
Log Dist CBD	0.581***	0.0832***	0.0654***	0.673***	0.264***
	(0.0330)	(0.0177)	(0.00981)	(0.0360)	(0.0550)
Redlined	0.0752**	0.0132	-0.158***	-0.0928**	0.252***
	(0.0383)	(0.0261)	(0.0216)	(0.0472)	(0.0932)
R-squared	0.151	0.095	0.099	0.171	0.066
Dep. Var Mean (1960)	3403	555 (2010$)	0.880	2874	488
CBSA FE	Yes	Yes	Yes	Yes	Yes
Geo Controls	Yes	Yes	Yes	Yes	Yes
Observations	11,923	11,923	11,923	11,923	11,923

Notes: Unit of observation is census tract. Data comes from 1950, 1960, and 1970 tract-level aggregates retrieved from IPUMS NHGIS. The first difference is either over 1950 to 1960 or 1960 to 1970 depending on when highway construction started in the CBSA and stacked into one panel. Tracts are limited to those within 5 miles of the nearest constructed route and 30 miles from the central business district. Fixed effects are at the CBSA (Core-based statistical area) level. Conley standard errors for spatial correlation within a 1km radius are reported. Panels B-D have as controls log distance from rivers, lakes, shores, ports, historical railroads, canals, and historical large urban roads. All specifications include the gradient (Dist CBD/Dist Highway) as a control. Redlined tracts are those where more than 80% of the area is redlined. Kleibergen-Paap rk Wald F statistics are reported for the first-stage. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Elasticity of Population to Commuter Access by Race

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$\Delta \log L_{igr}$ (Δ Log Population 1960-1970)								
	OLS			OLS + Redlining FE			IV		
Variables	Geo Cont	+ CBSA FE	+ BH (2023)	Geo Cont	+ CBSA FE	+ BH (2023)	HW	Plans	Rays
$\Delta \log CMA_{igr}$									
Black	0.416***	0.0907	0.0941	0.596***	0.273***	0.273***	-2.026***	-1.757	-3.062*
	(0.101)	(0.0968)	(0.0968)	(0.105)	(0.100)	(0.100)	(0.469)	(1.276)	(1.564)
	[0.114]	[0.110]	[0.110]	[0.119]	[0.116]	[0.116]	[0.564]	[1.490]	[1.838]
White	1.207***	1.401***	1.410***	0.958***	1.069***	1.083***	0.166	0.648**	0.719**
	(0.109)	(0.115)	(0.115)	(0.119)	(0.125)	(0.125)	(0.143)	(0.323)	(0.340)
	[0.126]	[0.142]	[0.143]	[0.137]	[0.157]	[0.157]	[0.190]	[0.456]	[0.458]
R-squared	0.088	0.113	0.113	0.094	0.117	0.118	0.0923	0.099	0.072
CBSA FE		Yes	Yes		Yes	Yes	Yes	Yes	Yes
Geo Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rounded Obs	60500	60500	60500	60500	60500	60500	60500	60500	60500
C-D F-Stat							1543	203	139
K-P F-stat							520	35	25

Notes: Unit of observation is census tract by race and education. Data comes from the first difference of 1960 to 1970 using restricted Census microdata. Fixed effects are at the CBSA (Core-based statistical area) level. Standard errors are cluster-robust with clusters at the tract-level. Conley standard errors for spatial correlation within a 1km radius are reported in brackets. The geographic controls are log distance from the central business district, rivers, lakes, shores, ports, historical railroads, canals, and historical large urban roads, all interacted with race. In Column 3, the Borusyak and Hull (2023) control for CMA in large roads is interacted with race. Redlining fixed effects are interacted with race. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. IV specifications include the Borusyak and Hull (2023) control for CMA in large roads interacted with race and CBSA fixed effects. Kleibergen-Paap rk Wald and Cragg-Donald Wald F statistics for weak instruments are reported. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5: Commuting Gravity Equation

	(1)	(2)	(3)	(4)
Race x Educ	Black <HS	Black HS Grad	White <HS	White HS Grad
$\nu_{gr} = \kappa_{gr}\phi$	*Panel A* – Log Commuting Share			
Log Commute Time	-4.201*** (0.119)	-3.609*** (0.120)	-4.664*** (0.0640)	-4.134*** (0.0496)
R-squared	0.692	0.623	0.574	0.579
	Panel B – Log Commuting Share – IV Plans			
Log Commute Time	-4.206*** (0.126)	-3.671*** (0.120)	-4.707*** (0.0673)	-4.168*** (0.0505)
R-squared	0.232	0.182	0.377	0.367
	Panel C – Log Commuting Share – IV Rays			
Log Commute Time	-4.197*** (0.127)	-3.645*** (0.122)	-4.708*** (0.0674)	-4.154*** (0.0503)
R-squared	0.232	0.182	0.377	0.367
Rounded Obs	7000	8000	21500	25000
	Panel D – Poisson Pseudo Maximum Likelihood			
Log Commute Time	-4.703*** (0.0819)	-3.929*** (0.0599)	-3.877*** (0.0471)	-3.247*** (0.0359)
Rounded Obs	20500	21000	26000	27000
POR X Year FE	Yes	Yes	Yes	Yes
POW X Year FE	Yes	Yes	Yes	Yes

Notes: Unit of observation is Place of Work Zone by Place of Work Zone pair by year where commuting flows from residential tracts are aggregated up to the Place of Work Zone geography. Data comes from the restricted Census microdata in 1960 and 1970. Fixed effects are for POR (Place of Residence) by year at the Place of Work Zone scale although it does not represent workplace but rather residential location. POW by year fixed effects are for workplace at the Place of Work Zone level. The conditional commuting share is the share from a residential location that commutes to a workplace. The observation counts are lower for the Black population as some residences and workplaces have zero Black population (while PPML addresses zeros in bilateral flows, it does not address zeros in entire rows or columns). Standard errors are cluster-robust with clusters at the Place of Work Zone by Place of Work Zone level. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Residential Elasticity and Preferences for Racial Composition

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta \log L_{igr}$ (Δ Log Population 1960-1970)							
	OLS		IV Hausman		IV Davis		IV CMA	
Variables	Base & Geo Cont	+ SES Cont	Plans	Rays	Plans	Rays	Plans	Rays
θ_r: $\Delta \log CMA_{igr}$								
Black	0.119	0.224	0.353	0.362	1.281***	1.284***	0.00593	0.412
	(0.172)	(0.171)	(0.324)	(0.291)	(0.374)	(0.324)	(0.378)	(0.290)
	[0.196]	[0.196]	[0.579]	[0.563]	[0.626]	[0.585]	[0.672]	[0.494]
White	0.802***	0.802***	0.420**	0.777***	0.576***	0.918***	0.228	0.493**
	(0.183)	(0.183)	(0.185)	(0.190)	(0.167)	(0.161)	(0.244)	(0.203)
	[0.213]	[0.214]	[0.247]	[0.270]	[0.225]	[0.237]	[0.282]	[0.253]
$\tilde{\rho}_r = \theta_r \rho_r$: $\Delta \log$ Pct White								
Black	-0.364***	-0.283***	-0.0650	-0.0973**	-0.0616	-0.0766	-0.0737	-0.0418
	(0.0546)	(0.0523)	(0.0532)	(0.0481)	(0.111)	(0.0894)	(0.151)	(0.137)
	[0.0722]	[0.0720]	[0.0968]	[0.0919]	[0.176]	[0.147]	[0.238]	[0.214]
White	1.049***	1.066***	1.239***	1.234***	1.202***	1.173***	1.170***	1.016***
	(0.0244)	(0.0246)	(0.0664)	(0.0671)	(0.0556)	(0.0526)	(0.181)	(0.154)
	[0.0436]	[0.0434]	[0.0963]	[0.0943]	[0.0957]	[0.0939]	[0.221]	[0.193]
R-squared	0.190	0.202	0.531	0.528	0.531	0.527	0.565	0.568
CBSA X Group FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Base Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SES Controls		Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rounded Obs	56500	56500	56000	56000	56000	56000	38000	38000
S-W F-Stat θ_B			10.66	12.14	6.45	10.09	5.76	14.90
S-W F-Stat θ_W			23.81	23.74	26.26	28.55	11.96	16.98
S-W F-Stat $\tilde{\rho}_B$			7.37	8.41	4.26	5.55	2.68	3.53
S-W F-Stat $\tilde{\rho}_W$			18.45	19.45	17.22	18.61	5.66	6.02

Notes: Unit of observation is census tract by race and education. Data comes from the first difference of 1960 to 1970 using restricted Census microdata. Fixed effects are at the CBSA (Core-based statistical area) by race and education group level. Standard errors are cluster-robust with clusters at the tract-level. Conley standard errors for spatial correlation within a 1km radius are reported in brackets. The base controls are change in log rent and 5 binary indicators for distance from highways built between 1960 and 1970 in 1-mile wide bins, all interacted with race and education. Redlining fixed effects are included in all specifications. The socio-economic status (SES) controls are change in log income, change in log percentage high school graduate, change in log percentage bottom income quintile, change in log percentage top income quintile, change in log home values, all interacted with race and education. The geographic controls are log distance from the central business district, rivers, lakes, shores, ports, historical railroads, canals, and historical large urban roads, all interacted with race and education. All specifications include the Borusyak and Hull (2023) control for CMA in large roads interacted with race and education. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. Sanderson-Windmeijer multivariate F statistics for weak instruments with multiple endogenous regressors are reported. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Change in Amenities over Distance from Highway

	(1)	(2)	(3)	(4)	(5)
		$\Delta \log b_{igr}$		$\Delta \log B_{igr}$	$\Delta \log b_{igr}$
Variables	OLS	OLS	0.5 mi bins	OLS	Placebo
Dist Highway (mi = 1)	-0.453***	-0.119**		-0.124**	0.00565
	(0.0501)	(0.0516)		(0.0517)	(0.137)
Dist Highway (mi = 2)	-0.379***	-0.0933*		-0.125**	-0.0707
	(0.0499)	(0.0515)		(0.0515)	(0.0997)
Dist Highway (mi = 3)	-0.223***	0.000345		-0.0343	-0.0752
	(0.0531)	(0.0552)		(0.0553)	(0.0910)
Dist Highway (mi = 4)	-0.0795	0.0824		0.0458	0.0307
	(0.0596)	(0.0604)		(0.0606)	(0.0899)
Dist Highway (mi = 5)	0.0369	0.143**		0.140**	0.0437
	(0.0642)	(0.0638)		0.0638)	(0.0793)
Dist Highway (mi = 0.5)			-0.191***		
			(0.0581)		
Dist Highway (mi = 1)			-0.0651		
			(0.0572)		
Dist Highway (mi = 1.5)			-0.0994*		
			(0.0588)		
Dist Highway (mi = 2)			-0.0888		
			(0.0573)		
Dist Highway (mi = 2.5)			0.0116		
			(0.0618)		
Dist Highway (mi = 3)			-0.0153		
			(0.0667)		
Dist Highway (mi = 3.5)			0.0703		
			(0.0738)		
Dist Highway (mi = 4)			0.0955		
			(0.0731)		
Dist Highway (mi = 4.5)			0.140*		
			(0.0808)		
Dist Highway (mi = 5)			0.146*		
			(0.0807)		
R-squared	0.028	0.052	0.052	0.050	0.069
CBSA X Group FE	Yes	Yes	Yes	Yes	Yes
Geo Controls		Yes	Yes	Yes	Yes
Rounded Obs	49500	49500	49500	49500	9000

Notes: Unit of observation is census tract by race and education. Data comes from the first difference of 1960 to 1970 using restricted Census microdata. Fixed effects are at the CBSA (Core-based statistical area) by race and education group level. Standard errors are cluster-robust with clusters at the tract-level. There are 5 binary indicators for distance from highways built between 1960 and 1970 in 1-mile wide bins (the value displayed is the upper end of the bin). In Column 3, the bins are split further into 0.5-mile wide bins. The geographic controls are log distance from the central business district, rivers, lakes, shores, ports, historical railroads, canals, and historical large urban roads, all interacted with race and education. The sample is limited to tracts within 10 miles of constructed highway routes. For the placebo specification in Column 5, the sample is restricted to not be within 5 miles of a highway and is limited to tracts within 10 miles of historical large urban roads. The geographic control for distance from historical large urban roads is dropped since it is now the endogenous variable. All specifications include redlining fixed effects. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. Parameters used to invert for dependent variables are the same as in Table 10. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Key Model Parameters

Parameters	Source
Labor Supply Elasticity $\phi = 3$	Ahlfeldt et al. (2015), Monte et al. (2018), Morten and Oliveira (2018), Severen (2021)
Commuting Elasticity $\nu_{LB} = 4.20$, $\nu_{HB} = 3.65$, $\nu_{LW} = 4.71$, $\nu_{HW} = 4.15$ $\kappa_{LB} = 1.4$, $\kappa_{HB} = 1.22$, $\kappa_{LW} = 1.57$, $\kappa_{HW} = 1.38$	Estimated in Table 5 Implied with $\phi = 3$
Residential Elasticity $\theta_B = 0.35$, $\theta_W = 0.8$	Estimated in Table 6
Racial Preferences $\rho_B = 0$, $\rho_W = 1$	Estimated in Table 6
Non-Housing Consumption Share $\beta_{LB} = 0.66$, $\beta_{HB} = 0.78$, $\beta_{LW} = 0.70$, $\beta_{HW} = 0.79$	Calibrated to CEX in Appendix 5.1
Production Labor Share $\alpha = 0.7$	Greenwood et al. (1997)
Elasticity of Substitution by Race and Education $\sigma^r = 8$, $\sigma^g = 2$	Card (2009), Boustan (2009)
Agglomeration $\gamma^A = 0.07$	Rosenthal and Strange (2004), Kline and Moretti (2014)
Housing Supply Elasticity $\mu^{cbd} = 0.35$, $\mu^{sub} = 0.25$	Baum-Snow and Han (2021)
Highway Localized Costs $b^{HW} = 0.203$, $\eta = 0.612$	Estimated in Table 7
Institutional Exclusion $E = 0.073$	Estimated in Table 10

Notes: Parameters ν_{gr} come from Table 5 Panel C. ϕ is set to a value from the literature. β_{gr} come from Table A.16. θ_r come from the midpoint of estimates in Table 6. α is set to a value from the literature. σ^r and σ^g are set to values from the literature. μ^{cbd} and μ^{sub} are set to values from the literature. γ^A is set to a value from the literature. ρ_N is set to 0 since the estimates from Table 6 are not distinguishable from zero. ρ_W is set to be within the lower range of the confidence intervals from Table 6 and not greater than 1. b^{HW} and η come from fitting two values from Table 7 Column 5, and E comes from Table 10 Column 4 where $E = \exp(-0.914/\theta_N)$.

Table 9: Welfare Changes (%) by Race and Education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Race	Black	Black	White	White	By Race		By Educ	
x Educ	<HS	HS+	<HS	HS+	Black	White	<HS	HS+
	Panel A – The Impacts of the Interstates							
General Equilibrium								
Welfare Change	-1.45	-0.16	2.69	3.01	-1.04	2.86	2.07	2.79
Direct Impacts								
Commuting Benefits	6.21	7.03	8.07	7.75	6.47	7.90	7.79	7.70
Localized Costs	-8.05	-8.00	-6.46	-5.94	-8.03	-6.18	-6.70	-6.08
Welfare Change	-1.84	-0.97	1.61	1.81	-1.56	1.72	1.10	1.62
Reallocation Only								
Welfare Change	-1.33	-0.25	2.91	3.00	-0.98	2.96	2.28	2.78
	Panel B – Mechanisms							
Partial Equilibrium								
Welfare Change	-1.17	-0.13	2.85	2.97	-0.84	2.91	2.25	2.76
No Spillovers								
Welfare Change	-1.44	-0.21	2.79	3.00	-1.05	2.90	2.16	2.78
Highway Impacts Separately								
Commuting Benefits	7.32	8.77	10.16	9.87	7.78	10.01	9.74	9.80
Localized Costs	-8.08	-8.05	-6.57	-6.03	-8.07	-6.28	-6.79	-6.17
	Panel C – Interaction with Institutional Segregation							
General Eq, No BD								
Welfare Change	-0.99	0.11	2.69	3.01	-0.64	2.86	2.14	2.81
General Eq, Same Fund Amen								
Welfare Change	1.00	1.12	2.64	3.00	1.04	2.83	2.40	2.87

Notes: Welfare calculations are based on data from the restricted Census in 1960. Direct impacts come from the linear approximation in Section 6.1. The general equilibrium simulation allows wages to respond in equilibrium. The partial equilibrium simulation keeps wages fixed. No institutions adjusts fundamental amenities for Black households by parameter E in redlined areas. Same fundamental amenities sets fundamental amenities of Black households to those of White households. The general equilibrium simulation with no institutions (same fundamental amenities) adds the highway impacts in the counterfactual world with no institutions (same fundamental amenities). All values are rounded to four significant digits to meet Census disclosure rules. Columns 5–8 are weighted averages of the race by education welfare numbers using population weights from the bottom of Table 1.

Table 10: Border Discontinuity Decomposition

	(1)	(2)	(3)	(4)
	Panel A – Black			
Variables	$\log L_{igr}$	$\theta_r \log B_{igr}$	$\theta_r \log b_{igr}$	+ SES Cont
ψ_B: Border RD	1.425***	1.370***	1.266***	0.914***
	(0.226)	(0.238)	(0.209)	(0.181)
Bandwidth (mi)	0.495	0.447	0.509	0.496
Order of Poly.	1	1	1	1
Education FE	Yes	Yes	Yes	Yes
Border FE	Yes	Yes	Yes	Yes
Rounded Obs	13000	13000	13000	13000
	Panel B – White			
Variables	$\log L_{igr}$	$\theta_r \log B_{igr}$	$\theta_r \log b_{igr}$	+ SES Cont
ψ_W: Border RD	-0.546***	-0.603***	0.00305	0.132
	(0.122)	(0.112)	(0.0971)	(0.0937)
Bandwidth (mi)	0.358	0.398	0.297	0.305
Order of Poly.	1	1	1	1
Education FE	Yes	Yes	Yes	Yes
Border FE	Yes	Yes	Yes	Yes
Rounded Obs	13500	13500	13500	13500

Parameters

$\theta_B = \theta_W = 0.9$

$\theta_B \rho_B = -0.3, \theta_W \rho_W = 1.20$

$\beta_{LB} = 0.66, \beta_{HB} = 0.78, \beta_{LW} = 0.70, \beta_{HW} = 0.79$

Notes: Unit of observation is census tract by border in the redlining maps. Data comes from the 1960 restricted Census microdata. The dependent variable is residualized on fixed effects for education within race and on border fixed effects for all specifications. Controls are log percentage high school grad, log population density, log average income, log percentage top quintile, log percentage bottom quintile, and log home values. Coefficients on controls are estimated with redlining fixed effects. Sample is limited to tracts that are at least 0.1 miles away from possible physical barriers such as historical large urban roads, constructed highways in 1960, or historical railroads and also at least 0.1 miles away from a school district boundary. The bandwidth is chosen optimally following Calonico et al. (2014). Distance from the border is measured in miles. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. Parameters used to invert for dependent variables are displayed at the bottom. *** p<0.01, ** p<0.05, * p<0.1

Table 11: Summary Statistics by Race

	(1)	(2)	(3)	(4)
Panel A	White		Black	
Variable	Mean (SD)	Rounded N	Mean (SD)	Rounded N
HS Grad Rate	0.949		0.893	
	(0.221)	6122000	(0.309)	628000
College Grad Rate	0.386		0.236	
	(0.487)	6122000	(0.425)	628000
Adjusted Gross Income (2018 $K)	101700		49210	
	(344600)	25750000	(106200)	3662000
Wage & Salary Income (2018 $K)	88290		48090	
	(160200)	25200000	(65200)	3579000
Individual Earnings (2018 $K)	58240		37420	
	(303000)	23800000	(47260)	3644000
Child Household Income Rank	56.3		34.4	
	(27.8)	25750000	(24.9)	3662000
Child Individual Income Rank	54.4		43.2	
	(28.7)	23800000	(26.4)	3644000
Average Parental Income (2018 $K)	81110		49520	
	(160700)	27220000	(77120)	4218000
Parent Household Income Rank	55.5		34.4	
	(27.8)	27220000	(26.6)	4218000
Panel B	White		Black	
Upward-Downward Mobility	Par Quintile 1	Par Quintile 5	Par Quintile 1	Par Quintile 5
P(Child Quint = 1 | Par Quint = X)	0.249	0.072	0.446	0.207
P(Child Quint = 5 | Par Quint = X)	0.123	0.409	0.038	0.173
Percentage in Quintiles	Quintile 1	Quintile 5	Quintile 1	Quintile 5
P(Child Quintile = X)	0.136	0.246	0.366	0.068
P(Parent Quintile = X)	0.141	0.237	0.399	0.082

Panel C	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
All Race Groups	No Race	White	Black	Asian	Hispanic	Indigenous	NHPI	Other
Percentage	0.0838	0.7160	0.1109	0.0095	0.0720	0.0060	0.0009	0.0008
Population	3188000	27240000	4219000	361000	2739000	228000	34000	30000

Note: High school and college graduation rates are from ACS surveys. Adjusted Gross Income and Wage & Salary income are from the 1040 forms during the years in which the child is aged 35-39. Individual earnings are from W-2 forms during the years in which the child is aged 35-39, except for the birth cohorts of 1964-1969. Their earnings are measured during ages 41-45 as the W-2 data begins in 2005. Children are assigned percentile ranks relative to all other children in their birth cohort, while parents are ranked relative to all parents with children in the same birth cohort. Upward-downward mobility is calculated using household income. All racial groups exclude individuals of Hispanic ethnicity. NHPI is an abbreviation for Native Hawaiian and Pacific Islander. Counts are rounded in the following manner: numbers between 10,000 and 99,999 are rounded to the nearest 500; between 100,000 and 9,999,999 to the nearest 1,000; and above 10,000,000 to the nearest 10,000 to meet Census disclosure rules.

Table 12: Summary Statistics for County and Tract Movers

Panel A	(1)	(2)	(3)	(4)
	County Movers		Tract Movers	
Variable	1 Move	0 or 2+ Moves	1 Move	0 or 2+ Moves
High School Graduation Rate	0.947	0.950	0.955	0.953
SD	(0.224)	(0.218)	(0.207)	(0.212)
Rounded N	1376000	4791000	1501000	3183000
College Graduation Rate	0.379	0.386	0.421	0.406
SD	(0.485)	(0.487)	(0.494)	(0.491)
Rounded N	1376000	4791000	1501000	3183000
Adjusted Gross Income (2018 $K)	97.72	97.58	104.6	100.3
SD	(316.3)	(334.2)	(378.5)	(333.1)
Rounded N	6049000	20590000	6672000	14050000
Wage & Salary Income (2018 $K)	85.22	85.37	90.38	87.85
SD	(153.1)	(157.0)	(166.8)	(169.1)
Rounded N	5920000	20170000	6539000	13770000
Individual Earnings (2018 $K)	56.47	56.87	59.75	58.44
SD	(139.8)	(214.1)	(126.4)	(255.4)
Rounded N	5637000	19180000	6224000	13130000

Panel B	(1)	(2)	(3)	(4)	(5)	(6)
	County Movers - White			Tract Movers - White		
Mover Types	0 Moves	1 Move	2+ Moves	0 Movers	1 Move	2+ Moves
Percentage	0.609	0.232	0.160	0.380	0.323	0.297
Population	15200000	5791000	3994000	7349000	6247000	5744000
	County Movers - Black			Tract Movers - Black		
Mover Types	0 Moves	1 Move	2+ Moves	0 Movers	1 Move	2+ Moves
Percentage	0.683	0.196	0.121	0.316	0.310	0.373
Population	2286000	656000	405000	845000	829000	998000

Note: High school and college graduation rates come from the ACS surveys. Adjusted Gross Income and Wage & Salary income come from the 1040 forms during the years in which the child is aged 35-39. Individual earnings come from W-2 forms during the years in which the child is aged 35-39, except for the birth cohorts of 1964-1969. Their earnings are measured during ages 41-45 as the W-2 data begins in 2005. Children are assigned percentile ranks relative to all other children in their birth cohort, while parents are ranked relative to all parents with children in the same birth cohort. All racial groups exclude individuals of Hispanic ethnicity. Moves are calculated starting when the 1040 data is first available in 1974 up until age 23. Counts are rounded in the following manner: numbers between 10,000 and 99,999 are rounded to the nearest 500; between 100,000 and 9,999,999 to the nearest 1,000; and above 10,000,000 to the nearest 10,000 to meet Census disclosure rules.

Table 13: Rank-Rank Correlations by Race

	(1)	(2)	(3)	(4)	(5)	(6)
	Child Household Income Rank					
Variables	Pooled	White	Black	Hispanic	Asian	Indigenous
Par HH Income Rank	0.334***	0.295***	0.217***	0.241***	0.238***	0.260***
	(0.000161)	(0.000191)	(0.000498)	(0.000638)	(0.00167)	(0.00219)
Constant	34.98***	39.77***	26.81***	36.98***	49.11***	27.95***
	(0.00951)	(0.0120)	(0.0200)	(0.0294)	(0.120)	(0.0953)
R-squared	0.112	0.086	0.055	0.056	0.058	0.070
Rounded Obs	3.4e+07	2.57e+07	3.66e+06	2.47e+06	343000	198000

Note: Parental household income comes from the 1040 forms and is the average of the first four years of tax data available post-birth of the child. Household income of the child comes from the 1040 forms and is the average of the five years during which the child is aged 35-39. Children are assigned percentile ranks relative to all other children in their birth cohort, while parents are ranked relative to all parents with children in the same birth cohort. The pooled category includes all racial groups, not solely White and Black individuals. All racial groups exclude individuals of Hispanic ethnicity. Counts are rounded in the following manner: numbers between 10,000 and 99,999 are rounded to the nearest 500; between 100,000 and 9,999,999 to the nearest 1,000; and above 10,000,000 to the nearest 10,000 to meet Census disclosure rules. Robust standard errors are included in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 14: County and Tract Variation in Predicted Income Ranks (P25)

	(1)	(2)	(3)	(4)	(5)
	White		Black		
	County Percentile		County Percentile		Correlation
Panel A - County Mean	10th	90th	10th	90th	Across Race
Pred Household Income Rank	41.26	51.03	28.02	34.21	0.535
Pred Individual Income Rank	42.48	51.81	28.77	35.20	0.540
	Tract Percentile		Tract Percentile		Correlation
Panel B - Tract Mean	10th	90th	10th	90th	Across Race
Pred Household Income Rank	42.92	56.56	28.41	39.13	0.180
Pred Individual Income Rank	43.83	57.29	29.19	40.00	0.175

Note: Predicted income rank is computed by estimating linear rank-rank correlations for each racial group in each geographic unit (either county or tract) and then predicting the rank of children from the 25th percentile of the parent income distribution. The 10th and 90th percentiles of predicted ranks are displayed, and the correlation across race groups is calculated with analytical weights based on years spent in each location and by trimming the bottom and top 1% of predicted ranks. All racial groups exclude individuals of Hispanic ethnicity.

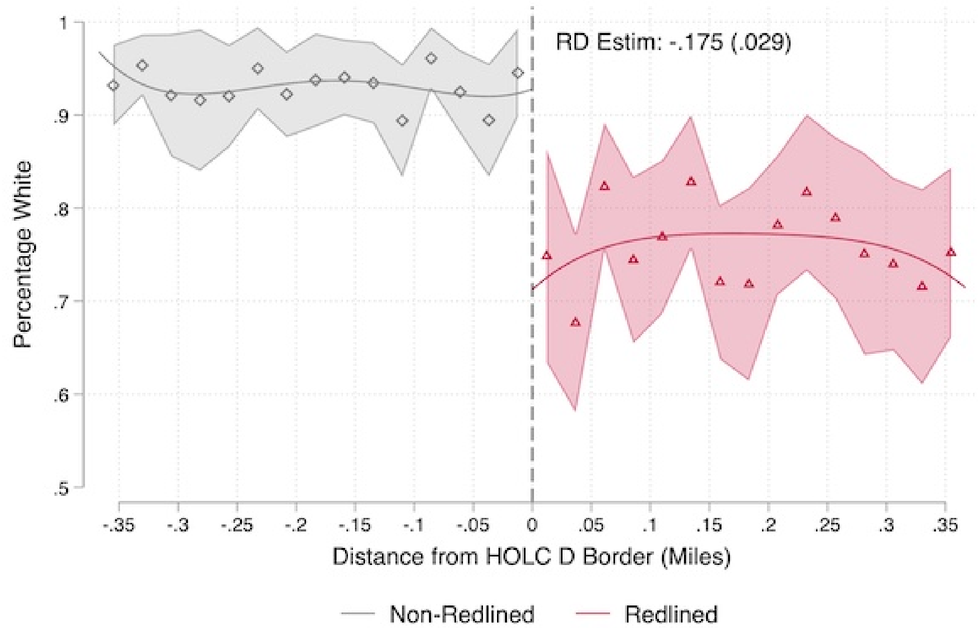
Table 15: Movers Exposure Effects By Race and Gender

	(1)	(2)	(3)	(4)
	White		Black	
Variables	Male	Female	Male	Female
$\leq$ 23 Exposure Slope	0.0128	0.0104	0.0171	0.0116
	(0.0017)	(0.0016)	(0.0029)	(0.0026)
$\leq$ 23 Intercept	0.533	0.594	0.301	0.366
	(0.030)	(0.029)	(0.040)	(0.040)
$>$ 23 Exposure Slope	0.0094	0.0182	0.0214	-0.0045
	(0.0081)	(0.0075)	(0.0260)	(0.0225)
$>$ 23 Intercept	0.424	0.549	0.316	0.286
	(0.0384)	(0.0291)	(0.085)	(0.082)
Rounded Obs	2597000	2628000	236000	301000

Note: Predicted income ranks of origin and destination counties are calculated by race with one-time movers removed to eliminate a mechanical correlation between children's income and the predicted income rank of the county. The specification assumes a linear relationship between years of exposure to the destination county relative to the origin county prior to age 23 and post age 23. One-time movers who move up until age 28 are included in the sample. All racial groups exclude individuals of Hispanic ethnicity. Counts are rounded in the following manner: numbers between 10,000 and 99,999 are rounded to the nearest 500; between 100,000 and 9,999,999 to the nearest 1,000; and above 10,000,000 to the nearest 10,000 to meet Census disclosure rules.

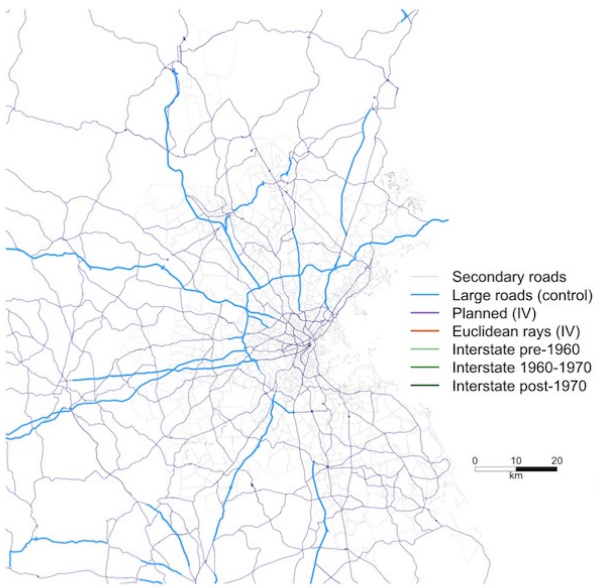
19 Figures

Figure 1: Border Discontinuity for Percentage White in 1960 at HOLC D Border

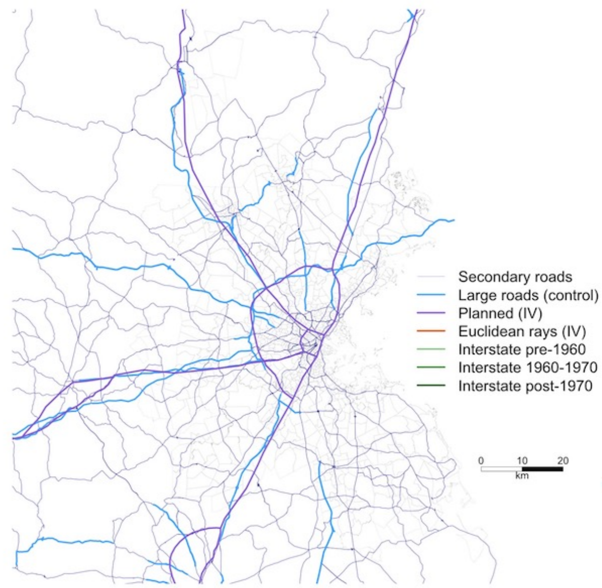


Notes: Unit of observation is census tract by HOLC border pair. Data comes from 1960 tract-level aggregates retrieved from IPUMS NHGIS. The left side of the discontinuity is non-redlined and the right side is redlined. The order of polynomial fit is 4 with optimal bandwith of 0.368 chosen following Calonico et al. (2014), and the kernel is Epanechnikov. Redlined tracts are tracts where more than 80% of the area is redlined. There are 15 bins on the left (N=752) and 15 bins on the right (N=752). The estimated coefficient is from the balanced sample RD shown in Table A.22 Panel A with the order of polynomial set to 1, the same optimal bandwidth of 0.368, and the same number of effective observations (however N=2957 enter into the regression for both sides).

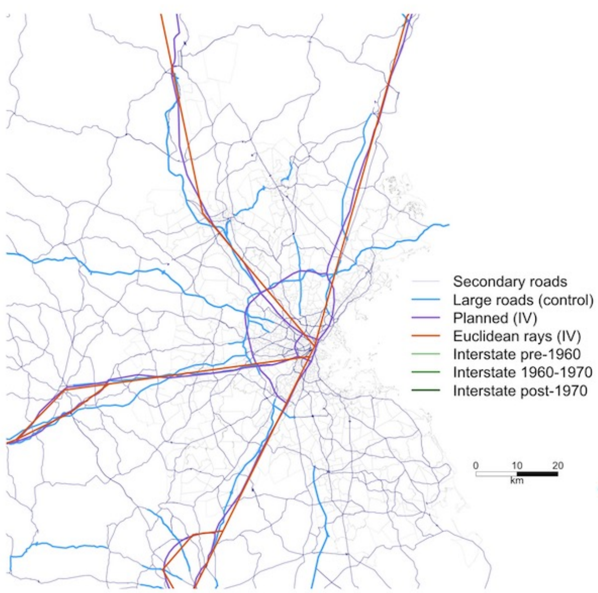
Figure 2: Historical Road Networks and Highway Routes for the Boston Metropolitan Area



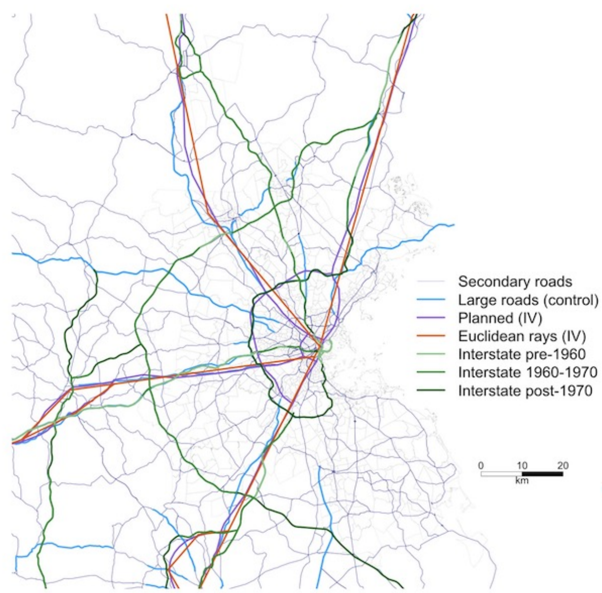
(a) Historical Urban Roads



(b) Planned Routes



(c) Euclidean Rays



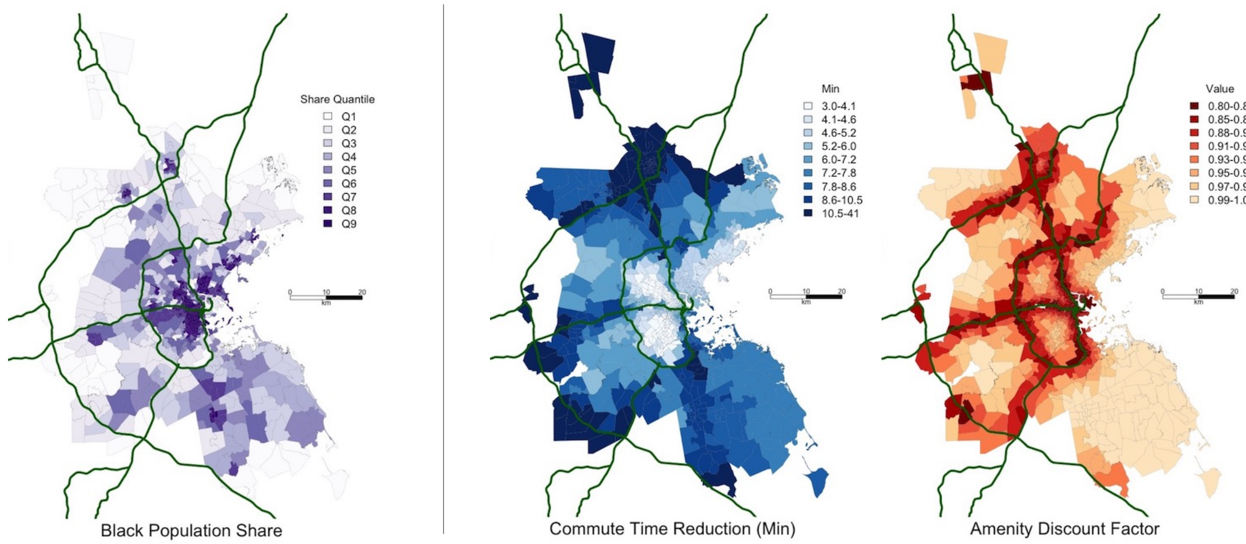
(d) Interstate Highways

Notes: Historical urban roads are split into two categories: smaller roads and large roads (superhighways in the legend of Shell Atlases) with large roads in light blue. These large roads were candidates for Interstate construction, and as is evident in Panel 6a compared to Panel 2d, Interstate routes were often built on top of these large roads. Planned routes are digitized from Yellow Book maps. Euclidean rays connect major cities in the 1947 highway plan. Interstate routes are the constructed Interstate network.

Figure 3: Black Population Relative to Highway Impacts for the Boston Metropolitan Area in 1960

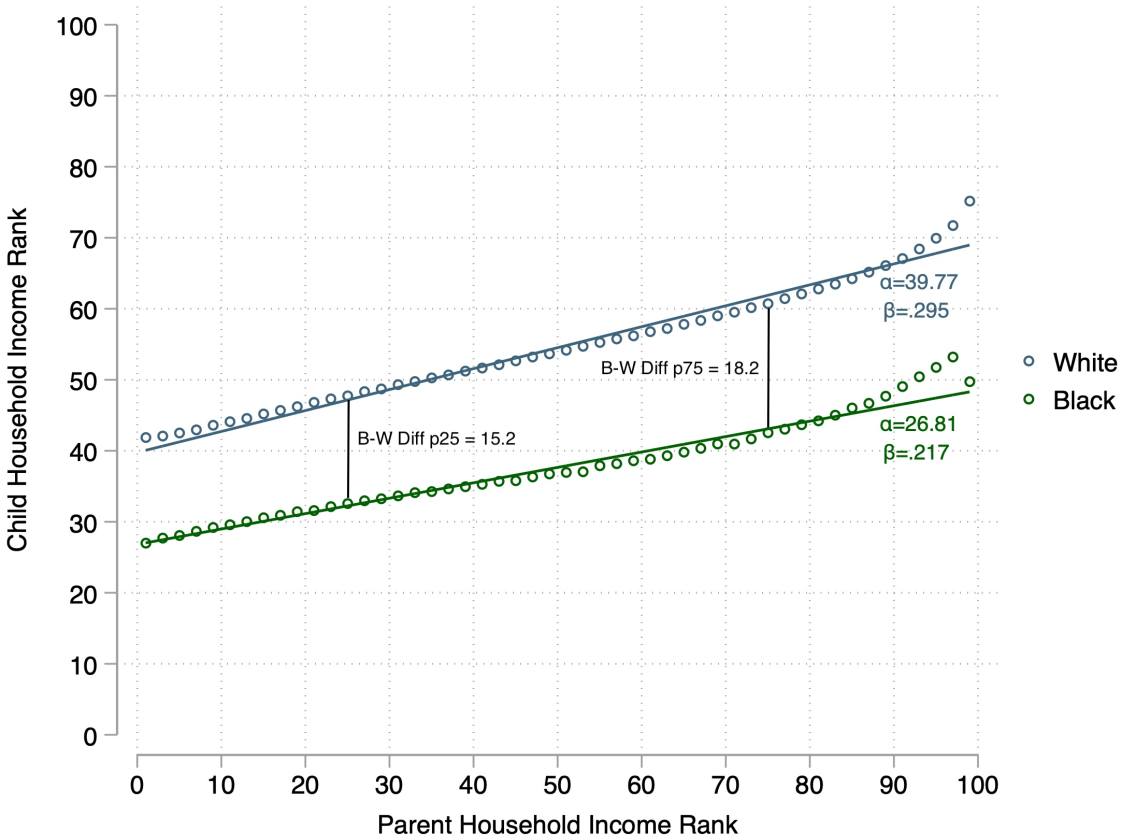
Impacts of Interstate highways

Black Population in 1960 | Commute time changes | Localized costs



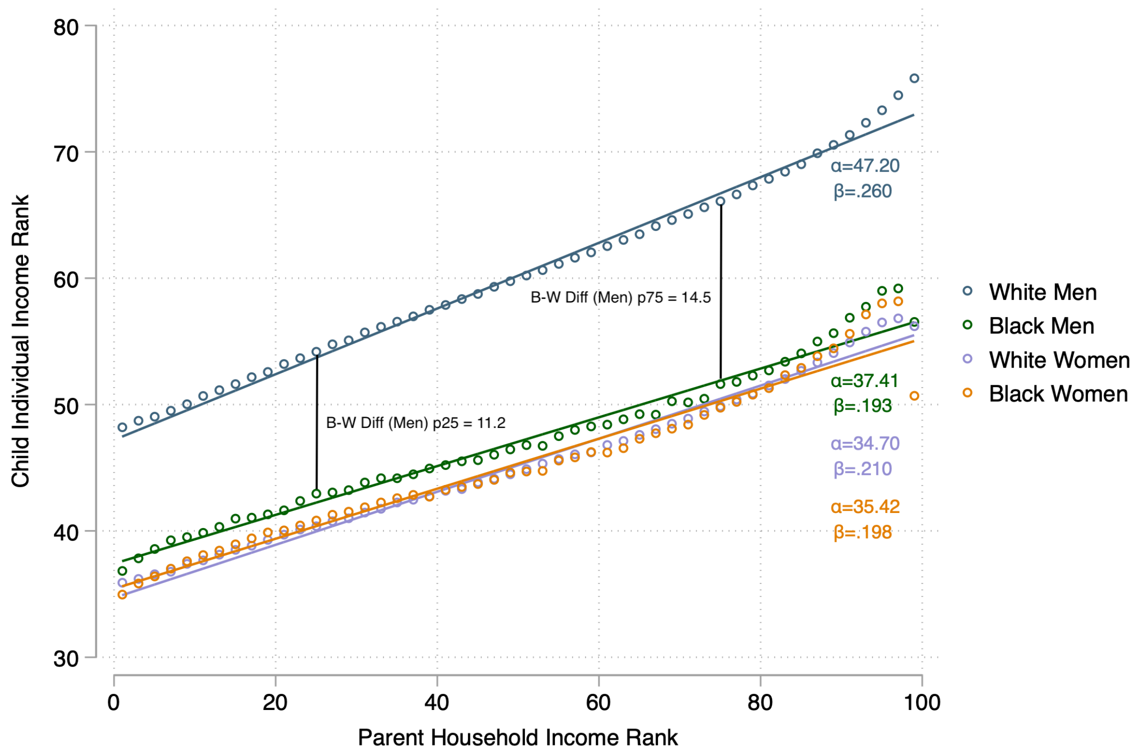
Notes: Unit of observation is census tract. Data for the Black population share comes from 1960 tract-level aggregates retrieved from IPUMS NHGIS. Commute time changes come from the author's calculations as the difference between commute times for the historical road network and for the entire Interstate network overlayed on the historical road network. Local costs are calculated by taking the estimate from Table 8 and applying it to census tracts using the distance of the centroid of the tract to the nearest Interstate highway. The sample of tracts is limited to those where population is observed in 1960.

Figure 4: Child Household Income Rank by Race



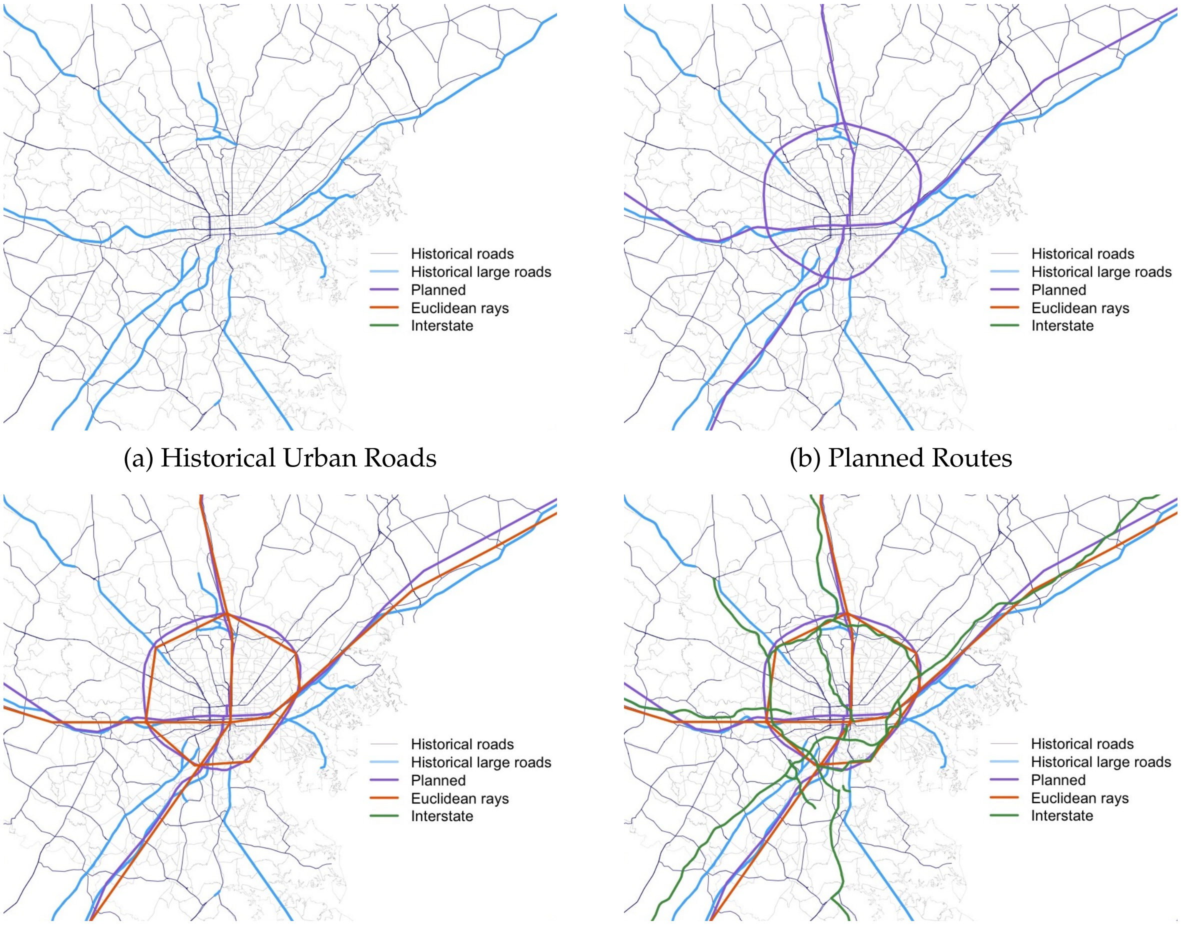
Note: Average child household income rank is measured over 50 bins across parental household income rank separately for White and Black individuals using the 1040 tax data. Parental household income comes from the 1040 forms and is the average of the first four years of tax data available post-birth of the child. Household income of the child comes from the 1040 forms and is the average of the five years during which the child is aged 35-39. Children are assigned percentile ranks relative to all other children in their birth cohort, while parents are ranked relative to all parents with children in the same birth cohort. All racial groups exclude individuals of Hispanic ethnicity. The best-fit lines are estimated using an OLS regression on the individual observations and are displayed in Table 13. The slopes β_r and intercepts α_r from these regressions are reported for each race. White-black differences in mean child household income rank are reported at the 25th and 75th percentiles of the parent income distribution.

Figure 5: Child Individual Income Rank by Race and Gender



Note: Average child individual income rank is measured over 50 bins across parental household income rank separately for White and Black individuals by gender using the 1040 tax data. Parental household income comes from the 1040 forms and is the average of the first four years of tax data available post-birth of the child. Individual earnings come from W-2 forms during the years in which the child is aged 35-39, except for the birth cohorts of 1964-1969. Their earnings are measured during ages 41-45 as the W-2 data begins in 2005. Children are assigned percentile ranks relative to all other children in their birth cohort, while parents are ranked relative to all parents with children in the same birth cohort. All racial groups exclude individuals of Hispanic ethnicity. The best-fit lines are estimated using an OLS regression on the individual observations and are displayed in Table A.28. The slopes β_r and intercepts α_r from these regressions are reported for each race and gender. White-black differences for men in mean child individual income rank are reported at the 25th and 75th percentiles of the parent income distribution.

Figure 6: Historical Road Networks and Highway Routes for the Baltimore Metropolitan Area



(a) Historical Urban Roads

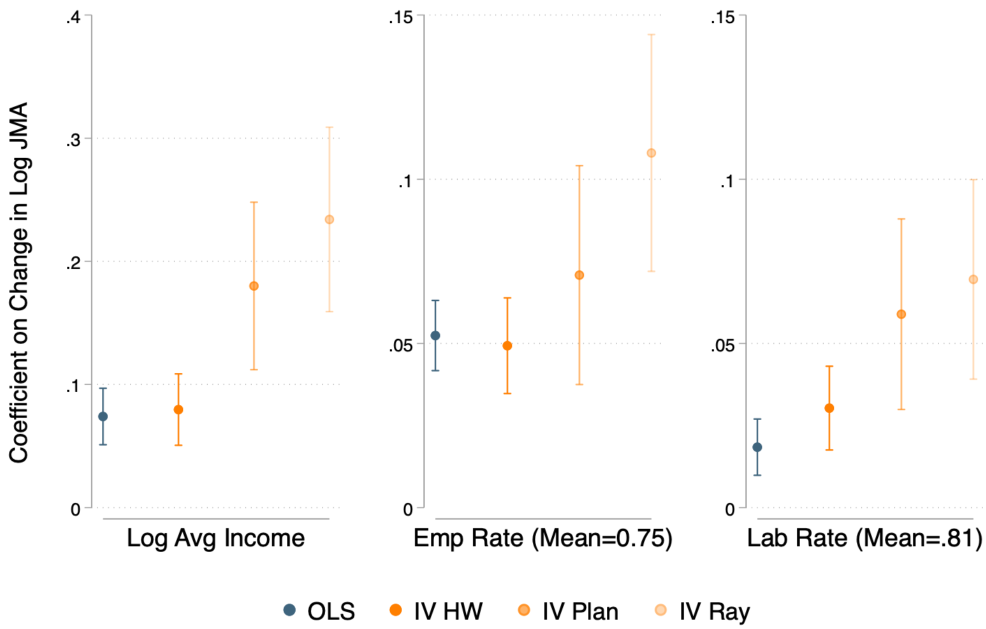
(b) Planned Routes

(c) Euclidean Rays

(d) Interstate Highways

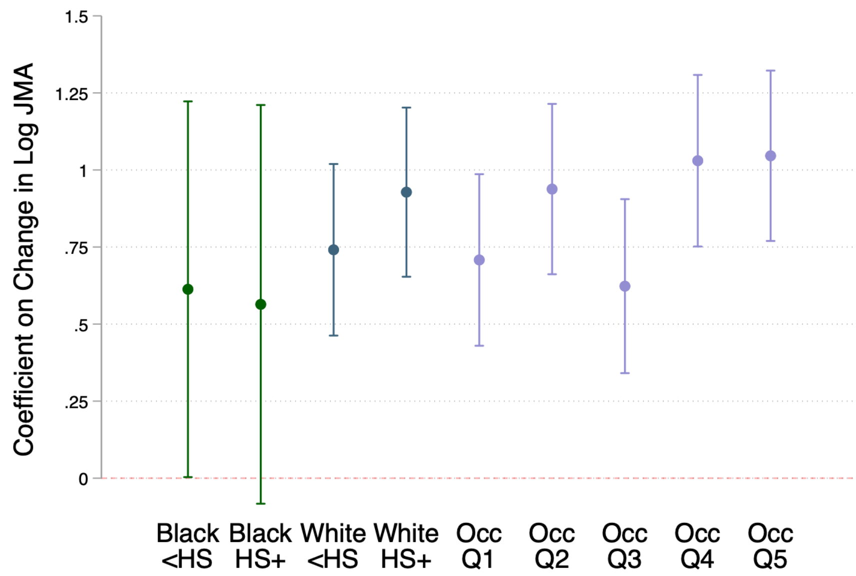
Note: Historical urban roads are split into two categories: smaller roads and large roads (superhighways in the legend of Shell Atlases) with large roads in light blue. These large roads were candidates for interstate construction, and as is evident in Panel 6a compared to Panel 6d, interstate routes were often built on top of these large roads. Planned routes are digitized from Yellow Book maps. Euclidean rays connect major cities in the 1947 highway plan. Interstate routes are the constructed interstate network.

Figure 7: The Effect of Job Market Access Improvements on Changes in Tract-Level Income, Employment Rate, and Labor Force Participation Rate (1960-1970)



Note: Tract characteristics are calculated using the Decennial Census in 1960 and 1970. Average income, employment rate, and labor force participation rate is calculated among men aged 16 and up. Employment rate has men aged 16+ as the denominator and employment among men aged 16+ as the numerator. Coefficient estimates, standard errors, dependent variable means and F-stats are reported in Table A.31

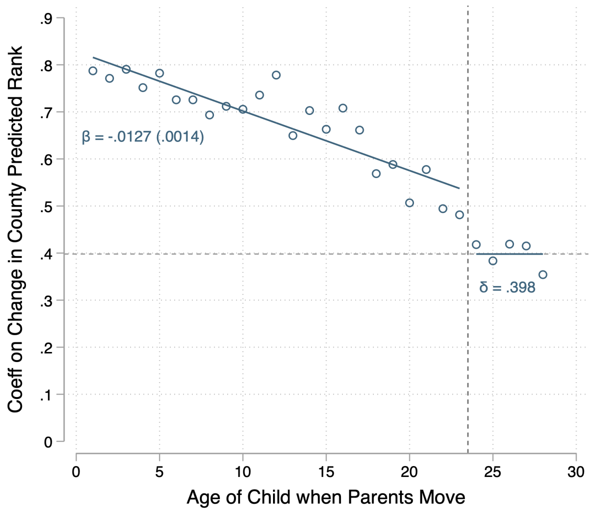
Figure 8: Population Responses to Job Market Access Improvements by Group (1960-1970)



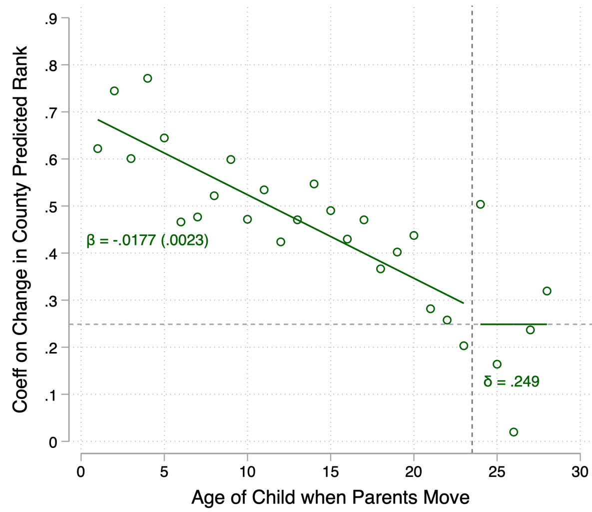
Note: Tract-level population is calculated using the Decennial Census in 1960 and 1970. Population by high school graduate status, race (White or Black), and occupational quintile is recorded among individuals aged 16 and up. Occupations are ordered based on nation-wide median income among the employed into five bins.

Figure 9: Exposure to County Predicted Income Rank over Age at Move for Movers

(a) Income Rank of White Males

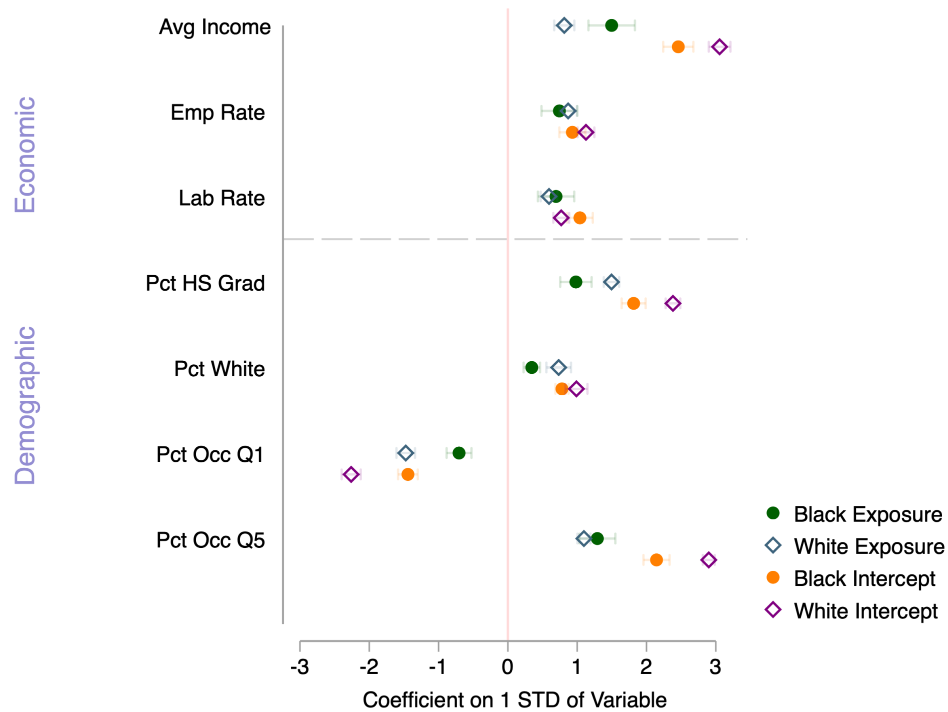


(b) Income Rank of Black Males



Note: Predicted income ranks of origin and destination counties are calculated by race with one-time movers excluded to eliminate a mechanical correlation between children's income and the predicted income rank of the county. Household income of the child comes from the 1040 forms and is the average of the five years during which the child is aged 35-39. Children are assigned percentile ranks relative to all other children in their birth cohort. The specification calculates the coefficients for child income rank in each age at move bin from age 1 up until age 28. The coefficients b_m can be interpreted as how children's income ranks change when they move at age m to a county with a 1 percentile higher predicted individual income rank in adulthood for children of the same race. Only movers who move once from birth until age 28 are included in the sample. Estimate β from a parametric specification assuming a linear relationship between children income rank and age at move bin coefficients up until age 23 are displayed (with standard errors in parentheses). The intercept δ is the mean of the age at move bin coefficients post age 23. All racial groups exclude individuals of Hispanic ethnicity.

Figure 10: Causal Impacts of Tracts and Tract-Level Characteristics



Note: Causal impacts of tracts come from a movers design along tract characteristics from origin to destination. Tract characteristics are calculated using the Decennial Census in 1970. Percentage high school graduate, percentage White, and percentages in each occupational quintile is calculated among individuals aged 16 and up. Occupations are ordered based on nation-wide median income among the employed into five bins. Average income, employment rate, and labor force participation rate is calculated among men aged 16 and up. Employment rate has men aged 16+ as the denominator and employment among men aged 16+ as the numerator.

Appendices

1 Tables

Table A.1: Housing Price Discrimination in Rents and Home Values in 1960

	(1)	(2)	(3)	(4)
Variables		*Panel A* – Log Rent		
Black	-0.155***	-0.0250***	-0.0295***	0.0260***
	(0.0117)	(0.00719)	(0.00939)	(0.00584)
Redlined	-0.340***		-0.212***	
	(0.0107)		(0.00804)	
Black x Redlined	0.183***	0.0845***	0.0929***	0.0508***
	(0.0174)	(0.0103)	(0.0138)	(0.00873)
Constant	4.272***			
	(0.00514)			
R-squared	0.104	0.433	0.394	0.592
Tract FE		Yes		Yes
Quality Controls			Yes	Yes
Rounded Obs	1729000	1729000	1729000	1729000
Variables		*Panel B* – Log Home Value		
Black	-0.332***	-0.0453***	-0.143***	-0.0146**
	(0.0175)	(0.0112)	(0.0121)	(0.00675)
Redlined	-0.361***		-0.205***	
	(0.0171)		(0.0134)	
Black x Redlined	0.142***	0.0602***	0.0967***	0.0509***
	(0.0319)	(0.0189)	(0.0245)	(0.0123)
Constant	9.625***			
	(0.00567)			
R-squared	0.078	0.522	0.404	0.658
Tract FE		Yes		Yes
Quality Controls			Yes	Yes
Rounded Obs	1562000	1562000	1562000	1562000

Notes: Unit of observation is household. Household level data comes from the 1960 Census microdata. Fixed effects are at the census tract level. Standard errors are cluster-robust with clusters at the tract level. The quality controls include categorical variables for availability of air conditioning, dryer, elevator, freezer, hot water, kitchen, shower, basement, toilet, and the type of heating, type of fuel for cooking, type of fuel for heat, type of fuel for water, source of water, source of water, sewage facilities, number of stories, number of rooms, number of bathrooms, number of bedrooms, and year built. Redlined tracts are tracts where more than 80% of the area is redlined. Observation counts are rounded to nearest 1000 to meet Census disclosure rules. *** p<0.01, ** p<0.05, * p<0.1

Table A.2: Percentage of the 1950 Population by Race and Education in HOLC D (Redlined) Areas for the 10 Most Populous Cities (Core-Based Statistical Areas)

Black		White		<HS		HS Grad	
CBSA	% HOLC D	CBSA	% HOLC D	CBSA	% HOLC D	CBSA	% HOLC D
New York	0.85	New York	0.36	New York	0.45	New York	0.28
Chicago	0.93	Chicago	0.32	Chicago	0.44	Los Angeles	0.20
Philadelphia	0.87	Los Angeles	0.25	Los Angeles	0.34	Chicago	0.26
Detroit	0.84	Detroit	0.30	Philadelphia	0.53	Boston	0.20
Baltimore	0.79	Philadelphia	0.40	Detroit	0.44	Detroit	0.25
Los Angeles	0.77	Boston	0.29	Boston	0.36	Philadelphia	0.31
St. Louis	0.78	Cleveland	0.31	St. Louis	0.36	San Francisco	0.28
New Orleans-	0.82	San Francisco	0.33	Cleveland	0.44	Cleveland	0.26
Cleveland	0.90	St. Louis	0.25	Baltimore	0.42	Pittsburgh	0.24
Memphis	0.75	Pittsburgh	0.32	Pittsburgh	0.41	Minneapolis	0.17

Notes: Data comes from 1960 tract-level aggregates retrieved from IPUMS NHGIS. Calculations are limited to tracts with HOLC grades. This encompasses most of the population as 85.3% of the less than high school, 83.1% of the high school graduate or more, 87.2% of the Black, and 83.3% of the White population lived in a tract with a HOLC grade.

Table A.3: Summary Statistics for Redlined vs. Non-Redlined Tracts in 1960

	(1)	(2)	(3)	(4)
	Redlined	Non-Redlined		
Variables	Mean (SD)	Mean (SD)	Difference (SE)	T-Stat
Demographic Variables				
Pct White	72.46	93.91	-21.45***	(-29.29)
	(34.40)	(14.92)	(0.73)	
Pct HS Grad	28.82	47.57	-18.75***	(-64.21)
	(12.76)	(16.00)	(0.29)	
Pct Bottom Quintile	27.47	16.97	10.49***	(34.01)
	(14.25)	(10.08)	(0.31)	
Pct Top Quintile	13.07	22.84	-9.77***	(-46.25)
	(8.92)	(13.56)	(0.21)	
Population	4463.4	2570.3	1893.1***	(30.67)
	(2838.9)	(2166.6)	(61.8)	
Economic Variables				
Rent (2010$)	449.8	556.5	-106.7***	(-37.52)
	(121.2)	(175.4)	(2.8)	
Home Value (2010$)	96236.6	117697.9	-21461.3***	(-23.67)
	(40617.5)	(42236.3)	(907.2)	
Dist CBD (Miles)	6.89	15.87	-8.98***	(-59.35)
	(6.11)	(11.11)	(0.15)	
Commuting Variables				
Pct Auto	0.41	0.69	-0.28***	(-52.06)
	(0.24)	(0.20)	(0.01)	
Pct Pub Trans	0.37	0.16	0.21***	(47.51)
	(0.21)	(0.17)	(0.00)	
Observations	2256	19445	21701	

Notes: Unit of observation is census tract. Data comes from 1960 tract-level aggregates retrieved from IPUMS NHGIS. Rents are monthly and CPI-adjusted to 2010 dollars from 1960 dollars. Home values are CPI-adjusted to 2010 dollars from 1960 dollars. Percentage automobile is the percentage of employed workers whose main mode of transport is private automobile which includes truck and van drivers. Percentage public transport is the percentage of employed workers whose main mode of transport is railroad, subway, elevated, bus, streetcar, or other public means. Sample standard deviations are included in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.4: Pre-Trends for Placement of Highway Routes (1940-1950, 1950-1960)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables	Log Distance from Highway Routes					Plans	Rays
Δ Pct White	0.931***				1.055***	-0.0153	0.0139
	(0.211)				(0.216)	(0.254)	(0.267)
Δ Pct HS Grad		-0.260			-0.585***	-0.366	-0.413*
		(0.179)			(0.174)	(0.229)	(0.224)
Δ Log Rent			-0.0446***		-0.0221**	-0.0275	-0.0232
			(0.0132)		(0.0110)	(0.0168)	(0.0155)
Δ Log Home Value				-0.0285***	-0.0194**	0.00178	-0.0140
				(0.0110)	(0.00977)	(0.0118)	(0.0136)
Log Dist CBD	0.178***	0.205***	0.195***	0.199***	0.238***	0.314***	0.439***
	(0.0242)	(0.0245)	(0.0241)	(0.0242)	(0.0344)	(0.0386)	(0.0407)
R-squared	0.032	0.026	0.028	0.027	0.163	0.140	0.151
CBSA FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,166	7,166	7,166	7,166	7,166	7,166	7,166

Notes: Unit of observation is census tract. Data comes from 1940, 1950 and 1960 tract-level aggregates retrieved from IPUMS NHGIS. The first difference is either over 1940 to 1950 or 1950 to 1960 depending on when highway construction started in the CBSA. Tracts are limited to those within 5 miles of the nearest constructed route. Fixed effects are at the CBSA (Core-based statistical area) level for all specifications. Conley standard errors for spatial correlation within a 1km radius are reported. Median income is missing in 1960 and so is not included in the pre-trends table. Change in percent bottom and top quintile are not shown as they are available only starting in 1950, and very few cities began construction on the Interstate highway system post-1960. All specifications have as controls log distance from the central business district, rivers, lakes, shores, ports, historical railroads, canals, and historical large urban roads. *** p<0.01, ** p<0.05, * p<0.1

Table A.5: First-Stage for Highway Placement

Variables	(1) Log Dist HW	(2) Log Dist HW	(3) Dist HW = 1 mi	(4) Dist HW = 1 mi
Log Dist Plans	0.325***			
	(0.0175)			
Log Dist Rays		0.246***		
		(0.0196)		
Dist Plans = 1 mi			0.426***	
			(0.0223)	
Dist Rays = 1 mi				0.312***
				(0.0292)
Log Dist CBD	0.0863***	0.0752***	-0.0379***	-0.0422***
	(0.0212)	(0.0241)	(0.00957)	(0.0101)
F-Stat	342.8	156.5	366.3	113.9
R-squared	0.291	0.224	0.251	0.174
CBSA FE	Yes	Yes	Yes	Yes
Geo Controls	Yes	Yes	Yes	Yes
Observations	31,627	31,627	31,627	31,627
No. Counties	467	467	467	467

Notes: Unit of observation is census tract. Tracts are limited to those within 5 miles of the nearest constructed route. Fixed effects are at the CBSA (Core-based statistical area) level. Standard errors are cluster-robust with clusters at the county level. All specifications have as controls log distance from the central business district, rivers, lakes, shores, ports, historical railroads, canals, and historical large urban roads. The reported F-stat comes from testing a single coefficient on the excluded instrument. *** p<0.01, ** p<0.05, * p<0.1

Table A.6: Change in Home Values over Distance from Central Business District and Distance from Highway (1950-1960, 1960-1970)

	(1)	(2)	(3)
	OLS	IV for Log Dist Highway	
Variables	Δ Log Home Value	Plans	Rays
Log Dist Highway	-0.0159	-0.0994***	-0.0469
	(0.0179)	(0.0373)	(0.0531)
Log Dist CBD	-0.168***	-0.148***	-0.161***
	(0.0389)	(0.0388)	(0.0410)
Redlined	0.284***	0.265***	0.277***
	(0.0648)	(0.0648)	(0.0667)
Dep. Var Mean (1960)	120,572 (2010$)		
R-squared	0.121	0.118	0.121
CBSA FE	Yes	Yes	Yes
Geo Controls	Yes	Yes	Yes
Observations	10,395	10,395	10,395
KP F-Stat		678.5	469.7

Notes: Unit of observation is census tract. Data comes from 1950, 1960, and 1970 tract-level aggregates retrieved from IPUMS NHGIS. The first difference is either over 1950 to 1960 or 1960 to 1970 depending on when highway construction started in the CBSA and stacked into one panel. Tracts are limited to those within 5 miles of the nearest constructed route and 30 miles from the central business district. Fixed effects are at the CBSA (Core-based statistical area) level. Conley standard errors for spatial correlation within a 1km radius are reported. All specifications have as controls log distance from rivers, lakes, shores, ports, historical railroads, canals, and historical large urban roads, and the gradient (Dist CBD/Dist Highway). Redlined tracts are those where more than 80% of the area is redlined. Kleibergen-Paap rk Wald F statistics are reported for the first-stage. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.7: First-Stage for Commuter Market Access Improvements

Variables	(1) Δ Log CMA	(2) Δ Log CMA	(3) Δ Log CMA	(4) Δ Log CMA HW	(5) Δ Log CMA HW
Δ Log CMA HW	0.639***				
	(0.0127)				
Δ Log CMA Plans		0.107***		0.382***	
		(0.0120)		(0.0088)	
Δ Log CMA Rays			0.0888***		0.336***
			(0.0108)		(0.0079)
F-Stat	2534	79.68	67.14	1884	1797
R-squared	0.313	0.262	0.262	0.505	0.484
CBSA FE	Yes	Yes	Yes	Yes	Yes
Geo Controls	Yes	Yes	Yes	Yes	Yes
Rounded Obs	60500	60500	60500	60500	60500

Notes: Unit of observation is census tract by race and education. Data comes from the first difference of 1960 to 1970 using restricted Census microdata. Fixed effects are at the CBSA (Core-based statistical area) level. Standard errors are cluster-robust with clusters at the tract-level. All specifications have as controls log distance from the central business district, rivers, lakes, shores, ports, historical railroads, canals, and historical large urban roads. All specifications include the Borusyak and Hull (2023) control for CMA in large roads and CBSA fixed effects. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. The reported F-stat comes from testing a single coefficient on the excluded instrument. *** p<0.01, ** p<0.05, * p<0.1

Table A.8: Elasticity of Population to Commuter Market Access – Additional Results

	(1)	(2)	(3)	(4)
	$\Delta \log L_{igr}$ (Δ Log Population 1960-1970)			
Variables	+BH (2023) Plans	+BH (2023) Rays	Unscaled CMA	Race x Educ
$\Delta \log CMA_{igr}$				
Black	0.104	0.107		
	(0.0973)	(0.0970)		
White	1.469***	1.458***		
	(0.121)	(0.121)		
$\Delta \log \Phi_{igr}$				
Black			0.0314	
			(0.0323)	
White			0.470***	
			(0.0385)	
$\Delta \log CMA_{igr}$				
Black <HS				0.218*
				(0.124)
Black HS Grad				-0.712***
				(0.155)
White <HS				0.958***
				(0.127)
White HS Grad				0.946***
				(0.141)
R-squared	0.113	0.113	0.113	0.139
CBSA FE	Yes	Yes	Yes	Yes
Geo Controls	Yes	Yes	Yes	Yes
Rounded Obs	60500	60500	60500	60500

Notes: Unit of observation is census tract by race and education. Data comes from the first difference of 1960 to 1970 using restricted Census microdata. Fixed effects are at the CBSA (Core-based statistical area) level. Standard errors are cluster-robust with clusters at the tract-level. The geographic controls are log distance from the central business district, rivers, lakes, shores, ports, historical railroads, canals, and historical large urban roads, all interacted with race for Columns 1–3 and with race and education for Column 4. Column 1 and Column 2 include the Borusyak and Hull (2023) control for CMA interacted with race when the planned network and the Euclidean ray network are built, respectively. Columns 3–4 include the Borusyak and Hull (2023) control for CMA in large roads where in Column 3 it is interacted with race and in Column 4 it is interacted with race and education. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. *** p<0.01, ** p<0.05, * p<0.1

Table A.9: Elasticity of Rents, Racial Composition, and Population to Commuter Market Access

Variables	(1) Δ Log Rent	(2) Δ Log Pop + Δ Log Rent Cont	(3) Δ Log Pct White	(4) Δ Log Pop + Δ Log Pct White Cont
$\Delta \log CMA_{igr}$	0.0432***		-0.0180	
	(0.00720)		(0.0139)	
$\Delta \log CMA_{igr}$				
Black		0.137		0.141
		(0.0974)		(0.0959)
White		1.267***		1.403***
		(0.118)		(0.114)
R-squared	0.225	0.121	0.071	0.146
CBSA FE	Yes	Yes	Yes	Yes
Geo Controls	Yes	Yes	Yes	Yes
Rounded Obs	59000	59000	60000	60000

Notes: Unit of observation is census tract by race and education. Data comes from the first difference of 1960 to 1970 using restricted Census microdata. Fixed effects are at the CBSA (Core-based statistical area) level. Standard errors are cluster-robust with clusters at the tract-level. The geographic controls are log distance from the central business district, rivers, lakes, shores, ports, historical railroads, canals, and historical large urban roads, all interacted with race in Columns 2 and 4 and with race and education in Columns 1 and 3. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. All specifications include the Borusyak and Hull (2023) control for CMA in large roads. Kleibergen-Paap rk Wald and Cragg-Donald Wald F statistics for weak instruments are reported. *** p<0.01, ** p<0.05, * p<0.1

Table A.10: Elasticity of Population to Commuter Market Access for Instruments

	(1)	(2)
	$\Delta \log L_{igr}$ (Δ Log Pop)	
Variables	Plans	Rays
$\Delta \log CMA_{igr}$		
Black	0.665*	0.624*
	(0.365)	(0.333)
	[0.587]	[0.550]
White	0.430**	0.745***
	(0.174)	(0.172)
	[0.578]	[0.628]
R-squared	0.517	0.513
CBSA X Group FE	Yes	Yes
Base Controls	Yes	Yes
Geo Controls	Yes	Yes
Rounded Obs	58000	58000
C-D F-Stat	313.7	259.5
K-P F-stat	26.20	25.97

Notes: Unit of observation is census tract by race and education. Data comes from the first difference of 1960 to 1970 using restricted Census microdata. Fixed effects are at the CBSA (Core-based statistical area) by race and education level. Standard errors are cluster-robust with clusters at the tract-level. Conley standard errors for spatial correlation within a 1km radius are reported in brackets. The base controls are change in log rent, change in log pct White, and 5 binary indicators for distance from highways built between 1960 and 1970 in 1-mile wide bins, all interacted with race and education. Redlining fixed effects are included in all specifications. The geographic controls are log distance from the central business district, rivers, lakes, shores, ports, historical railroads, canals, and historical large urban roads, all interacted with race and education. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. All specifications include the Borusyak and Hull (2023) control for CMA in large roads. Kleibergen-Paap rk Wald and Cragg-Donald Wald F statistics for weak instruments are reported. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.11: Commuting Gravity Equation – Additional Results

	(1)	(2)	(3)	(4)
Race x Educ	Black <HS	Black HS Grad	White <HS	White HS Grad
	Panel A – First-Stage – IV Plans			
Log Commute Time	0.988***	1.026***	1.023***	1.025***
	(0.00573)	(0.00475)	(0.00178)	(0.00145)
F-Stat (Rounded)	29710	46750	331400	501900
	Panel B – First-Stage – IV Rays			
Log Commute Time	0.999***	1.037***	1.027***	1.029***
	(0.00595)	(0.00492)	(0.00183)	(0.00152)
F-Stat (Rounded)	28150	44440	315400	455600
	Panel C – Log Commuting Share – IV Distance			
Log Commute Time	-5.198***	-4.325***	-5.205***	-4.485***
	(0.140)	(0.118)	(0.0502)	(0.0415)
R-squared	0.223	0.181	0.379	0.369
	Panel D – First-Stage – IV Distance			
Log Commute Time	0.492***	0.556***	0.662***	0.713***
	(0.00707)	(0.00609)	(0.00396)	(0.00404)
F-Stat (Rounded)	4840	8350	28020	31100
	Panel E – Log Commuting Share in Log Distance			
Log Distance	-2.556***	-2.406***	-3.447***	-3.199***
	(0.0642)	(0.0640)	(0.0336)	(0.0306)
R-squared	0.697	0.627	0.578	0.577
Rounded Obs	7000	8000	21500	25000
POR X Year FE	Yes	Yes	Yes	Yes
POW X Year FE	Yes	Yes	Yes	Yes

Notes: Unit of observation is Place of Work Zone by Place of Work Zone pair by year where commuting flows from residential tracts are aggregated up to the Place of Work Zone geography. Fixed effects are for POR (Place of Residence) by year at the Place of Work Zone scale although it does not represent workplace but rather residential location. POW by year fixed effects are for workplace at the Place of Work Zone level. The conditional commuting share is the share from a residential location that commutes to a workplace. Data comes from the restricted Census microdata in 1960 and 1970. The observation counts are lower for the Black population as some residences and workplaces have zero Black population. Standard errors are cluster-robust with clusters at the Place of Work Zone by Place of Work Zone level. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. The F-stat comes from testing a single coefficient on the excluded instrument and is rounded to four significant digits to meet Census disclosure rules. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.12: Commuting Gravity Equation – Poisson Pseudo Maximum Likelihood IV Results

	(1)	(2)	(3)	(4)
Race x Educ	Black <HS	Black HS Grad	White <HS	White HS Grad
$\nu_{gr} = \kappa_{gr}\phi$	*Panel A* – Commuting Share (PPML) – IV Plans			
Log Commute Time	-4.706*** (0.138)	-3.940*** (0.0857)	-3.888*** (0.0655)	-3.260*** (0.0526)
	Panel B – Commuting Share (PPML) – IV Rays			
Log Commute Time	-4.707*** (0.140)	-3.941*** (0.0879)	-3.883*** (0.0655)	-3.256*** (0.0522)
Rounded Obs	20500	21000	26000	27000
POR X Year FE	Yes	Yes	Yes	Yes
POW X Year FE	Yes	Yes	Yes	Yes

Notes: Unit of observation is Place of Work Zone by Place of Work Zone pair by year where commuting flows from residential tracts are aggregated up to the Place of Work Zone geography. Fixed effects are for POR (Place of Residence) by year at the Place of Work Zone scale although it does not represent workplace but rather residential location. POW by year fixed effects are for workplace at the Place of Work Zone level. The conditional commuting share is the share from a residential location that commutes to a workplace. Data comes from the restricted Census microdata in 1960 and 1970. The observation counts are lower for the Black population as some residences and workplaces have zero Black population (while PPML addresses zeros in bilateral flows, it does not address zeros in entire rows or columns). As the coefficient is estimated via the control function approach of Wooldridge (2015), to obtain the correct standard errors, I bootstrap 200 samples to calculate standard errors with clusters at the Place of Work Zone by Place of Work Zone level. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. *** p<0.01, ** p<0.05, * p<0.1

Table A.13: Change in Amenities over Distance from Highway – IV Results

	(1)	(2)
	$\Delta \log b_{igr}$	
	IV – 2 mi bins	
Variables	Dist Plans	Dist Rays
Dist Highway (mi = 2)	-0.0469	0.519
	(0.133)	(0.668)
Dist Highway (mi = 4)	0.249*	0.312
	(0.149)	(0.502)
Dist Highway (mi = 6)	0.0766	-0.149
	(0.279)	(1.385)
R-squared	0.047	0.036
CBSA X Group FE	Yes	Yes
Geo Controls	Yes	Yes
Rounded Obs	51000	47000
C-D F-Stat	686.2	18.38
K-P F-stat	110.8	4.55

Notes: Unit of observation is census tract by race and education. Fixed effects are at the CBSA (Core-based statistical area) by race and education group level. Data comes from the first difference of 1960 to 1970 using restricted Census microdata. Standard errors are cluster-robust with clusters at the tract-level. There are 3 binary indicators for distance from highways built between 1960 and 1970 in 2-mile wide bins to increase power. The geographic controls are log distance from the central business district, rivers, lakes, shores, ports, historical railroads, canals, and historical large urban roads, all interacted with race and education. Redlining fixed effects are included. The sample is limited to tracts within 10 miles of planned routes or the Euclidean ray network. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. Kleibergen-Paap rk Wald and Cragg-Donald Wald F statistics for weak instruments are reported. Parameters used to invert for dependent variables are the same as in Table 10. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.14: Environmental Pollution Index (PM 2.5) over Distance from Highway

	(1)	(2)	(3)
	Log Particulate Matter 2.5		
Variables	1 mi bins	+ Geo Cont	0.5 mi bins
Dist Highway (mi = 1)	0.0245***	0.0204***	
	(0.000983)	(0.000976)	
Dist Highway (mi = 2)	0.0231***	0.0196***	
	(0.000980)	(0.000965)	
Dist Highway (mi = 3)	0.0197***	0.0174***	
	(0.00102)	(0.00100)	
Dist Highway (mi = 4)	0.0146***	0.0133***	
	(0.00114)	(0.00111)	
Dist Highway (mi = 5)	0.0108***	0.0103***	
	(0.00129)	(0.00126)	
Dist Highway (mi = 0.5)			0.0201***
			(0.00107)
Dist Highway (mi = 1)			0.0207***
			(0.00101)
Dist Highway (mi = 1.5)			0.0200***
			(0.00103)
Dist Highway (mi = 2)			0.0191***
			(0.00106)
Dist Highway (mi = 2.5)			0.0176***
			(0.00110)
Dist Highway (mi = 3)			0.0172***
			(0.00117)
Dist Highway (mi = 3.5)			0.0145***
			(0.00130)
Dist Highway (mi = 4)			0.0119***
			(0.00136)
Dist Highway (mi = 4.5)			0.0118***
			(0.00154)
Dist Highway (mi = 5)			0.00850***
			(0.00167)
Dep Var Mean	13.50		
R-squared	0.962	0.964	0.964
CBSA FE	Yes	Yes	Yes
Geo Controls		Yes	Yes
Observations	32,833	32,833	32,833

Notes: Unit of observation is census tract. Fixed effects are at the CBSA (Core-based statistical area) level. Data comes from the CDC Environmental Health Census Tract-Level PM2.5 Concentrations, 2001-2005 measures. There are 5 binary indicators for distance from highways in 1-mile wide bins in Columns 1–2 and 10 binary indicators in 0.5-mile wide bins in Column 3 (the value displayed is the upper end of the bin). Included in all specifications are redlining fixed effects and log distance from the central business district. The geographic controls are log distance from rivers, lakes, shores, ports, historical railroads, canals, and historical large urban roads. The sample is limited to tracts within 10 miles of highway routes. Standard errors are heteroskedasticity robust. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.15: Housing Expenditure Function

Variables	(1) Housing Expenditures
Predicted Income	0.119***
	(0.00294)
Constant	353.3***
	(9.263)
R-squared	0.080
Observations	20,786

Notes: Unit of observation is individual. Data comes from the Consumer Expenditure Surveys Public-Use Microdata in the year 1980. Predicted income is a linear prediction of income using categorical variables in age, education, marital status, occupation, sex, race and region. Income and housing expenditure is for quarterly amounts. Standard errors are heteroskedasticity robust. *** p<0.01, ** p<0.05, * p<0.1

Table A.16: Predicted Share of Income Spent on Housing

Race x Educ	Black <HS	Black HS Grad	White <HS	White HS Grad
Housing Exp. Share	0.34	0.22	0.30	0.21
Observations	1441	1908	5885	14712

Notes: Unit of observation is individual. Data comes from the Consumer Expenditure Surveys Public-Use Microdata in the year 1980. Income and housing expenditure is in quarterly amounts. Predicted share of income spent on housing uses the linear housing expenditure function from A.15 and the average level of income of the four race by education groups.

Table A.17: Predicted and Observed Commute Flows in 1960 and 1970

	(1)	(2)	(3)	(4)
	Unweighted		Weighted	
Variables	Log Obs 1960	Log Obs 1970	Log Obs 1960	Log Obs 1970
Log Predicted 1960	0.855***		0.838***	
	(0.0260)		(0.0197)	
Log Predicted 1970		0.866***		0.950***
		(0.0178)		(0.0134)
Constant	0.456***	0.521***	2.018***	0.949***
	(0.136)	(0.102)	(0.111)	(0.186)
R-squared	0.560	0.613	0.861	0.950
Rounded Obs	12000	14000	12000	14000

Notes: Unit of observation is Place of Work Zone level by Place of Work Zone level. Data comes from restricted Census microdata in 1960 and 1970. Standard errors are cluster-robust with clusters at the CBSA (Core-based statistical area) level. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.18: Change in Log Productivity Over Distance from Highway (1960-1970)

	(1)
Variables	Δ Log Productivity
Dist from Highway (Miles)	-0.00269
	(0.00177)
Constant	-1.037***
	(0.0121)
R-squared	0.007
Rounded Obs	16000

Notes: Unit of observation is tract. Data comes from restricted Census microdata in 1960 and 1970. Standard errors are cluster-robust with clusters at the Place of Work Zone level because the variation in wages used to invert for productivity are at the Place of Work Zone level while housing prices used for inversion are at the tract level. Distance from the highway is in miles from segments of the highway network constructed between 1960 and 1960. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.19: Log Productivity Over Distance from Central Business District and Pct HOLC D in 1960 and 1970

	(1)	(2)
Variables	Log Prod 1960	Log Prod 1970
Pct HOLC D	0.0954	0.0358
	(0.0761)	(0.0719)
Dist from CBD (Miles)	0.00212	-0.00106
	(0.00171)	(0.00137)
Constant	5.797***	4.829***
	(0.0397)	(0.0332)
R-squared	0.020	0.011
Rounded Obs	17000	14000

Notes: Unit of observation is tract. Data comes from restricted Census microdata in 1960 and 1970. Standard errors are cluster-robust with clusters at the Place of Work Zone level because the variation in wages used to invert for productivity are at the Place of Work Zone level while housing prices used for inversion are at the tract level. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. *** p<0.01, ** p<0.05, * p<0.1

Table A.20: Changes in Equilibrium Outcomes (%) for Highway Impacts by Race and Education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Race	Black	Black	White	White	By Race		By Educ	
x Educ	<HS	HS+	<HS	HS+	Black	White	<HS	HS+
General Equilibrium								
Rent	0.32	0.47	0.23	0.14	0.37	0.18	0.24	0.16
Pct White	-0.47	-0.16	0.03	0.01	-0.37	0.02	-0.04	-0.00
Pct HOLC D	-0.19	-0.26	-1.06	-0.95	-0.21	-1.00	-0.93	-0.90
Amenities	-10.35	-9.43	-5.91	-4.82	-10.06	-5.33	-6.57	-5.13
Wages	-0.11	0.13	0.20	0.28	-0.03	0.24	0.15	0.27
Localized Costs	-7.99	-7.96	-6.37	-5.86	-7.98	-6.10	-6.61	-6.00
Dist from CBD	0.59	0.27	1.49	1.14	0.49	1.30	1.36	1.08
Commute Time	-3.72	-5.31	-3.47	-4.21	-4.23	-3.86	-3.51	-4.28
Commute Dist	4.81	3.95	6.16	4.91	4.53	5.50	5.96	4.84
General Equilibrium, No BD								
Rent	0.31	0.45	0.23	0.14	0.35	0.18	0.24	0.16
Pct White	-0.34	-0.15	0.03	0.01	-0.28	0.02	-0.03	-0.00
Pct HOLC D	-0.25	-0.33	-0.91	-0.86	-0.28	-0.88	-0.81	-0.82
Amenities	-8.67	-8.49	-5.99	-4.82	-8.61	-5.37	-6.39	-5.07
Wages	-0.13	0.12	0.20	0.27	-0.05	0.24	0.15	0.26
Localized Costs	-7.43	-7.55	-6.41	-5.88	-7.47	-6.13	-6.56	-5.99
Dist from CBD	0.61	0.27	1.50	1.15	0.50	1.31	1.37	1.09
Commute Time	-3.59	-5.15	-3.48	-4.22	-4.09	-3.87	-3.50	-4.28
Commute Dist	4.82	3.92	6.15	4.91	4.53	5.49	5.95	4.84
General Equilibrium, Same Fund Amen								
Rent	0.30	0.27	0.24	0.15	0.29	0.19	0.25	0.16
Pct White	0.09	0.08	-0.02	0.00	0.09	-0.01	-0.00	0.01
Pct HOLC D	-0.32	-0.29	-0.77	-0.77	-0.31	-0.77	-0.70	-0.74
Amenities	-6.00	-5.15	-6.17	-4.90	-5.73	-5.50	-6.14	-4.92
Wages	0.16	0.25	0.19	0.26	0.19	0.23	0.19	0.26
Localized Costs	-5.81	-5.80	-6.53	-5.99	-5.81	-6.24	-6.42	-5.98
Dist from CBD	0.67	0.47	1.52	1.18	0.61	1.34	1.39	1.13
Commute Time	-3.55	-4.16	-3.48	-4.21	-3.75	-3.87	-3.49	-4.21
Commute Dist	4.94	4.50	6.15	4.97	4.80	5.52	5.97	4.94

Notes: Equilibrium outcome calculations are based on data from the restricted Census in 1960. The general equilibrium simulation allows wages to respond in equilibrium. No institutions adjusts fundamental amenities for Black households by parameter E in redlined areas. The general equilibrium simulation with no institutions adds the highway impacts in the counterfactual world with no institutions. Parameter values are the same as in Table 9. All values are rounded to four significant digits to meet Census disclosure rules.

Table A.21: Alternative Exercises for Welfare Changes (%) by Race and Education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Race	Black	Black	White	White	By Race		By Educ	
x Educ	<HS	HS+	<HS	HS+	Black	White	<HS	HS+
General Equilibrium								
Baseline	-1.45	-0.16	2.69	3.01	-1.04	2.86	2.07	2.79
Full Interstate Network								
Welfare Change	15.36	23.45	25.53	27.59	17.95	26.62	24.01	27.31
Alternative Road Placements								
Planned Routes	14.83	23.15	25.23	27.67	17.49	26.52	23.68	27.36
Euclidean Rays	21.32	30.14	32.31	34.74	24.14	33.60	30.67	34.43
Alternative Spillovers								
$\rho'_W = 0$	-1.44	-0.19	2.79	3.03	-1.04	2.92	2.16	2.81
$\rho'_B = -0.2$	-0.90	-0.01	2.70	3.03	-0.62	2.87	2.16	2.82
Alternative Elasticities								
$\theta'_B = \theta_W = 0.8$	-1.47	-0.19	2.52	2.86	-1.06	2.70	1.93	2.65
$\theta'_r = 3\theta_r$	-1.26	0.02	3.28	3.52	-0.85	3.41	2.60	3.28
$\theta'_B = \theta'_W = 3\theta_W$	1.02	1.01	2.67	2.94	1.02	2.81	2.42	2.81

Notes: Welfare calculations are based on data from the restricted Census in 1960. All welfare changes are for the general equilibrium simulation of highway impacts but with different parameter values. $\rho'_N = -0.2$ comes from the estimate for $\tilde{\rho}_N = -0.07$ in Table 6 divided by the residential elasticity of $\theta_N = 0.35$. The alternative elasticities set the residential elasticity for Black and White households to the same values at the level of White households, to three times their original values, to three times the original level of White households. All values are rounded to four significant digits to meet Census disclosure rules.

Table A.22: Border Discontinuity on Additional Variables

	(1)	(2)	(3)	(4)	(5)
			Panel A – Pct White in 1960		
Variables	Standard	Balanced Sample	Border FE	Drop Roads, Schools (0.1 mi)	Drop Roads, Schools (0.3 mi)
Border RD	-0.189***	-0.175***	-0.180***	-0.171***	-0.166***
	(0.027)	(0.029)	(0.025)	(0.026)	(0.034)
Bandwidth (mi)	0.351	0.368	0.355	0.403	0.432
Border FE	No	No	Yes	Yes	Yes
Observations	12573	5914	12532	10703	5717
			Panel B – Socioeconomic Variables in 1960		
Variables	Pct HS	Pct Bottom Q5	Pct Top Q5	Home Value	Rent
Border RD	-0.064***	0.104***	-0.093***	-22248***	-110.06***
	(0.006)	(0.010)	(0.008)	(3309)	(12.05)
Dep. Var Mean	0.265	0.189	0.207	114238 (2010$)	534 (2010$)
Bandwidth (mi)	0.267	0.397	0.409	0.428	0.248
Border FE	Yes	Yes	Yes	Yes	Yes
Observations	12275	12310	12310	12260	12268
			Panel C – Pct White Over Time		
Variables	1950	1960	1970	1980	1990
Border RD	-0.141***	-0.187***	-0.185***	-0.151***	-0.146***
	(0.023)	(0.029)	(0.035)	(0.037)	(0.035)
Dep. Var Mean	0.945	0.911	0.852	0.773	0.668
Bandwidth (mi)	0.350	0.350	0.350	0.350	0.350
Border FE	Yes	Yes	Yes	Yes	Yes
Observations	9964	9964	9964	9964	9964
			Panel D – Land Types and Tree Cover		
Variables	Pct Open Water	Pct Woody Wetlands	Pct Decid Forest	Pct Highly Developed	Pct Tree Cover
Border RD	0.005	-0.004	-0.015**	0.033***	-0.056***
	(0.003)	(0.003)	(0.005)	(0.007)	(0.009)
Dep. Var Mean	0.014	0.021	0.050	0.063	0.197
Bandwidth (mi)	0.406	0.315	0.351	0.347	0.461
Border FE	Yes	Yes	Yes	Yes	Yes
Observations	11529	11529	11529	11529	11529

Notes: Unit of observation is census tract by redlining border. Data comes from 1950, 1960, 1970, 1980, and 1990 tract-level aggregates retrieved from IPUMS NHGIS. The dependent variable is residualized on border fixed effects for many specifications. The balanced sample has the same number of tracts on the redlined and non-redlined sides. The "Drop Roads, Schools" sample is limited to tracts that are at least 0.1 (or 0.3) miles away from possible physical barriers such as historical large urban roads, constructed highways in 1960, or historical railroads and 0.1 (or 0.3) miles away from a school district boundary. The bandwidth is chosen optimally following Calonico et al. (2014) except for in Panel C, where the bandwidth is set to 0.35 so the effective sample remains the same across decades. The order of polynomial is 1 for all specifications. Distance from the border is measured in miles. *** p<0.01, ** p<0.05, * p<0.1

Table A.23: Border Discontinuity Decomposition – Reduced Form Approach

	(1)	(2)	(3)	(4)
	Panel A – Black			
Variables	log L_{igr}	Controls 1	Controls 2	Controls 3
ψ_B: Border RD	1.425***	1.414***	0.555***	0.489**
	(0.226)	(0.227)	(0.212)	(0.204)
Bandwidth (mi)	0.495	0.488	0.414	0.365
Order of Poly.	1	1	1	1
Education FE	Yes	Yes	Yes	Yes
Border FE	Yes	Yes	Yes	Yes
Rounded Obs	13000	13000	13000	13000
	Panel B – White			
Variables	log L_{igr}	Controls 1	Controls 2	Controls 3
ψ_W: Border RD	-0.546***	-0.556***	-0.101	0.0205
	(0.122)	(0.122)	(0.0855)	(0.0794)
Bandwidth (mi)	0.358	0.364	0.380	0.397
Order of Poly.	1	1	1	1
Education FE	Yes	Yes	Yes	Yes
Border FE	Yes	Yes	Yes	Yes
Rounded Obs	13500	13500	13500	13500

Notes: Unit of observation is census tract by border in the redlining maps. Data comes from the 1960 restricted Census microdata. The dependent variable is residualized on fixed effects for education within race and on border fixed effects for all specifications. Controls 1 are log rent and log commuter access. Controls 2 includes Controls 1 and adds log percentage white. Controls 3 includes Controls 2 and adds log percentage high school grad, log population density, log average income, log percentage top quintile, log percentage bottom quintile, and log home values. Coefficients on controls are estimated with redlining fixed effects. Sample is limited to tracts that are at least 0.1 miles away from possible physical barriers such as historical large urban roads, constructed highways in 1960, or historical railroads and also at least 0.1 miles away from a school district boundary. The bandwidth is chosen optimally following Calonico et al. (2014). Distance from the border is measured in miles. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. *** p<0.01, ** p<0.05, * p<0.1

Table A.24: Changes in Welfare and Equilibrium Outcomes (%) for Removal of Institutional Barriers by Race and Education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Race	Black	Black	White	White	By Race		By Educ	
x Educ	<HS	HS+	<HS	HS+	Black	White	<HS	HS+
No BD								
Welfare Change	—	—	0.20	-0.50	—	-0.17	—	—
Rent	3.23	3.25	-0.17	-0.06	3.24	-0.11	0.34	0.17
Pct White	10.60	3.86	-0.29	-0.46	8.44	-0.38	1.33	-0.17
Pct HOLC D	-25.80	-31.02	6.06	5.51	-27.47	5.77	1.31	3.02
Dist from CBD	11.49	7.27	-0.57	-0.20	10.14	-0.37	1.23	0.31
Amenities	—	—	-0.19	-0.80	—	-0.51	—	—
Same Fund Amen								
Welfare Change	—	—	-0.03	-3.27	—	-1.75	—	—
Rent	9.35	11.62	-0.56	-0.33	10.08	-0.44	0.92	0.48
Pct White	136.3	124.0	-5.04	-6.12	132.36	-5.61	16.02	2.75
Pct HOLC D	-52.74	-46.97	12.67	14.01	-50.89	13.38	2.93	9.85
Dist from CBD	93.37	94.16	-4.48	-3.25	93.62	-3.83	10.10	3.39
Amenities	—	—	-1.87	-5.64	—	-3.87	—	—

Notes: Equilibrium outcome calculations are based on data from the restricted Census in 1960. The general equilibrium simulation allows wages to respond in equilibrium. No institutions adjusts fundamental amenities for Black households by parameter E in redlined areas. The general equilibrium simulation with no institutions adds the highway impacts in the counterfactual world with no institutions. Parameter values are the same as in Table 9. All values are rounded to four significant digits to meet Census disclosure rules.

Table A.25: Match Rates by Birth Year

	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled		White		Black	
Variable	Match Rate	Population	Match Rate	Population	Match Rate	Population
Birth Year						
1964	0.58	4094000	0.62	2827000	0.52	487000
1965	0.59	3831000	0.63	2619000	0.53	467000
1966	0.60	3677000	0.65	2511000	0.53	449000
1967	0.60	3594000	0.65	2448000	0.54	438000
1968	0.61	3582000	0.67	2437000	0.55	430000
1969	0.70	3688000	0.74	2502000	0.65	442000
1970	0.71	3834000	0.76	2580000	0.66	469000
1971	0.72	3670000	0.77	2431000	0.67	459000
1972	0.73	3384000	0.79	2203000	0.66	431000
1973	0.74	3264000	0.81	2104000	0.66	416000
1974	0.76	3294000	0.84	2120000	0.66	410000
1975	0.64	3280000	0.70	2087000	0.58	412000
1976	0.66	3302000	0.72	2092000	0.58	415000
1977	0.67	3451000	0.74	2195000	0.59	441000
1978	0.69	3447000	0.77	2178000	0.58	446000
1979	0.71	3607000	0.79	2267000	0.57	468000
All Years	0.67	57000000	0.72	37600000	0.60	7080000

		White		Black	
Variable	Gender	Match Rate	Population	Match Rate	Population
Birth Year					
All Years	Men	0.73	18980000	0.60	3316000
All Years	Women	0.72	18620000	0.59	3764000

Note: The pooled match rates are for the entire U.S. and includes White individuals, Black individuals, and other racial groups. All racial groups exclude individuals of Hispanic ethnicity. There is a discrete jump in match rates for the birth cohorts of 1969 to 1974. Individuals with birth years between 1964-1974 were matched to the 1974 IRS 1040 form, and individuals with birth years between 1969-1979 were matched to the 1979 IRS 1040 form. Therefore the 1969-1974 cohorts were given two chances to be matched to at least one tax filing. As these children's parents do not consistently file for taxes across years, some appear in the 1974 form and not the 1979 form or vice versa. Counts are rounded in the following manner: numbers between 10,000 and 99,999 are rounded to the nearest 500; between 100,000 and 9,999,999 to the nearest 1,000; and above 10,000,000 to the nearest 10,000 to meet Census disclosure rules.

Table A.26: Representativeness of Unmatched vs. Matched Children

	(1)	(2)	(3)	(4)
	Unmatched	Matched	Difference	
Variable	Mean	Mean	Raw Diff	Race + Year FE
HS Grad Rate	0.901	0.936	0.0358***	0.0266***
SD	(0.299)	(0.244)	(0.000175)	(0.000177)
Rounded N	3165000	7626000	10790000	10360000
College Grad Rate	0.310	0.360	0.0498***	0.0288***
SD	(0.462)	(0.480)	(0.000318)	(0.000328)
Rounded N	3165000	7626000	10790000	10360000
Adjusted Gross Income (2018 $K)	81.65	92.25	10.60***	5.604***
SD	(324.2)	(321.5)	(0.0995)	(0.104)
Rounded N	15200000	34000000	49200000	46730000
Wage & Salary Income (2018 $K)	71.81	81.04	9.230***	5.404***
SD	(132.6)	(148.2)	(0.0448)	(0.0468)
Rounded N	14840000	33240000	48080000	45710000
Individual Earnings (2018 $K)	48.26	54.02	5.754***	3.781***
SD	(125.8)	(302.6)	(0.0819)	(0.0885)
Rounded N	14740000	31910000	46650000	44110000

Note: High school and college graduation rates come from the ACS surveys. Adjusted Gross Income and Wage & Salary income come from the 1040 forms during the years in which the child is aged 35-39. Individual earnings come from W-2 forms during the years in which the child is aged 35-39, except for the birth cohorts of 1964-1969. Their earnings are measured during ages 41-45 as the W-2 data begins in 2005. Race and birth year fixed effects are included in Column (4) for the calculation of the difference between matched and unmatched children. Counts are rounded in the following manner: numbers between 10,000 and 99,999 are rounded to the nearest 500; between 100,000 and 9,999,999 to the nearest 1,000; and above 10,000,000 to the nearest 10,000 to meet Census disclosure rules. Robust standard errors are included in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.27: Intergenerational Elasticities

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Log Child Household Income						
	By Race			By Race and Gender			
Variables	Pooled	White	Black	White Men	White Women	Black Men	Black Women
Log Par HH Inc	0.435***	0.400***	0.244***	0.390***	0.411***	0.240***	0.245***
	(0.000265)	(0.000323)	(0.000689)	(0.000457)	(0.000456)	(0.00111)	(0.000875)
Constant	6.190***	6.661***	7.838***	6.770***	6.553***	7.922***	7.800***
	(0.00292)	(0.00359)	(0.00731)	(0.00508)	(0.00508)	(0.0118)	(0.00925)
R-squared	0.092	0.074	0.036	0.072	0.076	0.031	0.039
Rounded Obs	3.400e+07	2.570e+07	3.660e+06	1.280e+07	1.290e+07	1.590e+06	2.070e+06

Note: Parental household income comes from the 1040 forms and is the average of the first four years of tax data available post-birth of the child. Household income of the child comes from the 1040 forms and is the average of the five years during which the child is aged 35-39. The pooled category includes all racial groups, not solely White and Black individuals. All racial groups exclude individuals of Hispanic ethnicity. Counts are rounded in the following manner: numbers between 10,000 and 99,999 are rounded to the nearest 500; between 100,000 and 9,999,999 to the nearest 1,000; and above 10,000,000 to the nearest 10,000 to meet Census disclosure rules. Robust standard errors are included in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.28: Rank-Rank Correlations by Race and Gender

	(1)	(2)	(3)	(4)
Panel A	Child Household Income Rank			
Race x Gender	White Men	White Women	Black Men	Black Women
Parent HH Income Rank	0.288***	0.302***	0.207***	0.222***
	(0.000269)	(0.000271)	(0.000765)	(0.000655)
Constant	39.94***	39.61***	28.85***	25.33***
	(0.0168)	(0.0172)	(0.0316)	(0.0256)
R-squared	0.084	0.089	0.048	0.059
Rounded Obs	1.280e+07	1.290e+07	1.600e+06	2.070e+06
Panel B	Child Individual Income Rank			
Race x Gender	White Men	White Women	Black Men	Black Women
Parent HH Income Rank	0.260***	0.210***	0.193***	0.198***
	(0.000280)	(0.000294)	(0.000799)	(0.000686)
Constant	47.20***	34.70***	37.41***	35.42***
	(0.0177)	(0.0174)	(0.0342)	(0.0281)
R-squared	0.068	0.043	0.036	0.043
Rounded Obs	1.230e+07	1.150e+07	1.660e+06	1.980e+06

Note: Parental household income comes from the 1040 forms and is the average of the first four years of tax data available post-birth of the child. Household income of the child comes from the 1040 forms and is the average of the five years during which the child is aged 35-39. Individual earnings come from W-2 forms during the years in which the child is aged 35-39, except for the birth cohorts of 1964-1969. Their earnings are measured during ages 41-45 as the W-2 data begins in 2005. Children are assigned percentile ranks relative to all other children in their birth cohort, while parents are ranked relative to all parents with children in the same birth cohort. All racial groups exclude individuals of Hispanic ethnicity. Counts are rounded in the following manner: numbers between 10,000 and 99,999 are rounded to the nearest 500; between 100,000 and 9,999,999 to the nearest 1,000; and above 10,000,000 to the nearest 10,000 to meet Census disclosure rules. Robust standard errors are included in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.29: Summary Statistics of County and Tract Characteristics in 1970

	(1)	(2)	(3)
Panel A	County Characteristics		
Variable	White Pop	Black Pop	Std Dev
Pct HS Grad	0.532	0.493	0.126
Pct White	0.925	0.804	0.137
Pct Occup Q1	0.328	0.346	0.054
Pct Occup Q5	0.097	0.095	0.025
Avg Income	46210	44880	7430
Employment Rate	0.699	0.690	0.092
Labor Force Participation Rate	0.769	0.764	0.078
Employment Rate (White)	0.707	0.707	0.091
Labor Force Participation Rate (White)	0.776	0.779	0.076
Employment Rate (Black)	0.565	0.614	0.245
Labor Force Participation Rate (Black)	0.674	0.702	0.244
Panel B	Tract Characteristics		
Variable	White Pop	Black Pop	Std Dev
Pct HS Grad	0.583	0.485	0.159
Pct White	0.964	0.648	0.201
Pct Occup Q1	0.302	0.371	0.083
Pct Occup Q5	0.120	0.076	0.072
Avg Income	53450	43030	17310
Employment Rate	0.752	0.706	0.121
Labor Force Participation Rate	0.812	0.776	0.103
Employment Rate (White)	0.753	0.686	0.131
Labor Force Participation Rate (White)	0.813	0.751	0.114
Employment Rate (Black)	0.646	0.681	0.313
Labor Force Participation Rate (Black)	0.741	0.763	0.284

Note: County and tract characteristics are calculated using the Decennial Census in 1970. Columns by race weight the location characteristic with population by race. The standard deviation of the characteristic across counties or tracts is also reported. Percentage high school graduate, percentage White, and percentages in each occupational quintile is calculated among individuals aged 16 and up. Occupations are ordered based on nation-wide median income among the employed into five bins. Average income, employment rate, and labor force participation rate is calculated among men aged 16 and up. Employment rate has men aged 16+ as the denominator and employment among men aged 16+ as the numerator. Employment rate and labor force participation rate are also calculated just among White and Black men.

Table A.30: Job Market Access and Tract-Level Income, Employment Rate, and Labor Force Participation Rate in Levels (1970)

	(1)	(2)	(3)
		Panel A – OLS	
Variables	Log Avg Income	Employment Rate	Labor Force Participation Rate
Log JMA, 1970	0.0474***	0.0194***	0.0183***
	(0.0102)	(0.00349)	(0.00306)
R-squared	0.178	0.230	0.206
		Panel B – IV Highway [KP Wald F-Stat = 2873]	
Log JMA, 1970	0.0585***	0.0197***	0.0183***
	(0.0110)	(0.00369)	(0.00326)
R-squared	0.138	0.138	0.135
		Panel C – IV Plans [KP Wald F-Stat = 2383]	
Log JMA, 1970	0.0810***	0.0281***	0.0260***
	(0.0125)	(0.00421)	(0.00371)
R-squared	0.138	0.137	0.134
		Panel D – IV Rays [KP Wald F-Stat = 2366]	
Log JMA, 1970	0.0889***	0.0332***	0.0310***
	(0.0127)	(0.00429)	(0.00379)
R-squared	0.137	0.137	0.133
CBSA FE	Yes	Yes	Yes
Rounded Obs	20500	20500	20500

Note: Tract characteristics are calculated using the Decennial Census in 1970. Percentage high school graduate, percentage White, and percentages in each occupational quintile is calculated among individuals aged 16 and up. Occupations are ordered based on nation-wide median income among the employed into five bins. Average income, employment rate, and labor force participation rate is calculated among men aged 16 and up. Employment rate has men aged 16+ as the denominator and employment among men aged 16+ as the numerator. Kleibergen-Paap rk Wald statistics are reported for the first-stage. Counts are rounded in the following manner: numbers between 10,000 and 99,999 are rounded to the nearest 500; between 100,000 and 9,999,999 to the nearest 1,000; and above 10,000,000 to the nearest 10,000 to meet Census disclosure rules. Robust standard errors are included in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A.31: The Effect of Job Market Access Improvements on Changes in Tract-Level Income, Employment Rate, and Labor Force Participation Rate (1960-1970)

	(1)	(2)	(3)
		Panel A – OLS	
Variables	Δ Log Avg Income	Δ Employment Rate	Δ Labor Force Participation Rate
Δ Log JMA	0.0740***	0.0524***	0.0184***
	(0.0117)	(0.00545)	(0.00437)
R-squared	0.129	0.100	0.0978
		Panel B – IV Highway [KP Wald F-Stat = 1261]	
Δ Log JMA	0.0796***	0.0493***	0.0303***
	(0.0148)	(0.00743)	(0.00650)
R-squared	0.0580	0.0625	0.0600
		Panel C – IV Plans [KP Wald F-Stat = 621]	
Δ Log JMA	0.180***	0.0708***	0.0589***
	(0.0347)	(0.0170)	(0.0148)
R-squared	0.0533	0.0618	0.0565
		Panel D – IV Rays [KP Wald F-Stat = 562]	
Δ Log JMA	0.234***	0.108***	0.0695***
	(0.0382)	(0.0184)	(0.0155)
R-squared	0.0472	0.0566	0.0542
CBSA FE	Yes	Yes	Yes
Rounded Obs	20500	20500	20500

Note: Tract characteristics are calculated using the Decennial Census in 1960 and 1970. Percentage high school graduate, percentage White, and percentages in each occupational quintile is calculated among individuals aged 16 and up. Occupations are ordered based on nation-wide median income among the employed into five bins. Average income, employment rate, and labor force participation rate is calculated among men aged 16 and up. Employment rate has men aged 16+ as the denominator and employment among men aged 16+ as the numerator. Kleibergen-Paap rk Wald statistics are reported for the first-stage. Counts are rounded in the following manner: numbers between 10,000 and 99,999 are rounded to the nearest 500; between 100,000 and 9,999,999 to the nearest 1,000; and above 10,000,000 to the nearest 10,000 to meet Census disclosure rules. Robust standard errors are included in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

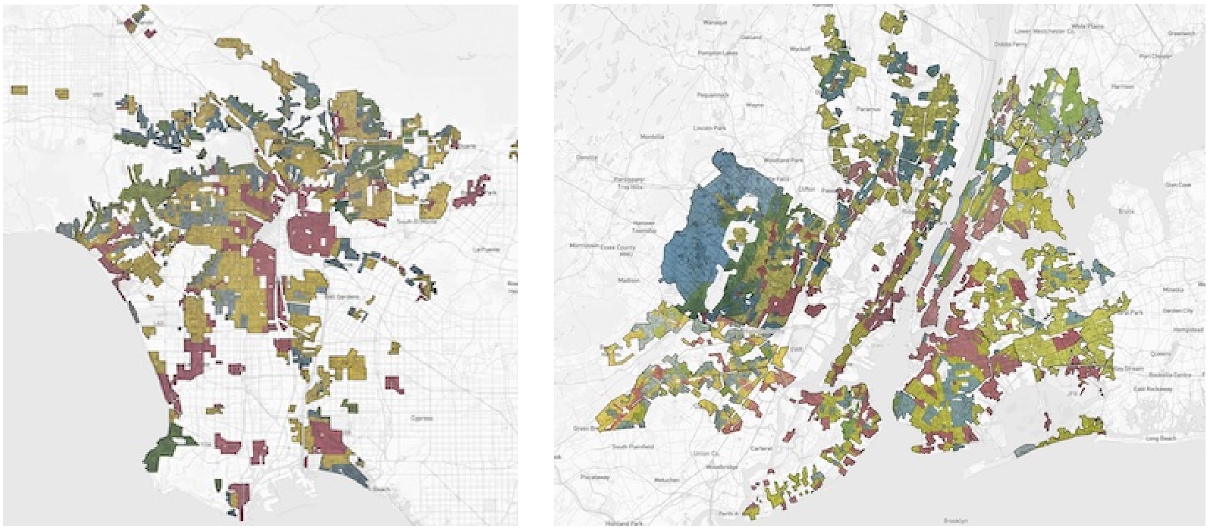
Table A.32: Parent Movers Panel - Two-way FE Income Changes in JMA

	(1)	(2)
	Log Income of Parents	
Variables	OLS	IV HW
Log JMA, 1970	-0.0203***	-0.0214***
	(0.00295)	(0.00311)
R-squared	0.581	0.0865
Rounded Obs	19800000	19800000
Person FE	Yes	Yes
Year FE	Yes	Yes
CBSA FE	Yes	Yes

Note: Parents who move once starting in the first year the 1040 data is available in 1974 up until the year their child is age 23 are included in the sample. Job market access is calculated in 1970 with the Decennial Census data. The instrument for job market access aggregates over wages and employment in 1960 discounted by commute costs induced by the Interstate highway system in 1970. Counts are rounded in the following manner: numbers between 10,000 and 99,999 are rounded to the nearest 500; between 100,000 and 9,999,999 to the nearest 1,000; and above 10,000,000 to the nearest 10,000 to meet Census disclosure rules.

2 Figures

Figure B.1: Redlining Maps for Los Angeles and New York City



(a) Los Angeles, California

(b) New York City, New York

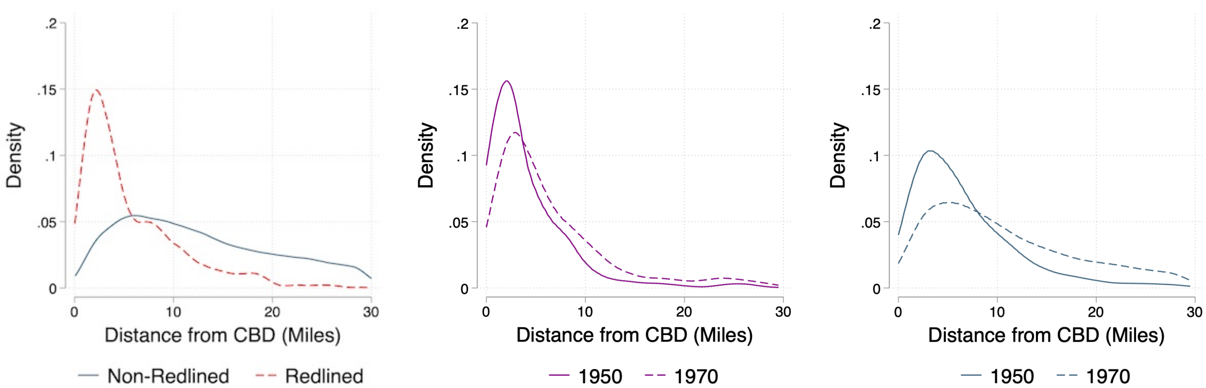
Notes: HOLC Maps for Los Angeles and New York City from Mapping Inequality: Redlining in New Deal America.

Figure B.2: Spatial Distribution Population By Race in 1950 and 1970 and
Redlined vs. Non-Redlined Tracts Over Distance from Central Business District

(a) Redlining

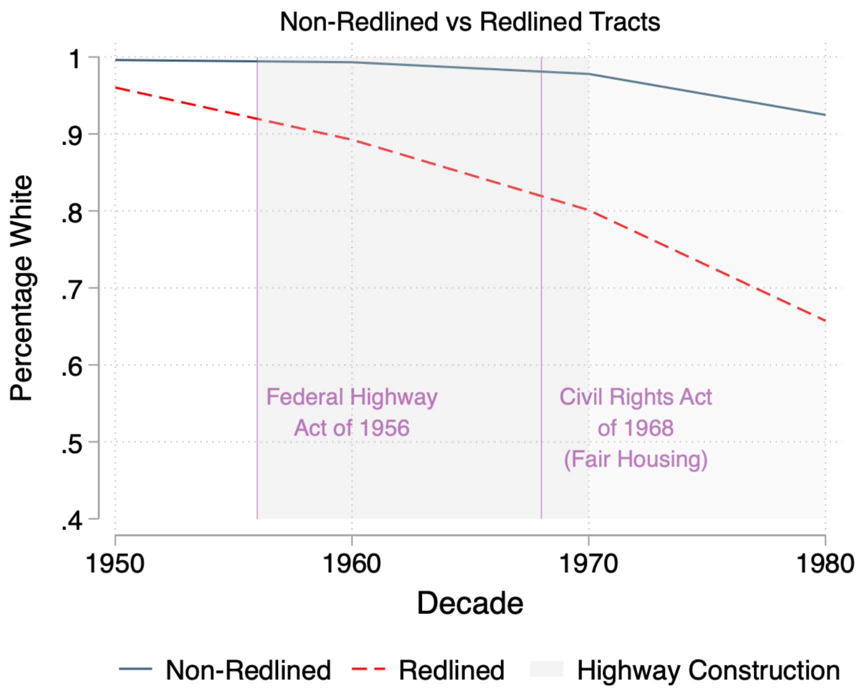
(b) Black

(c) White



Notes: Data comes from IPUMS NHGIS in 1950 and 1970. Epanechnikov kernel density estimation for the spatial distribution of the Black vs. White population is over 60 bins of distance from the central business district in miles with population weights. Kernel density for redlined vs. non-redlined tracts uses the raw counts of tracts. Redlined tracts are tracts where more than 80% of the area is redlined.

Figure B.3: Racial Composition of Median Tract 1950-1990



Notes: Data comes from IPUMS NHGIS tract-level aggregates from 1950 to 1990 for a constant sample of tracts that have population by race in 1950 and are in the sample of 100 cities with Yellow Book maps. Gray areas in the graph are for periods of highway construction starting in 1956.

Figure B.4: Yellow Book Maps for Atlanta and Cleveland

(a) Atlanta, Georgia (b) Cleveland, Ohio

Notes: Yellow Book (General Location of National System of Interstate Highways Including All Additional Routes at Urban Areas Designated in September 1955) maps retrieved from the Bureau of Public Roads.

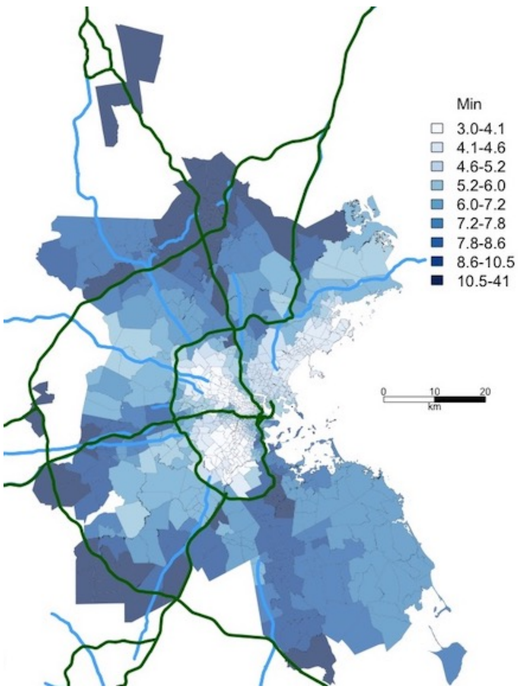
Figure B.5: Historical Roads from Shell Atlases for Baltimore and San Francisco

(a) Baltimore, Maryland

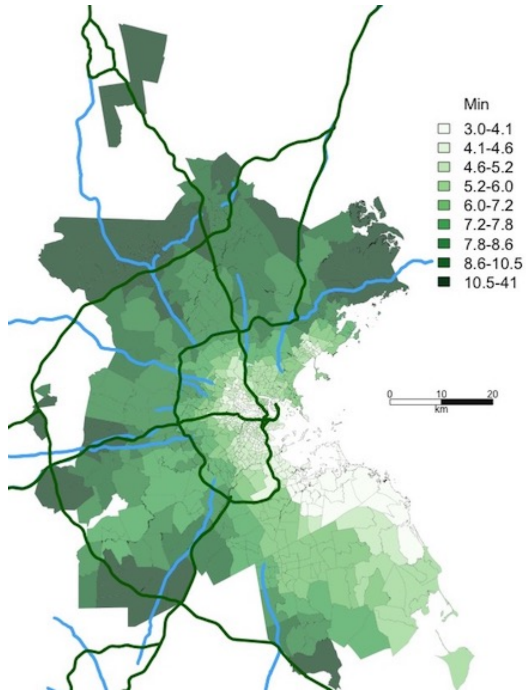
(b) San Francisco, California

Notes: Shell Atlases by the H.M. Gousha Company in 1956 retrieved from the Rumsey Collection for Baltimore and San Francisco.

Figure B.6: Commute Time Reductions with Historical Roads as Interstates



(a) Interstate Roads

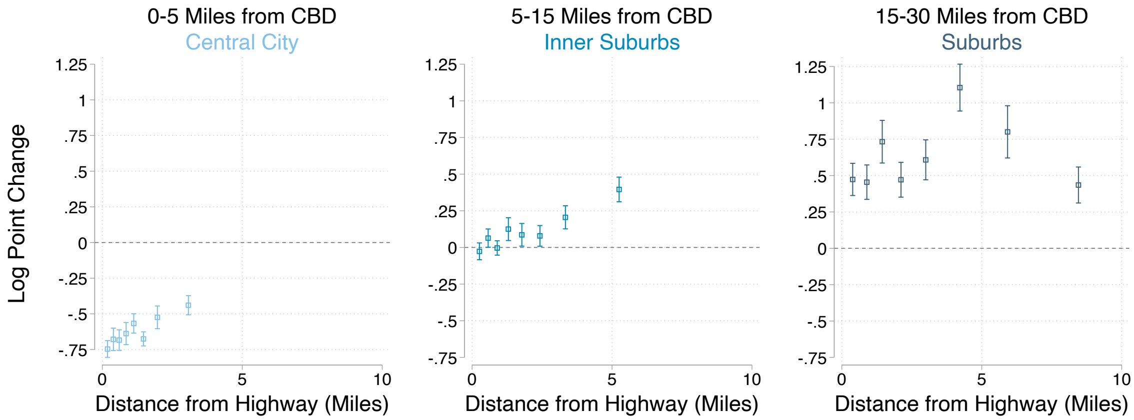


(b) Historical Large Roads as Interstates

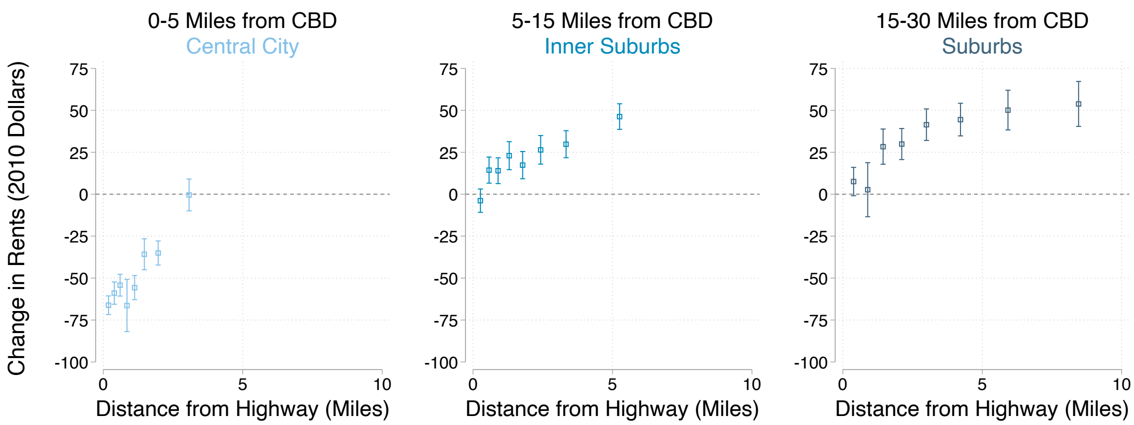
Notes: In Panel (a), commute time changes come from the author's calculations as the difference between commute times for the historical road network and for the entire Interstate network overlayed on the historical road network. In Panel (b) commute time changes come from the author's calculations as the difference between commute times for the historical road network and for the development of large roads as interstate highways.

Figure B.7: Changes Over Distance from Central Business District and Distance from Highway (1950-1960, 1960-1970)

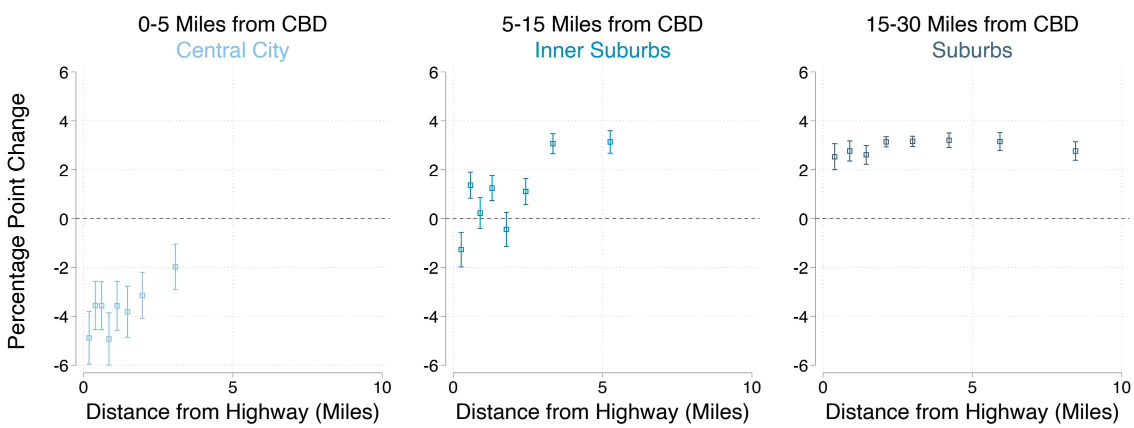
(a) Change in Log Population



(b) Change in Rents

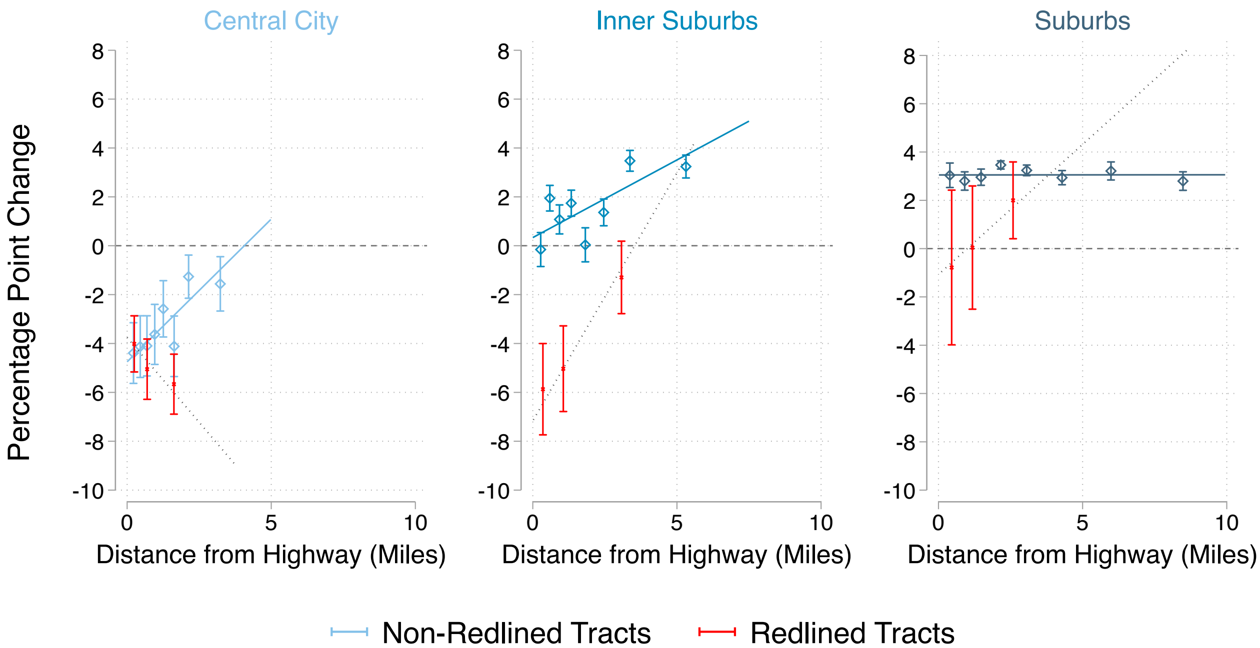


(c) Change in Percentage White



Notes: Unit of observation is census tract. Data comes from 1950, 1960, and 1970 tract-level aggregates retrieved from IPUMS NHGIS. The first difference is either over 1950 to 1960 or 1960 to 1970 and stacked into one panel depending on when highway construction started in the CBSA. All changes over time are de-meaned within CBSA. The sample of tracts for the central city panel is tracts within 5 miles of the constructed highway network, for the inner suburbs panel is tracts within 7.5 miles of the constructed highway network, and for the suburbs panel is tracts within 10 miles of the constructed highway network for legibility.

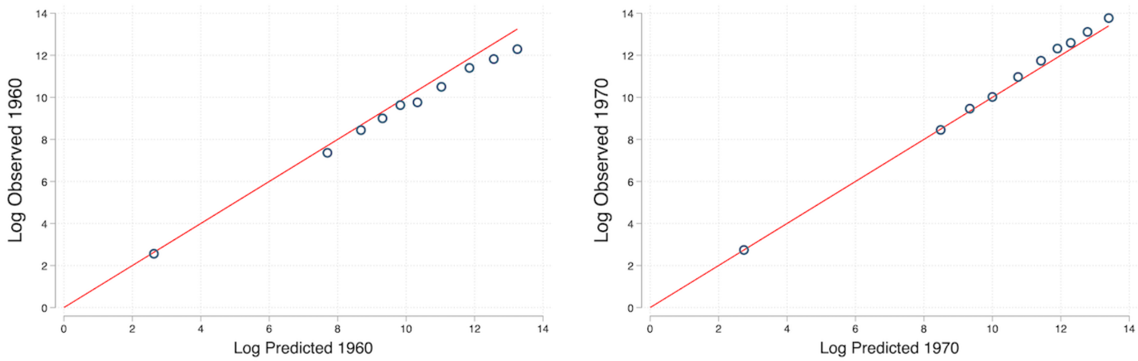
Figure B.8: Change in Percentage White by Redlining
Over Distance from Central Business District and Distance from Highway (1950-1960, 1960-1970)



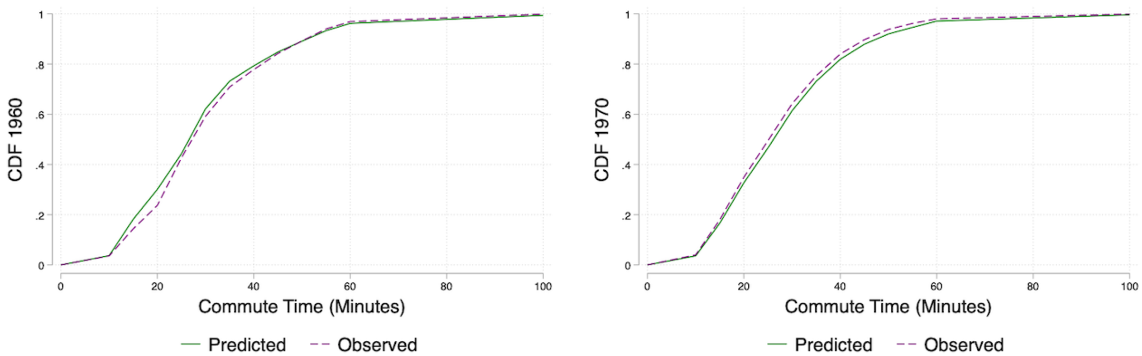
Notes: Unit of observation is census tract. Data comes from 1950, 1960, and 1970 tract-level aggregates retrieved from IPUMS NHGIS. The first difference is either over 1950 to 1960 or 1960 to 1970 and stacked into one panel depending on when highway construction started in the CBSA. All changes over time are de-meaned within CBSA. The sample of tracts for the central city panel is those within 5 miles of the constructed highway network, for the inner suburbs panel is those within 7.5 miles of the constructed highway network, and for the suburbs panel is those within 10 miles of the constructed highway network for legibility. Redlined tracts are those where more than 80% of the area is redlined.

Figure B.9: Predicted vs. Observed Commute Flows in 1960 and 1970

(a) Linear Fit and Scatter

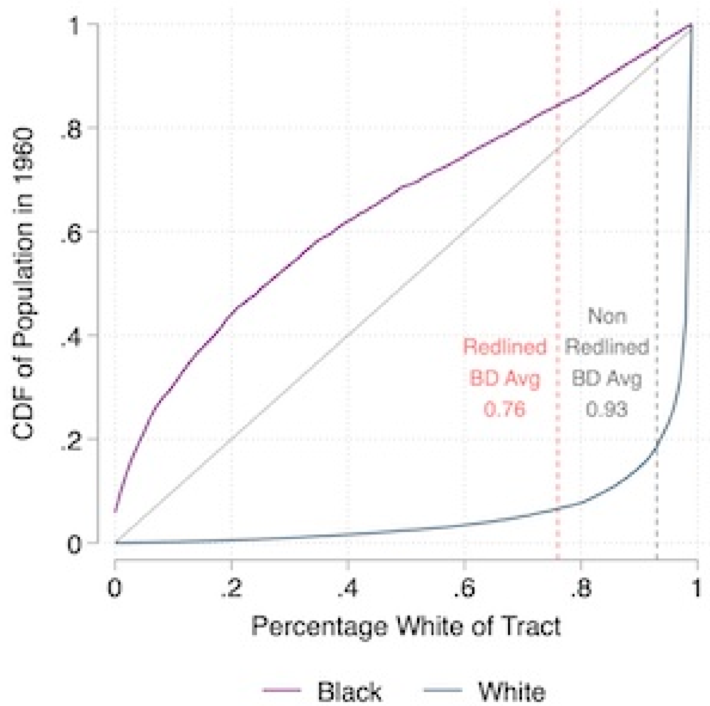


(b) Cumulative Distribution Function



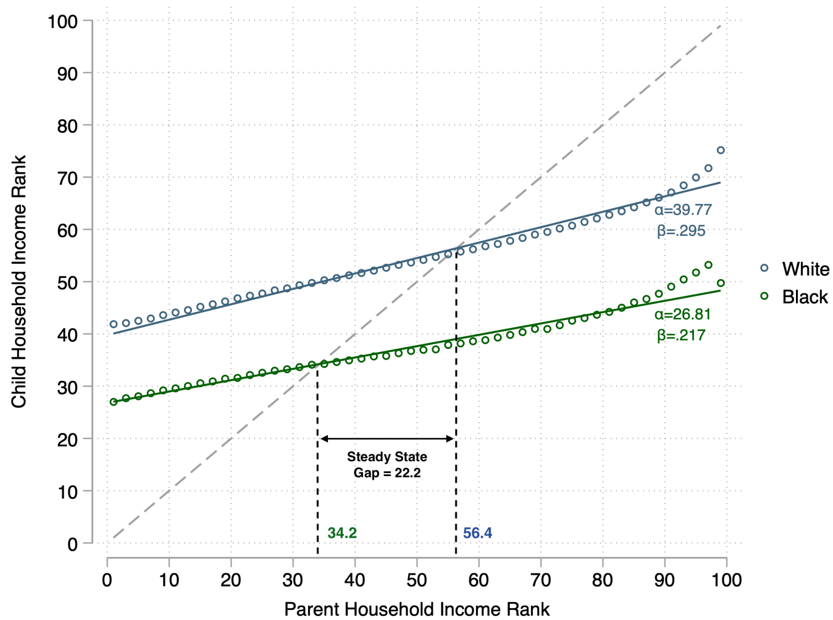
Notes: Unit of observation is Place of Work Zone level by Place of Work Zone level. Data comes from restricted Census microdata in 1960 and 1970. In Panel (a), the scatter plot is created with 10 quantiles of predicted flows with analytical weights on the level of the observed commute flows. The red line is the 45 degree line. The linear fit is shown in Table A.17. In Panel (b), the cumulative distribution function over commute time in minutes is in predicted flows for the green line and in observed flows for the purple line.

Figure B.10: CDF of Black and White Population in 1960 over Racial Composition of Neighborhood and Neighborhood Average by Redlining Borders



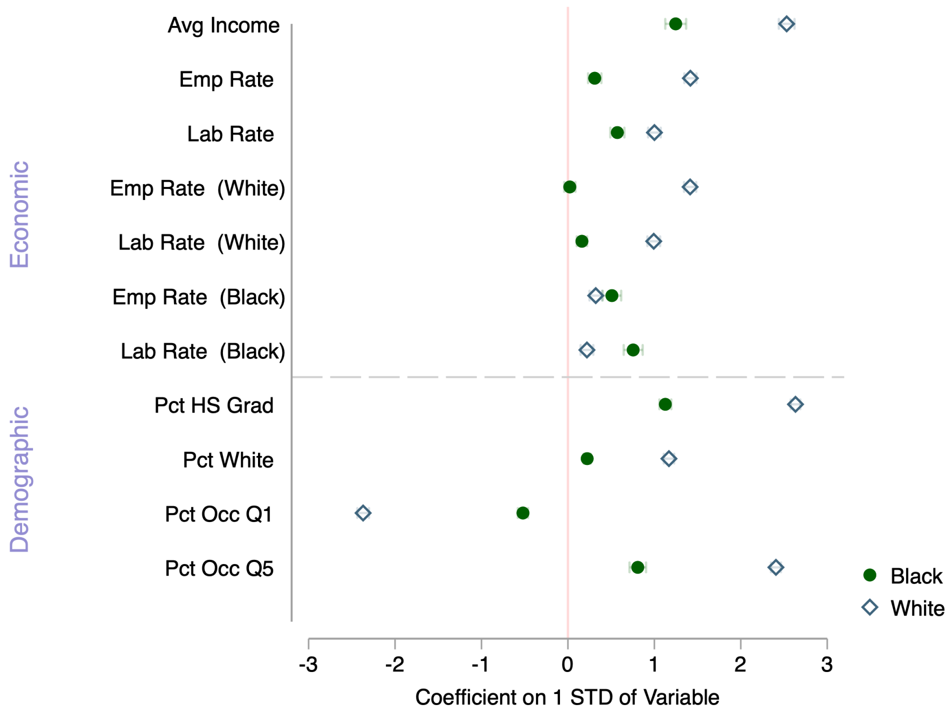
Notes: Data comes from the 1960 tract-level aggregates retrieved from IPUMS NHGIS. The cumulative distribution function is calculated by applying analytical weights of each race group's population to the racial composition of census tracts. The average percentage White of redlined neighborhoods and non-redlined neighborhoods is calculated from the sample by the border of Figure 1.

Figure B.11: Child Household Income Rank by Race - Steady State



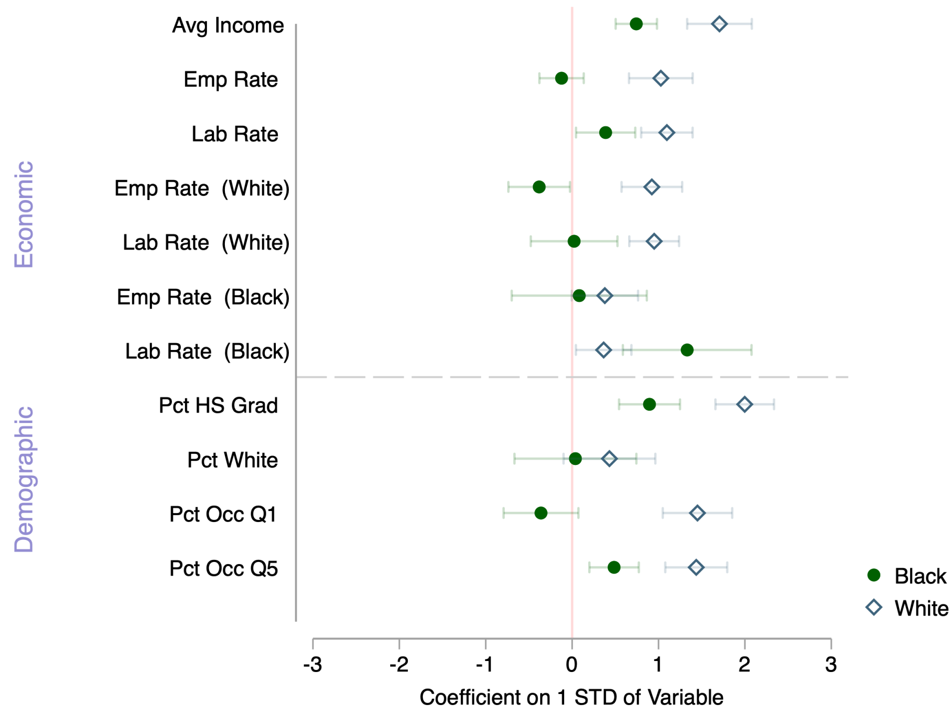
Note: Average child household income rank is measured over 50 bins across parental household income rank separately for White and Black individuals using the 1040 tax data. Parental household income comes from the 1040 forms and is the average of the first four years of tax data available post-birth of the child. Household income of the child comes from the 1040 forms and is the average of the five years during which the child is aged 35-39. Children are assigned percentile ranks relative to all other children in their birth cohort, while parents are ranked relative to all parents with children in the same birth cohort. All racial groups exclude individuals of Hispanic ethnicity. The best-fit lines are estimated using an OLS regression on the individual observations and are displayed in Table 13. The slopes β_r and intercepts α_r from these regressions are reported for each race. White-black differences in mean child household income rank are reported at the 25th and 75th percentiles of the parent income distribution. The 45-degree line is also plotted, and where it intersects the best-fit lines by race gives the steady state income ranks if the rank-rank relationships by race persist over time. Therefore, the steady-state income rank Black-White gap is 22.2 ranks.

Figure B.12: Descriptive Correlations between Predicted Child Income Rank (P25) and Tract-Level Characteristics



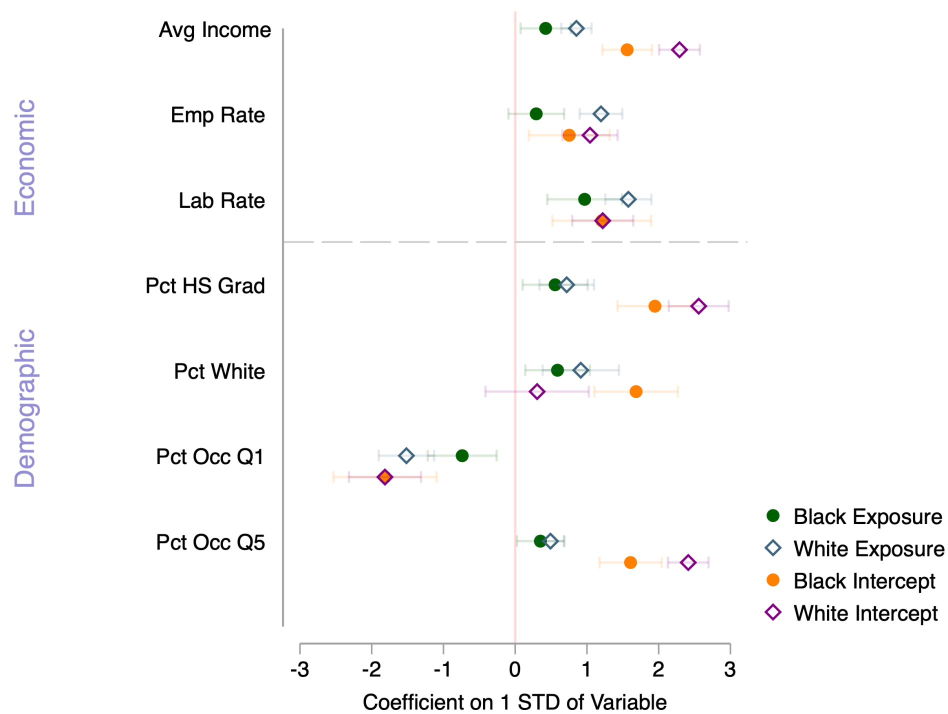
Note: Predicted income rank is computed by estimating linear rank-rank correlations for each racial group in each tract and then predicting the rank of children from the 25th percentile of the parent income distribution. Tract characteristics are calculated using the Decennial Census in 1970. Percentage high school graduate, percentage White, and percentages in each occupational quintile is calculated among individuals aged 16 and up. Occupations are ordered based on nation-wide median income among the employed into five bins. Average income, employment rate, and labor force participation rate is calculated among men aged 16 and up. Employment rate has men aged 16+ as the denominator and employment among men aged 16+ as the numerator. Employment rate and labor force participation rate are also calculated just among White and Black men.

Figure B.13: Descriptive Correlations between Predicted Child Income Rank (P25) and County-Level Characteristics



Note: Predicted income rank is computed by estimating linear rank-rank correlations for each racial group in each county and then predicting the rank of children from the 25th percentile of the parent income distribution. County characteristics are calculated using the Decennial Census in 1970. Percentage high school graduate, percentage White, and percentages in each occupational quintile is calculated among individuals aged 16 and up. Occupations are ordered based on nation-wide median income among the employed into five bins. Average income, employment rate, and labor force participation rate is calculated among men aged 16 and up. Employment rate has men aged 16+ as the denominator and employment among men aged 16+ as the numerator. Employment rate and labor force participation rate are also calculated just among White and Black men.

Figure B.14: Causal Impacts of Counties and County-Level Characteristics



Note: Causal impacts of counties come from a movers design along county characteristics from origin to destination. County characteristics are calculated using the Decennial Census in 1970. Percentage high school graduate, percentage White, and percentages in each occupational quintile is calculated among individuals aged 16 and up. Occupations are ordered based on nation-wide median income among the employed into five bins. Average income, employment rate, and labor force participation rate is calculated among men aged 16 and up. Employment rate has men aged 16+ as the denominator and employment among men aged 16+ as the numerator.

3 Descriptive Results

3.1 First-Stage

To test that the planned routes and Euclidean rays are predictive of highway placement, I estimate two types of first-stage equations of the forms below where $IV \in \{Plans, Rays\}$.

$$\log(DistHW_i) = \phi \log(DistIV_i) + \mathbf{X}_i\theta + \lambda_{m(i)} + \nu_i$$
$$\mathbf{1}\{DistHW_i = 1\} = \pi\mathbf{1}\{DistIV_i = 1\} + \mathbf{X}_i\sigma + \delta_{m(i)} + \xi_i$$

The first compares log distance from the constructed routes to log distance from either the planned routes or the Euclidean rays, and the second compares a binary indicator for whether tracts are within 1 mile of the constructed route to the same indicator for the planned routes and rays to study placement at finer spatial scales. All the earlier controls and city fixed effects are included in the estimation. Results are shown in Table A.5. In both forms of first-stage regressions, the instruments are highly correlated with highway location as F-statistics on the excluded instrument are all above 100.

4 Quantitative Model

4.1 Model Extensions

Stone-Geary

An alternative approach to incorporating non-homotheticity in housing consumption is to allow for Stone-Geary preferences. The consumer maximization problem then includes a minimum amount of housing consumption $\bar{l}_{gr}$ that can differ by group.

$$\max_{c_{ij}(o),l_i(o)} \quad \frac{z_i(o)\epsilon_j(o)B_{igr}}{d_{ijgr}}\left(\frac{c_{ij}(o)}{\beta_{gr}}\right)^{\beta_{gr}}\left(\frac{l_i(o)-\bar{l}_{gr}}{1-\beta_{gr}}\right)^{1-\beta_{gr}}$$
$$\text{s.t.} \quad c_{ij}(o)+Q_i l_i(o)=\frac{w_{jgr}\epsilon_j(o)}{d_{ijgr}}$$

which leads to the indirect utility function of

$$u_{ijgr}(o)=B_{igr}z_i(o)Q_i^{\beta_{gr}-1}\left(\frac{w_{jgr}\epsilon_j(o)}{d_{ijgr}}-Q_i\bar{l}_{gr}\right)$$

Home Ownership

To incorporate home ownership into the incomes of households, the budget constraint of the household is modified to include an additional source of income from housing rents that are redistributed across groups. The consumer maximization problem is then

$$\max_{c_{ij}(o),l_i(o)} \quad \frac{z_i(o)\epsilon_j(o)B_{igr}}{d_{ijgr}}\left(\frac{c_{ij}(o)}{\beta_{gr}}\right)^{\beta_{gr}}\left(\frac{l_i(o)}{1-\beta_{gr}}\right)^{1-\beta_{gr}}$$
$$\text{s.t.} \quad c_{ij}(o)+Q_i l_i(o)=w_{jgr}(1+\nu_{gr})$$

which leads to the indirect utility function of

$$u_{ijgr}(o)=\frac{z_i(o)\epsilon_j(o)B_{igr}Q_i^{\beta_{gr}-1}w_{jgr}(1+\nu_{gr})}{d_{ijgr}}$$

In the form above, homeownership enters into utility in a multiplicative way and is constant across neighborhoods for each group within each city. Expected income is multiplied by the same factor following

$$\overline{w}_{igr}(1+\nu_{gr})=\Big(\sum_j \pi_{j|igr}w_{jgr}\Big)(1+\nu_{gr})$$

Total expenditure on housing for neighborhood i by group gr is then

$$Exp_{igr}=(1-\beta_{gr})\overline{w}_{igr}(1+\nu_{gr})L_{igr}$$

Let total income by group gr be the sum of total labor income and total rental income from the redistribution of housing rents to each group based on the share of home values that the group owns in the portfolio of the city.

$$\underbrace{(1+\nu_{gr})\sum_i \overline{w}_{igr}L_{igr}}_{\text{total income}} = \underbrace{\sum_i \overline{w}_{igr}L_{igr}}_{\text{total labor income}} + \underbrace{\sum_i \hat{o}_{igr}\sum_{g,r} Exp_{igr}}_{\text{total rental income}}$$

$$\Rightarrow \nu_{gr} = \frac{\sum_i \hat{o}_{igr}\sum_{g,r} Exp_{igr}}{\sum_i \overline{w}_{igr}L_{igr}}$$

The share of home values $\hat{o}_{igr}$ is observed in the data and is defined as the proportion of home values in homes owned by group gr households out of total home values in a neighborhood.

Nested Frechet

With the indirect utility function defined as before, suppose that $z_i(o)$ is instead distributed with a nested Frechet structure where the cumulative distribution is

$$F(z_i(o)) = \exp\Big(-\Big(\sum_n\Big(\sum_{i\in S_n} z_i(o)^{-\theta_r}\Big)^{-\lambda_r/\theta_r}\Big)\Big)$$

where λ_r governs the substitutability across types of neighborhoods. When $\lambda_r = \theta_r$, the expression returns to the usual Frechet distribution as before. In this setting, suppose $\lambda_B < \theta_B$ where for Black households, there is more substitutability within type of neighborhood than across, and $\lambda_W = \theta_W$ where for White households, their choice behavior is not nested across the type of neighborhoods.

The choice probabilities are then

$$\begin{aligned}\pi_{igr} &= \pi_{n,gr}\pi_{igr|n}\\ &= \frac{\left(\sum_{s\in S_n}\left(B_{sgr}CMA_{sgr}Q_s^{\beta_{gr}-1}\right)^{\theta_r}\right)^{\lambda_r/\theta_r}}{\sum_m\left(\sum_{s\in S_m}\left(B_{sgr}CMA_{sgr}Q_s^{\beta_{gr}-1}\right)^{\theta_r}\right)^{\lambda_r/\theta_r}}\frac{\left(B_{igr}CMA_{igr}Q_i^{\beta_{gr}-1}\right)^{\theta_r}}{\sum_{s\in S_n}\left(B_{sgr}CMA_{sgr}Q_s^{\beta_{gr}-1}\right)^{\theta_r}}\end{aligned}$$

following a two step process. First, there is a choice of type n of neighborhoods, and then conditional on type, there is a choice of neighborhood i within type n. Define $V_{n,gr} = \left(\sum_{s\in S_m}\left(B_{sgr}CMA_{sgr}Q_s^{\beta_{gr}-1}\right)^{\theta_r}\right)^{1/\theta_r}$ as the inclusive value of living in type n neighborhoods. The share living in type n follows the usual gravity share formula with the shape parameter λ_r.

$$\pi_{n,gr} = \frac{V_{n,gr}^{\lambda_r}}{\sum_m V_{m,gr}^{\lambda_r}}$$

The population elasticity to CMA $\frac{\partial \pi_{igr}}{\partial CMA_{igr}}$ using the product rule and the definition of $\pi_{igr} = \pi_{n,gr}\pi_{igr|n}$ is

$$\begin{aligned}
\frac{\partial \pi_{igr}}{\partial CMA_{igr}} &= \frac{\partial \pi_{n,gr}}{\partial CMA_{igr}}\pi_{igr|n} + \frac{\partial \pi_{igr|n}}{\partial CMA_{igr}}\pi_{n,gr} \\
\frac{\partial \pi_{igr}}{\partial CMA_{igr}} &= \frac{\lambda_r \pi^2_{igr|n}}{CMA_{igr}}\pi_{n,gr}(1-\pi_{n,gr}) + \frac{\theta_r \pi_{n,gr}}{CMA_{igr}}\pi_{igr|n}(1-\pi_{igr|n}) \\
\Rightarrow \frac{\partial \pi_{igr}}{\partial CMA_{igr}}\frac{CMA_{igr}}{\pi_{igr}} &= \lambda_r \pi_{igr|n}(1-\pi_{n,gr}) + \theta_r(1-\pi_{igr|n})
\end{aligned}$$

When $\lambda_r = \theta_r$ (as is the case for White households), then the elasticity becomes

$$\frac{\partial \pi_{igr}}{\partial CMA_{igr}}\frac{CMA_{igr}}{\pi_{igr}} = \theta_r \pi_{igr|n}(1-\pi_{n,gr}) + \theta_r(1-\pi_{igr|n}) = \theta_r(\pi_{igr|n} - \pi_{igr}) + \theta_r(1-\pi_{igr|n}) = \theta_r(1-\pi_{igr})$$

which is the same as when there are no nests for types.

When $\lambda_r < \theta$ (as is the case for Black households), then the elasticity is lower

$$\lambda_r \pi_{igr|n}(1-\pi_{n,gr}) + \theta_r(1-\pi_{igr|n}) < \theta_r(1-\pi_{igr})$$

even though the conditional elasticity (conditional on type of neighborhood) is still approximately θ.

$$\frac{\partial \pi_{igr|n}}{\partial CMA_{igr}}\frac{CMA_{igr}}{\pi_{igr|n}} = \theta_r(1-\pi_{igr|n})$$

5 Inversion and Estimation

5.1 Parameter Estimation

Estimation of discriminatory pricing

Discriminatory pricing will be measured directly from home values and rents in the microdata. To test for differential pricing by race across neighborhoods, I look at the coefficient from the interaction of race and redlining. The estimating equation is across observations for each household h with either log home value or log rent as the dependent variable.

$$\log Q_h = \alpha_i + \alpha_r + \phi_1 D_i^{red} + \phi_2 D_i^{red} \times D_h^{non-white} + X_h + \epsilon_h$$

α_i is for neighborhood fixed effects, α_r is for race fixed effects, D_i^{red} is a dummy for being in a redlined neighborhood, D_h^{black} is a dummy for the household head being Black, and X_h is a set of household level characteristics on the quality of the home such as the availability of air conditioning, a freezer, a toilet, or a bathtub. The coefficient ϕ_2 is the differential increase in price black households have to pay to live outside of redlined neighborhoods compared to white households. In Table A.1, it appears that Black households pay less than White households for similar quality housing in non-redlined neighborhoods and more in redlined neighborhoods. These results do not suggest pricing is the reason why Black households are more likely to live in redlined areas.

Linear prediction of housing consumption share from CEX data

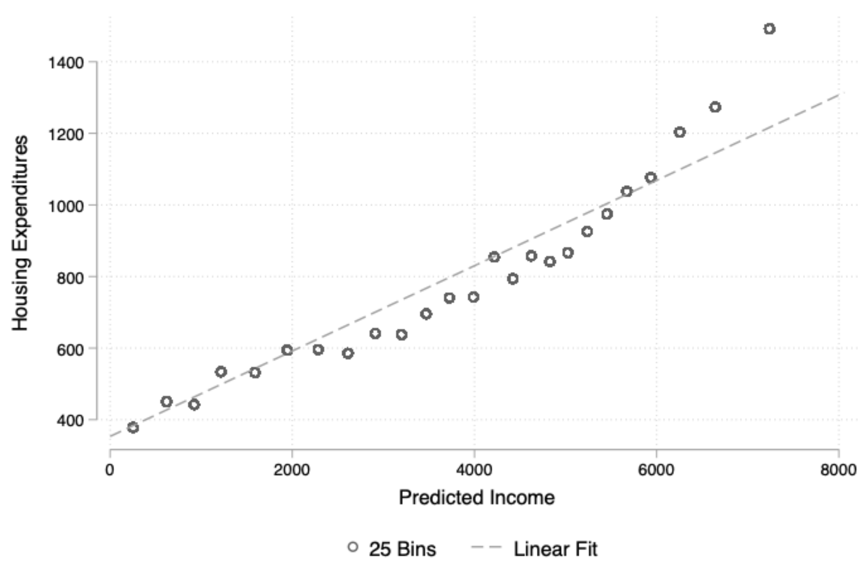
I assign the housing consumption share for each race by education by first estimating a linear function for housing expenditure over income from the Consumer Expenditure Surveys Public-Use Microdata in the year 1980.

$$E_i^{hous} = \beta_{gr}^0 + \beta_{gr}^1 PredIncome_i + \epsilon_i$$

E_i^{hous} is quarterly housing expenditure and $PredIncome_i$ is quarterly income predicted using categorical variables in age, education, marital status, occupation, sex, race and region. I use predicted rather than observed income given the variability in observed income that would lead to downward biased estimates of β_{gr}^1.

As depicted in Figure E.15, the assumption of linearity for the housing consumption Engel curve appears to be satisfied. From this expenditure function, I calculate the predicted housing expenditure for each race by education group given their average income and use the ratio of predicted housing expenditure to average income as the housing consumption share.

Figure E.15: Linear Housing Expenditure Function Over Income



Notes: Unit of observation is individual. Data comes from the Consumer Expenditure Surveys Public-Use Microdata in 1980. Predicted income is a linear prediction of income using categorical variables in age, education, marital status, occupation, sex, race and region. Income and housing expenditure is for quarterly amounts. 25 equally sized bins in predicted income (when predicted income is greater than zero) are created for the scatter. The linear fit uses the estimated coefficients from Table A.15.

5.2 Model Inversion

Parameters Estimated During Model Inversion

- $\alpha_{jg}, \alpha_{jgr}$ are labor intensities for the CES nested labor aggregate
- T_{jgr} is the scale parameter for workplaces

Observed data sources

- Observed wages $\hat{w}_{jgr}$ come from the Decennial microdata

Step 1 – Given $\{L_{Fjgr}, L_{igr}, t_{ijgr}\}$ and the semi-elasticity of commuting parameter $\{\nu_{gr}\}$, I invert for composite transformed wages $\omega_{jgr} = T_{jgr} w_{jgr}^{\phi}$ from the labor supply equation following

$$
\begin{aligned}
L_{Fjgr} &= \sum_i \frac{T_{jgr}(w_{jgr}/d_{ijgr})^{\phi}}{\sum_s T_{sgr}(w_{sgr}/d_{is})^{\phi}} L_{igr} \\
&= \sum_i \frac{\omega_{jgr}/\exp(\nu_{gr} t_{ijgr})}{\sum_s \omega_{sgr}/\exp(\nu_{gr} t_{isgr})} L_{igr}
\end{aligned}
$$

Commuting costs are in terms of commute times t_{ijgr} following $d_{ijgr} = t_{ijgr}^{\kappa_{gr}}$, therefore $d_{ijgr}^{\phi} = t_{ijgr}^{\nu_{gr}}$ with $\nu_{gr} = \kappa_{gr}\phi$. Labor supply is in the second line rewritten as a function of composite transformed wages ω_{jgr}. Wages are solved for iteratively following the process of Ahlfeldt et al. (2015) and wages are only identified up to a scaling factor.

Step 2 – Given $\{\omega_{jgr}\}$, the Frechet shape parameter for labor supply ϕ, and observed wages $\{\hat{w}_{jgr}\}$, I back out the Frechet scale parameter T_{jgr}. Following that $\omega_{jgr} = T_{jgr} w_{jgr}^{\phi}$, then the Frechet scale parameter is $T_{jgr} = \omega_{jgr}/w_{jgr}^{\phi}$. Compared to existing work where wages are not directly observed, this additional data allows for separately identifying the workplace amenity value T_{jgr} from the scale wages component ω_{jgr}.

Residential Side

Step 3 – Given $\{Q_i, \omega_{jgr}, t_{ijgr}, L_{igr}\}$ and the parameters $\{\beta_{gr}, \phi, \kappa_{gr}, \theta_r\}$, I can recover residential amenities B_{igr}. Returning to the residential choice equation, the share of each demographic group that lives in a location i follows

$$
\frac{L_{igr}}{\mathbb{L}_{gr}} = \frac{\left(B_{igr} CMA_{igr} Q_i^{\beta_{gr}-1}\right)^{\theta_r}}{\sum_t \left(B_{tgr} CMA_{tgr} Q_{tr}^{\beta_{gr}-1}\right)^{\theta_r}}
$$

which can be rearranged using the welfare equation (28).

$$
\left(\frac{L_{igr}}{\mathbb{L}_{gr}}\right)^{1/\theta_r} = \frac{B_{igr} CMA_{igr} Q_i^{\beta_{gr}-1}}{U_{gr}}
$$

Choosing units for amenities such that the geometric mean $\bar{B}_{igr} = \left[\prod_{i=1}^{S} B_{igr}\right]^{1/S} = 1$, and continuing with the bar notation for geometric mean, I calibrate amenities following

$$
\frac{B_{igr}}{\bar{B}_{igr}} = \left(\frac{L_{igr}}{\bar{L}_{Rigr}}\right)^{1/\theta_r} \left(\frac{Q_i}{\bar{Q}_{ir}}\right)^{1-\beta_{gr}} \left(\frac{CMA_{igr}}{\overline{CMA}_{igr}}\right)^{-1}
$$

Workplace Side

Step 4 – Given $\{L_{Fjgr}, w_{jgr}\}$ and the parameters $\{\sigma^r, \sigma^g\}$, I estimate the parameters α_{jg}, α_{jgr} with the following procedure. Using the labor demand equation from (8), the share of labor employed in a location and in an education group g that is of race r is

$$\frac{L_{Fjgr}}{L_{Fjg}} = \frac{(w_{jgr}/\alpha_{jgr})^{-\sigma^r}}{\sum_s (w_{jgs}/\alpha_{jgs})^{-\sigma^r}}$$

The share of labor and wages are observed, so this equation allows for determining α_{jgr} with the constraint that $\sum_r \alpha_{jgr} = 1$.

With a similar process, I solve for α_{jg}. First, I calculate $N_{jg} = \left(\sum_r \alpha_{jgr} L_{Fjgr}^{\frac{\sigma^r-1}{\sigma^r}}\right)^{\frac{\sigma^r}{\sigma^r-1}}$ which is a function of observed and previously estimated values. Using the CES demand form, I arrive at the equation

$$\frac{N_{jg}}{\sum_h N_{jh}} = \frac{(\omega_{jg}/\alpha_{jg})^{-\sigma^g}}{\sum_h (\omega_{jh}/\alpha_{jh})^{-\sigma^g}}$$

which is an equation for the unknown α_{jg} with the constraint that $\sum_g \alpha_{jg} = 1$. Recall that $\omega_{jg} = \left(\sum_r \alpha_{jgr}^{\sigma^r} w_{jgr}^{1-\sigma^r}\right)^{\frac{1}{1-\sigma^r}}$, which is a function of known values.

Step 5 – Given $\{q_i, w_{jgr}\}$ and the parameters $\{\alpha, \alpha_{jg}, \alpha_{jgr}\}$, I recover workplace productivity A_i. Productivity for each location i is inferred from the zero profit equation.

$$q_i = (1-\alpha)\left(\frac{\alpha}{W_j}\right)^{\frac{\alpha}{1-\alpha}} A_i^{\frac{1}{1-\alpha}} \text{ for } i \in Tracts_j$$

where $W_j = \left(\sum_g \alpha_{jg}^{\sigma^g} \omega_{jg}^{1-\sigma^g}\right)^{\frac{1}{1-\sigma^g}}$ the price index for labor is calculated after backing out wages w_{jgr} and the $\alpha_{jg}, \alpha_{jgr}$ relative productivity parameters at the POW zone j. Since prices are observable at the tract level for tract i, I assume that wages are the same for all tracts in POW zone j which is the set $Tracts_j$.

Housing Supply and Allocation

Step 6 – Given $\{Q_i, \omega_{jgr}, t_{ijgr}, L_{igr}, q_i, A_j, L_{Fjgr}\}$, the parameters $\{\beta_{gr}, \phi, \kappa_{gr}, \alpha, \alpha_{jg}, \alpha_{jgr}\}$, I recover total housing supply H_i and floorspace allocation θ_i across commercial and residential uses. Returning to the residential and commercial demand for floorspace equations, we have for the residential side

$$\begin{aligned} H_{Ri} = \theta_i H_i &= \sum_{g,r} \frac{Exp_{igr}}{Q_i} \\ Exp_{igr} &= (1-\beta_{gr})\overline{w}_{igr} L_{igr} \\ &= (1-\beta_{gr})\left(\sum_j \pi_{j|igr} w_{jgr}\right) L_{igr} \end{aligned}$$

and for the commercial side using that the production function is Cobb-Douglas

$$H_{Fi} = \left(\frac{W_j}{\alpha}\right)\left(\frac{1-\alpha}{q_i}\right)\frac{N_j}{S_j} \text{ for } i \in Tracts_j$$

$$\text{where } N_j = \left(\sum_g \alpha_{jg} N_{jg}^{\frac{\sigma^g-1}{\sigma^g}}\right)^{\frac{\sigma^g}{\sigma^g-1}} \text{ and } N_{jg} = \left(\sum_r \alpha_{jgr} L_{Fjgr}^{\frac{\sigma^r-1}{\sigma^r}}\right)^{\frac{\sigma^r}{\sigma^r-1}}$$

As the geographic unit on the workplace side is a POW zone while the geographic unit on the residential side is a census tract, I assume that on the workplace side, labor is supplied evenly across all the tracts within a POW zone.

Finally, θ_i where i is at the tract-level is set to follow

$$\theta_i = \frac{H_{Ri}}{H_i} = \frac{H_{Ri}}{H_{Ri} + H_{Fi}}$$

with housing supply at the tract level $H_i = H_{Ri} + H_{Fi}$.

Step 7 – Given $\{H_i, Q_i\}$ and the parameter μ, I recover the scaled amount of land used for development k_i as a location fundamental following profit maximization of the housing construction sector. Demand for capital can be derived in a straightforward manner as $M_i = Q_i H_i (1-\mu)/p$. Substituting this equation into the housing production function gives

$$H_i = K_i Q_i^{\frac{1-\mu}{\mu}} \left(\frac{1-\mu}{p}\right)^{1-\mu}$$

$$H_i = k_i Q_i^{\frac{1-\mu}{\mu}} \qquad \text{where } k_i = K_i \left(\frac{1-\mu}{p}\right)^{1-\mu}$$

In the quantitative implementation, I allow μ (the capital intensity of housing construction) which determines the housing supply elasticity to price to differ in the suburbs versus the central city following recent work by Baum-Snow and Han (2021) and Saiz (2010). Let $\mu = 0.3$ for neighborhoods within 5 miles of the CBD and $\mu = 0.2$ for neighborhoods further than 5 miles from the CBD, corresponding to floorspace supply elasticities of 2.33 and 4, respectively.

5.3 Endogenous Amenities/Racial Preferences – Instruments for Estimation

There are two endogenous variables $\{\Delta \log CMA_{igr}, \Delta \log(L_{iW}/L_i)\}$ with corresponding instruments. The instrument for CMA changes over time for all specifications comes from CMA where the planned routes or ray network are converted into Interstate highways. For example, CMA with the planned network is defined as $\mathbb{Z}_{igr}^{Plans} = \frac{1}{\phi}\left(\log \sum_j \omega_{jgr,1960}/d_{ijgr}^{Plans\,\phi}\right) - \frac{1}{\phi}\left(\log \sum_j \omega_{jgr,1960}/d_{ijgr,1960}{}^{\phi}\right)$ and $\mathbb{Z}_{igr}^{Rays}$ is the same measure with commute times in the post period from the ray network. The rest of the instruments described below are for racial composition changes.

Hausman Instruments – Instruments following Hausman (1996); Berry et al. (1995) are changes in rental prices and commuter market access (CMA) within the rings of 3-5 miles, 5-10 miles, and 10-15 miles away from each neighborhood. Commute times come from converting the planned routes and ray network into Interstate highways to produce more exogenous variation. For example, CMA changes with the planned routes for neighborhoods 3-5 miles away is denoted $\mathbb{Z}_{igr,3-5}^{Plans} = \{\mathbb{Z}_{sgr}^{Plans}\ \forall s :$

$\text{dist}(i,s) \geq 3,\ \text{dist}(i,s) < 5\}$. Therefore, the set of instruments for racial composition changes $\Delta \log(L_{iw}/L_i)$ are $\{\Delta \log Q_{i,3-5}, \Delta \log Q_{i,5-10}, \Delta \log Q_{i,10-15}, \mathbb{Z}^{Plans}_{igr,3-5}, \mathbb{Z}^{Plans}_{igr,5-10}, \mathbb{Z}^{Plans}_{igr,10-15}\}$ or the corresponding set with rays.

Davis Instruments – Instruments following Davis et al. (2019) come from a 3-step process. The first step requires estimating Equation 6 with all the base and geographic controls and city effects. In addition, racial composition changes are included as a control rather than an endogenous variable of interest. The highway variation is only used for estimating residential elasticity θ_r as the coefficient on $\Delta \log CMA_{igr}$ where the instruments for CMA changes are $\mathbb{Z}^{Plans}_{igr}$ (or $\mathbb{Z}^{Rays}_{igr}$) as well as the CMA Hausman instruments for additional power $\{\mathbb{Z}^{Plans}_{igr,3-5}, \mathbb{Z}^{Plans}_{igr,5-10}, \mathbb{Z}^{Plans}_{igr,10-15}\}$ (or the set with rays). The elasticity estimates are presented in Table A.10 with values from Column 2 entering into the next step. Setting $\theta_N = 0.62, \theta_W = 0.75$ and taking the estimate of local costs from Brinkman and Lin (2022) where $b_{HW} = 0.175, \eta = 1.28$, I solve the quantitative model where endogenous amenities are removed. I simulate the construction of Interstate highways only for segments between 1960 and 1970. This counterfactual predicts racial composition changes under the assumption that other fundamentals of amenities and productivity are unchanged $\Delta \bar{b}_{igr} = 0$, $\Delta a_j = 0$. The prediction for racial composition L'_{iW}/L'_i is used for the calculation of racial composition changes in the instrument $\Delta \log(L_{iW}/L_i)' = \log(L'_{iW}/L'_i) - \log(L_{iW,1960}/L_{i,1960})$. The final set of instruments also includes the CMA Hausman instruments for additional power $\{\mathbb{Z}^{Plans}_{igr,3-5}, \mathbb{Z}^{Plans}_{igr,5-10}, \mathbb{Z}^{Plans}_{igr,10-15}\}$ (or the set with rays) in the last step.

CMA Instruments – Following that there are race-specific responses to CMA, the final set of instruments are CMA for each group separately. Variation in commute times again comes from either the planned routes or ray network for exogeneity. The instruments are $\{\mathbb{Z}^{Plans}_{iLN}, \mathbb{Z}^{Plans}_{iHN}, \mathbb{Z}^{Plans}_{iLW}, \mathbb{Z}^{Plans}_{iHW}\}$ or the corresponding set with the ray network.

6 Welfare and Counterfactuals

6.1 Derivation of Direct Impacts to Welfare

This section derives the approximation of changes in welfare from total differentiating Equation (28) with respect to the two variables that are changing due to the Interstate highway system: commute times t_{ijgr} and amenities B_{igr}. Assuming that commute times only affect commuter access and amenities do not affect any other indirect residential characteristics such as prices, the approximation is then

$$d \log U_{gr} = \sum_{i,j} \frac{\partial \log U_{gr}}{\partial t_{ijgr}} \Delta t_{ijgr} + \sum_{i} \frac{\partial \log U_{gr}}{\partial B_{igr}} \Delta B_{igr}$$

For ease of notation, define the location-specific utility shifter for neighborhoods, ignoring the idiosyncratic shock, as $V_{igr} = B_{igr} CMA_{igr} Q_i^{\beta_{gr}-1}$. Calculating the partial derivative for amenities first, the expression is as follows

$$\begin{aligned}
\frac{\partial \log U_{gr}}{\partial B_{igr}} &= \frac{1}{\theta_r}\frac{\partial V_{igr}^{\theta_r}/\partial B_{igr}}{\sum_s V_{sgr}^{\theta_r}} \\
&= \frac{V_{igr}^{\theta_r-1}}{\sum_s V_{sgr}^{\theta_r}} CMA_{igr} Q_i^{\beta_{gr}-1} \\
&= \pi_{igr}/B_{igr}
\end{aligned}$$

where the last step substitutes in the residential share. A similar first step precedes calculating the partial derivative for commute times.

$$\begin{aligned}
\frac{\partial \log U_{gr}}{\partial t_{ijgr}} &= \frac{1}{\theta_r}\frac{\partial V_{igr}^{\theta_r}/\partial t_{ijgr}}{\sum_s V_{sgr}^{\theta_r}} \\
&= \frac{V_{igr}^{\theta_r-1}}{\sum_s V_{sgr}^{\theta_r}} B_{igr} Q_i^{\beta_{gr}-1} \left(\partial CMA_{igr}/\partial t_{ijgr}\right) \\
&= \frac{V_{igr}^{\theta_r-1}}{\sum_s V_{sgr}^{\theta_r}} B_{igr} Q_i^{\beta_{gr}-1} \left(-\Phi_{igr}^{\frac{1}{\phi}} \frac{T_{jgr}(w_{jgr}/d_{ijgr})^{\phi}}{\Phi_{igr}} \frac{\kappa_{gr}}{t_{ijgr}}\right) \\
&= -\frac{V_{igr}^{\theta_r}}{\sum_s V_{sgr}^{\theta_r}} \frac{T_{jgr}(w_{jgr}/d_{ijgr})^{\phi}}{\Phi_{igr}} \frac{\kappa_{gr}}{t_{ijgr}} \\
&= -\pi_{igr}\pi_{j|igr}\frac{\kappa_{gr}}{t_{ijgr}}
\end{aligned}$$

where the last step substitutes in the residential share and the conditional commuting share. Finally, note that $\Delta B_{igr}/B_{igr} = (-b_{highway}\exp(-\eta d_{i,highway}))$ so the direct impact to welfare is

$$d\log U_{gr} = -\sum_{i,j} \pi_{igr}\pi_{j|igr}\kappa_{gr}\Delta t_{ijgr}/t_{ijgr} - \sum_i \pi_{igr} b_{highway}\exp(-\eta d_{i,highway})$$

6.2 Solving for the Partial Equilibrium Counterfactual

To solve for the model counterfactuals, I employ a combination of observed data on travel times and city-level population $\{t_{ijgr}, \mathbb{L}_{gr}\}$, model parameters $\{\beta_{gr}, \kappa_{gr}, \phi, \theta_r, \mu\}$ with the externality parameter $\{\rho_r\}$, location fundamentals from the inversion process $\{b_{igr}\}$, and other location characteristics inferred during model inversion $\{k_i\}$. Equilibrium objects on the workplace side are fixed to their initial values for $\{w^0_{jgr}, \theta^0_i\}$. I assume starting values for the endogenous variables that correspond to the observed equilibrium for housing prices and the partially endogenous amenities $\{Q^0_i, B^0_{igr}\}$. From these starting values, I iterate following the equilibrium conditions of the model to reach a new equilibrium $\{Q^1_i, B^1_{igr}\}$.

$$\pi^1_{Rigr} = \frac{\left(B^0_{igr} CMA_{igr} (Q^0_i)^{\beta_{gr}-1}\right)^{\theta_r}}{\sum_t \left(B^0_{tgr} CMA_{tgr} (Q^0_{tr})^{\beta_{gr}-1}\right)^{\theta_r}} \quad \text{with } CMA_{igr} = \Phi_{igr}^{\frac{1}{\phi}}$$

$$\Phi_{igr} = \sum_s T_{sgr} (w^0_{sgr}/d_{isgr})^{\phi}$$

$$L^1_{Rigr} = \pi^1_{Rigr} \mathbb{L}_{gr}$$

$$Q^1_i = \left(\frac{Exp_i}{\theta^0_i k_i}\right)^{\mu} \quad \text{with } Exp_i = \sum_{g,r} (1-\beta_{gr}) \left(\sum_j \pi_{j|igr} w^0_{jgr}\right) L^1_{Rigr}$$

$$\text{and } \pi_{j|igr} = \frac{T_{jgr}(w^0_{jgr}/d_{ijgr})^{\phi}}{\sum_s T_{sgr}(w^0_{sgr}/d_{isgr})^{\phi}}$$

$$B^1_{igr} = b_{igr} (L^1_{RiW}/L^1_{Ri})^{\rho^R_r}$$

I continue the iterative procedure until the endogenous variables converge such that

$$\left\| \{Q^0_i, B^0_{igr}\} - \{Q^1_i, B^1_{igr}\} \right\| < \epsilon$$

for some tolerance level ϵ. Before I reach that point, I update the endogenous variables as weighted averages of the initial values and the predicted values with $\lambda \in (0,1)$ following

$$Q^2_i = \lambda Q^1_i + (1-\lambda) Q^0_i$$
$$B^2_{igr} = \lambda B^1_{igr} + (1-\lambda) B^0_{igr}$$

6.3 Solving for the General Equilibrium Counterfactual

To solve for the model counterfactuals, I employ a combination of observed data on travel times and city-level population $\{t_{ijgr}, \mathbb{L}_{gr}\}$, model parameters $\{\beta_{gr}, \kappa_{gr}, \phi, \theta_r, \alpha, \alpha_{jg}, \alpha_{jgr}, \sigma^g, \sigma^r, \mu\}$ with the externality parameters $\{\rho_r, \gamma^A\}$, location fundamentals from the inversion process $\{b_{igr}, a_i\}$, and other location characteristics inferred during model inversion $\{k_i, T_{jgr}\}$. I assume starting values for the endogenous variables that correspond to the observed equilibrium for wages, prices, distribution of rents, floorspace allocation and the partially endogenous amenities and productivity $\{w^0_{jgr}, Q^0_i, \theta^0_i, B^0_{igr}, A^0_i\}$. From these starting values, I iterate following the equilibrium conditions of the model to reach a new equilibrium $\{w^1_{jgr}, Q^1_i, \theta^1_i, B^1_{igr}, A^1_j\}$.

$$\pi^1_{Rigr} = \frac{\left(B^0_{igr} CMA_{igr} (Q^0_i)^{\beta_{gr}-1}\right)^{\theta_r}}{\sum_t \left(B^0_{tgr} CMA_{tgr} (Q^0_{tr})^{\beta_{gr}-1}\right)^{\theta_r}} \quad \text{with } CMA_{igr} = \Phi_{igr}^{\frac{1}{\phi}}$$

$$\Phi_{igr} = \sum_s T_{sgr} (w^0_{sgr}/d_{isgr})^{\phi}$$

$$L^1_{Rigr} = \pi_{igr} \mathbb{L}_{gr}$$

$$\pi^1_{j|igr} = \frac{T_{jgr}(w^0_{jgr}/d_{ijgr})^{\phi}}{\sum_s T_{sgr}(w^0_{sgr}/d_{isgr})^{\phi}}$$

$$L^1_{Fjgr} = \sum_i \pi_{j|igr} L^1_{Rigr}$$

$$Y^1_i = A_i N_i^{\alpha} H_{Fi}^{1-\alpha} \quad \text{with } N_{jg} = \left(\sum_r \alpha_{jgr} (L^1_{Fjgr})^{\frac{\sigma^r-1}{\sigma^r}}\right)^{\frac{\sigma^r}{\sigma^r-1}} \quad N_j = \left(\sum_g \alpha_{jg} N_{jg}^{\frac{\sigma^g-1}{\sigma^g}}\right)^{\frac{\sigma^g}{\sigma^g-1}}$$

$$N_i = N_j/S_j \text{ for } i \in Tracts_j$$

$$H_{Fi} = (1-\theta^0_i) k_i (Q^0_i)^{\frac{1-\mu}{\mu}}$$

$$Q^1_i = \left(\frac{Exp_i + (1-\alpha)Y^1_i}{k_i}\right)^{\mu} \quad \text{with } Exp_i = \sum_{g,r} (1-\beta_{gr}) \left(\sum_j \pi^1_{j|igr} w^0_{jgr}\right) L^1_{Rigr}$$

$$\theta^1_i = \frac{Exp_i}{(Q^1_i)^{\frac{1}{\mu}} k_i}$$

$$w^1_{jgr} = (\alpha_{jgr}\omega_{jg}) \left(\frac{\alpha_{jg} W_j}{\omega_{jg}}\right)^{\frac{\sigma^g}{\sigma^r}} \left(\frac{\alpha Y^1_j}{W_j L^1_{Fjgr}}\right)^{\frac{1}{\sigma^r}} \quad \text{with } \omega_{jg} = \left(\sum_r \alpha_{jgr}^{\sigma^r} (w^0_{jgr})^{1-\sigma^r}\right)^{\frac{1}{1-\sigma^r}}$$

$$W_j = \left(\sum_g \alpha_{jg}^{\sigma^g} \omega_{jg}^{1-\sigma^g}\right)^{\frac{1}{1-\sigma^g}}$$

$$Y^1_j = \sum_{i \in Tracts_j} Y^1_i$$

$$B^1_{igr} = b_{igr} (L^1_{RiW}/L^1_{Ri})^{\rho_r}$$

$$A^1_i = a_i (L^1_{Fj}/K_j)^{\gamma^A} \text{ for } i \in Tracts_j$$

I continue the iterative procedure until the endogenous variables converge such that

$$\left\| \{w_{jgr}^0, Q_i^0, \theta_i^0, B_{igr}^0, A_j^0\} - \{w_{jgr}^1, Q_i^1, \theta_i^1, B_{igr}^1, A_j^1\} \right\| < \epsilon$$

for some tolerance level ϵ. Before I reach that point, I update the endogenous variables as weighted averages of the initial values and the predicted values with $\lambda \in (0,1)$ following

$$\begin{aligned} w_{jgr}^2 &= \lambda w_{jgr}^1 + (1-\lambda) w_{jgr}^0 \\ Q_i^2 &= \lambda Q_i^1 + (1-\lambda) Q_i^0 \\ \theta_i^2 &= \lambda \theta_i^1 + (1-\lambda)\theta_i^0 \\ B_{igr}^2 &= \lambda B_{igr}^1 + (1-\lambda) B_{igr}^0 \\ A_j^2 &= \lambda A_j^1 + (1-\lambda) A_j^0 \end{aligned}$$

6.4 Sufficient Conditions for Uniqueness of Equilibria

First, I rewrite the equilibrium conditions in a form that adheres to the constant elasticity system of Allen et al. (2022) where spillovers are of an exponential form. I further allow the elasticities to differ by the type of the agent.

$$x_{ih} = f_{ijh}(x_j) = \sum_{j \in \mathcal{N}} K_{ijh} \prod_{h' \in \mathcal{H}} x_{jh'}^{\alpha_{ihh'}}$$

In this setting, type $\mathcal{N}$ can be a combination of location $i \in \{1, \ldots, S\}$, education $g \in \{L, H\}$, and race $r \in \{B, W\}$. The set of economic interactions $\mathcal{H}$ include population, prices, amenities, and productivity. Define the city-level constant following $\lambda_{gr} = \mathbb{L}_{gr} U_{gr}^{-\theta_r}$. The equilibrium conditions are the stacked set of equations

$$\begin{aligned} L_{igr} &= \lambda_{gr} \left(B_{igr} Q_i^{\beta_{gr}-1} \Phi_{igr}^{\frac{1}{\phi}} \right)^{\theta_r} \\ L_{Fjgr} &= \lambda_{gr} T_{jgr} w_{jgr}^{\phi} \Phi_{Fjgr} \\ \Phi_{igr} &= \lambda_{gr} \sum_j d_{ijgr}^{-\phi} \frac{L_{Fjgr}}{\Phi_{Fjgr}} \\ \Phi_{Fjgr} &= \lambda_{gr} \sum_i d_{ijgr}^{-\phi} \frac{L_{igr}}{\Phi_{igr}} \\ N_i^{(\sigma_g-1)/\sigma_g} &= \sum_g \alpha_{ig} N_{ig}^{(\sigma_g-1)/\sigma_g} \\ N_{ig}^{(\sigma_r-1)/\sigma_r} &= \sum_r \alpha_{igr} L_{Figr}^{(\sigma_r-1)/\sigma_r} \\ Y_i &= A_i^{1/\alpha} N_i (1-\alpha)^{(1-\alpha)/\alpha} Q_i^{(\alpha-1)/\alpha} \\ \overline{w}_{igr} &= \sum_j T_{jgr} w_{jgr}^{\phi+1} d_{ijgr}^{-\phi} \Phi_{igr}^{-1} \end{aligned}$$

$$
\begin{aligned}
{}_{i}^{\frac{1}{\mu}} &= k_i^{-1}\left(\sum_{g,r}(1-\beta_{gr})\overline{w}_{igr}L_{igr} + (1-\alpha)Y_i\right) \\
w_{jgr} &= \alpha_{jgr}(\omega_{jg})^{\frac{\sigma_r-\sigma_g}{\sigma_r}}\alpha_{jg}^{\frac{\sigma^g}{\sigma^r}}W_j^{\frac{\sigma^g-1}{\sigma^r}}\left(\alpha Y_j\right)^{\frac{1}{\sigma^r}}L_{Fjgr}^{-\frac{1}{\sigma_r}} \\
\omega_{jg}^{1-\sigma^r} &= \sum_r \alpha_{jgr}^{\sigma^r}w_{jgr}^{1-\sigma^r} \\
W_j^{1-\sigma^g} &= \sum_g \alpha_{jg}^{\sigma^g}\omega_{jg}^{1-\sigma^g} \\
L_{iW} &= \sum_g L_{igW} \\
L_i &= \sum_{g,r} L_{igr} \\
B_{igr} &= b_{igr}L_{iW}^{\rho_r}L_i^{-\rho_r} \\
A_j^{1/\gamma^A} &= \frac{a_j^{1/\gamma^A}}{K_j}\sum_{g,r}L_{Fjgr}
\end{aligned}
$$

Almost all of the elasticities $\epsilon_{ijh,jh'}(x_j) = \frac{\partial \log f_{ijh}(x_j)}{\partial \log x_{jh'}}$ are of the form $\epsilon_{ijh,jh'}(x_j) = \alpha_{hh'}$ except for the elasticity to price and the spillovers of racial composition on residential location choices. Let $\beta = \min_{g,r}\{\beta_{gr}\}$ where $\beta_{g,r} > 0$, $\rho = \max_r\{|\rho_r|\}$, and $\theta = \max_r\{\theta_r\}$. Then the $H \times H$ matrix $(\mathbf{A})_{hh'}$ where $H = 16$ is

$$
\begin{bmatrix}
0 & 0 & \theta/\phi & 0 & 0 & 0 & 0 & 0 & \mu(\beta-1)\theta & 0 & 0 & 0 & 0 & 0 & \theta & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & \phi & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{(\sigma_g-1)\sigma_r}{(\sigma_r-1)\sigma_g} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & (\frac{\sigma_r-1}{\sigma_r}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & (\frac{\sigma_g}{\sigma_g-1}) & 0 & 0 & 0 & \mu(\frac{\alpha-1}{\alpha}) & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\gamma^A}{\alpha} \\
0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & \phi+1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -\frac{1}{\sigma_r} & 0 & 0 & 0 & 0 & \frac{1}{\sigma_r} & 0 & 0 & 0 & \frac{\sigma_r-\sigma_g}{\sigma_r(1-\sigma_r)} & -\frac{1}{\sigma_r} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-\sigma_r & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1-\sigma_g}{1-\sigma_r} & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \rho^R & -\rho^R & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

Following Allen et al. (2022) Theorem 1 Part (i), a sufficent condition for uniqueness is that the spectral radius $\rho(A) < 1$. At the current parameter values in Table 8, $\rho(A) > 1$. However, unique equilibria may exist with the listed parameters as the above condition is sufficient but not necessary for uniqueness.

7 Data

Decennial Census Microdata

The Decennial Census is the data source for all of the quantitative estimation. Residential population, workplace population, commute flows, rental prices, and other characteristics of locations come from the Decennial Census microdata in 1960 and 1970. The decade between these two censuses covers a substantial portion of highway construction as by 1960, 20% of the national network was completed, and by 1970, 71% was completed.

- **Residences** – Residential geographic units are census tracts that represent neighborhoods with the usual tract containing 2,500 to 8,000 people. In 1960, there were 42,689 tracts across the United States but not all of the country was contained within tracts. By 1990, the entire country fell under some tract definition, and in 2010, there were 73,175 tracts. Tracts are re-defined across census surveys as population levels across neighborhoods change, so I interpolate all tract-level aggregates to consistent tract definitions with the Longitudinal Tract Database to match 2010 Census delineations (see more below) (Logan et al., 2014). The shapefile for 2010 census tract definitions is retrieved from IPUMS NHGIS.

- **Workplaces** – I construct geographic units which I define as Place of Work (POW) Zones from the Journey to Work questions of the 1960 Census, the first survey in which the Census Bureau asked for location of employment. County and municipality of place of work are reported as 1960-specific Universal Area Codes (UAC), and from these UACs, I calculate the smallest intersection of county and municipality to create the POW Zone. These POW Zones are then overlayed on 1960 tract definitions to create a spatial unit and mapped into 2010 tract boundaries with a crosswalk between 1960 and 2010 tracts. As the UAC for place of work is missing for some observations, I reweight the microdata by calculating inverse probability weights based on observed demographic variables of age, age squared, educational attainment, employment status, total income, wages, industry, occupation, a poverty indicator, race, gender, mode of transport, weeks worked, and a urban/rural indicator. In 1970, place of work is available for UACs, although 1970 UACs are different units from 1960 UACs. For some observations, place of work at the tract level is observed. The inverse probability weights for the 1970 Census are based on whether UAC is observed. For those with tract-level place of work, I assign them to the tract. For those with only UAC, I evenly distribute them across the tracts that are in the UAC. The 1970 tract reweighted sums are then mapped into the 1960 POW zones using a crosswalk between 1970 tracts and 2010 tracts to create a panel of workplace data from 1960 to 1970.

- **Cities** – Cities are represented by Metropolitan Statistical Areas out of the Core-Based Statistical Areas (CBSAs) from 2010 Census definitions. The quantitative analysis requires granular data on commute flows to workplaces from the Decennial microdata in 1960 and 1970. To create the POW Zone, the sample of cities is smaller. While some cities have many unique counties and municipalities, others have very few. For there to be sufficient spatial granularity in place of work, I limit the sample of cities to 25 of the largest, and these cities in total contain 406 POW Zones. I provide the list of cities with available data in Appendix Table G.34. For the motivating empirical analysis, the sample of cities is limited to the 100 cities with Yellow Book maps using public-use tract-level aggregates from NHGIS (see below).

- **Commuting** – With residences as tracts and workplaces as POW zones, commute flows are constructed from population counts over the cross-product of residences and workplaces and are comparable to the widely used Census Transportation Planning Package (CTPP) for commuting after 1990. Starting with the 1980 census, commute times are reported in the Journey to Work section. While individuals may be using non-automobile modes of transport during this time, the lack of data on public transit across a large set of cities makes analysis of other modes difficult. I account for commuting through other methods by assuming public transit systems and walking have not changed in speed from 1960 to 1980 and take reported commute times from the 1980 Decennial Census microdata, the first census survey with commute time data. I non-parametrically estimate non-automobile commute times over 15 bins of Euclidean distance for bilateral pairs of tract of residence and POW Zone and 3 bins of distance from the CBD for both residences and workplaces. The 15 bins of Euclidean distance are fully interacted with the 3 residential bins and 3 workplace bins. Adjusting for distance from the central city captures how car usage is greater when workers live in the suburbs or commute to the suburbs for employment. For each race and education group, I similarly create mode of transport weights over the interaction of bins of Euclidean distance in 1960 and 1970 and bins of distance from the CBD for residences and workplaces. Weighted commute times are averages using the weight for automobile modes (private auto, carpool, van or truck) with the computer generated commuting times for the road network (see below) and the weight for non-automobile modes with the binned commute times from above.

- **Race by Education Tabulations** – To tabulate the population counts, the Census Long-Form person-level sample (25% in 1960 and 15% in 1970) is limited to workers and divided into race and education categories. Person-level sampling weights from the Census are used for all tabulations. Race is divided into White and Non-White as finer splits of race leave too few counts for smaller geographic units. Education is also separated into two categories where those with a high school degree or higher are considered highly-educated and those without a high school degree are considered less-educated. Wages are then calculated for each geographic unit by race and education.

- **Housing Prices** – Housing price data come from the household-level sample with household sampling weights used for all tabulations. Quality-adjusted rents per unit are calculated by taking the rent and residualizing out housing characteristics of the number of rooms, bedrooms, and bathrooms,the availability of a basement, kitchen, heat, hot water, shower/bathtub, indoor toilet, and the year built (setting as the base price the average over the fitted values of housing characteristics for the CBSA and then adding in the neighborhood fixed effects for each neighborhood).

Digitized Roads and Highway Routes

- **Historical Urban Roads** – To capture commuting on the road network prior to Interstate construction, I digitize maps of historical U.S. and state highways and major roads from Shell Atlases in 1951 and 1956 (Rumsey, 2020). To create maps of the historical roads, I start with a highly accurate digital map of modern day major roads from ESRI (2019). I remove Interstate highways and keep major roads less than a freeway, other major roads, and secondary roads as a starting point for the historical map roads. I then georeference the Shell map images in ArcMap and edit the modern day major roads file to match the historical roads maps. I categorize the historical roads into two groups: Superhighways and other major roads following the legend of the Shell Atlases. Maps from 71 cities were digitized as shown in Appendix Table G.34.

- **Yellow Book Plans** – I retrieve maps of the planned routes from the *General Location of National System of Interstate Highways Including All Additional Routes at Urban Areas Designated in September 1955*, commonly known as the Yellow Book, for plans of Interstate highways within cities.

While maps for 100 cities are available, some cities are located within the same CBSA (e.g. Dallas and Fort Worth) and some are Micropolitan Statistical Areas. For these reasons, in Appendix Table G.34, only 96 cities shown. These planned maps were originally used by Brinkman and Lin (2022), and I manually digitized them for this project in ArcMap by georeferencing the map images and creating the spatial lines.

- **National 1947 Plan** – I digitize a map of the 1947 plan of national highway routes from Baum-Snow (2007). This map has less spatial granularity compared to the Yellow Book plans but conveys the direction of routes between cities and which cities the Interstate system was designed to connect. The 1947 plan and Yellow Book maps are consolidated into one planned network.

- **Euclidean Rays** – I construct an additional network of highway routes following the planned routes where I connect cities and towns in the planned maps with straight line rays. This network follows the "inconsequential units" approach where neighborhoods that happen to be located between major cities are treated by the Interstate highway system.

- **Constructed Highways** – The constructed Interstate system comes from MIT Libraries' file of Interstate Highways in 1996 (ESRI, 1996). I exploit the panel variation in when different segments were built by combining this constructed network map with the PR-511 database on dates of construction from Baum-Snow (2007) to examine changes only on routes constructed between 1960 and 1970.

Commuting Networks

- **Speeds** – To calculate commute times on the road networks, I assume speeds for different segments of the routes. For the historical urban roads, large roads (superhighways) are set to have a speed of 40 mph while other major roads are set to have a speed limit of 30 mph following travel surveys conducted during the 1950-1960 period (Gibbons and Proctor, 1954; Walters, 1961). For constructed highways, I use the speed limit for each segment of the highway. The consolidated planned routes of the 1947 plan and Yellow Book maps do not have associated speed limits, so I assign each 2500 meter segment the speed limit of the nearest constructed highway. The Euclidean ray spanning network is set to 60 mph. Minor errors in assignment of speed limits should not affect the results too much given that for urban highways, speed limits cover a narrow range of 55 mph to 65 mph.

- **Commuting Matrices** – For the period prior to highway construction, I calculate commuting times from each 2010 delineated tract centroid to other tract centroids within the same CBSA using ArcNetwork Analyst. The only road network that is traversible are the major roads from the historical road maps. For the period during highway construction, I retrieve the highway network at two stages mid-construction: for all routes built before 1960 and for all routes built before 1970. I overlay these semi-completed highway networks on the historical road network to calculate commuting times during these intermediate periods to align with the years when data is available from the Decennial Census. Using the planned maps and Euclidean ray networks, I construct commute times for the instruments by overlaying the planned and ray networks instead of the Interstate routes on the historical road network. Since there is some distance from tract centroids to the nearest road, and ArcGIS sets the starting point as the point on the traversible network that is closest to the centroid, I add in the additional travel time from the centroid to the road assuming a travel speed of 20 mph. Least cost travel times between tracts are then generated following Dijkstra's algorithm in ArcGIS Network Analyst for 49 million pairwise comparisons. I validate that the computer generated commute times for the fully constructed highway network overlayed on the historical road network are closely correlated with reported commute times by automobile in the 1980 Census (despite possible further road development) in Table G.33. The 1980 Census is the first census survey with commute time data.

Table G.33: Commuting Time Comparison in 1980

Variables	Reported 1980 Commute Time (Minute)
Generated Commute Time (Minute)	0.683***
	(0.0122)
Constant	10.52***
	(0.395)
R-squared	0.537
Correlation Coefficient	0.733
Rounded Obs	11500

Notes: Unit of observation is Place of Work Zone by Place of Work Zone. Data comes from the 1980 Census for survey reported commute times of workers whose mode of transport is private automobile. Computer generated commute times use the full constructed highway network and historical urban roads. Observation counts are rounded to the nearest 500 to meet Census disclosure rules. Robust standard errors are included in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Geographic Features

- **Historical Rail, Canals, Rivers** – Historical railroad networks from 1826-1911, 19th century canals, and 19th century steam-boat navigated rivers are included as controls (Atack, 2015, 2016, 2017).
- **Natural Features** – Distances to natural features including lakes, shores, and ports come from Lee and Lin (2017) and are included as controls.
- **Central Business Districts** – Centroids of the central business districts of MSAs come from Holian and Kahn (2015) although their list does not cover the full list of cities studies in this paper. To obtain the location of other central business districts, I search for where central business districts are in the modern day (assuming most downtowns do not change their location) for cities in Google Earth.
- **HOLC Redlining** – Redlining maps for the Home Owners' Loan Corporation come from a group of digital historians at Mapping Inequality: Redlining in New Deal America (Nelson et al., 2020).
  - **Borders** – To calculate distances from tracts to borders in the HOLC maps for the border discontinuity, I find for each tract the distance to all HOLC map borders. I keep all borders that are within 2 km of the tract centroid. If the tract is redlined, then it has a positive distance from the redlining border. If the tract is non-redlined, the distance is negative.
  - **Redlined/Non-Redlined** – I calculate the percentage of each census tract that is redlined by overlaying the 2010 tract boundaries on the HOLC redlining maps. Tracts that are more

than 80% covered by HOLC grade D areas are considered redlined. The results are not sensitive to the percentage cut-off as 70 percent of tracts are either 100% or 0% graded D. At the 80% cutoff, 3050 out of 13436 tracts are redlined while at a 50% cutoff, 3761 tracts are redlined.

- **School District Borders** – School district boundaries used for the border design are acquired from the National Center for Education Statistics (NCES) for the 1989-1990 school year.
- **Distances to Features** – I calculate the distance from tract centroids to each of the geographic features above. For the POW zones, I take the average of the distances from tract centroids for the tracts within a POW zone.

Natural Amenities

- **Land Cover** – The National Land Cover Database (NLCD) from the U.S. Geological Survey provides data nationwide on land cover types at high spatial resolution (30m). I obtain the dataset for 2011 and limit the characteristics to the land cover types of open water, woody wetlands, developed high-intensity, and deciduous forest. Other land cover types that are available include barren land, cultivated crops, and perennial ice and snow.
- **Tree Canopy Cover** – The U.S. Forest Service Science provides a dataset on Tree Canopy Cover (TCC), and I obtain the 2011 version. It is a 30m spatial resolution file with one variable representing the percentage of canopy cover.
- **Overlap of Tracts with Natural Amenities** – I calculate the overlap between each 30m square from the NLCD and TCC datasets with each tract from the census tracts (with 2010 boundaries) shapefile. A weighted average is computed across the squares that overlap with census tracts.

Air Pollution Index

- **Environmental Pollution** – The Centers for Disease Control and Prevention (CDC) National Environmental Public Health Tracking Network generates air quality measures at the census tract-level using the Environmental Protection Agency (EPA)'s Downscaler model for the mean predicted concentration of PM 2.5. I obtain the 2001-2005 daily estimates and aggregate over the 5 years of data to create a tract-level average.

IPUMS NHGIS Public-Use Aggregates

- I construct a panel of tract-level characteristics from the public-use aggregates available at IPUMS NHGIS (Manson et al., 2017) starting from 1950 and ending in 2010. Aggregates include tract-level population by education, race, income, and housing rents and home values. This dataset is interpolated to be consistent with 2010 tract definitions and spans the full set of cities with planned (Yellow Book) maps in the U.S. The panel is unbalanced however as it was not until 1990 that the Census defined tract geographic units for the entire United States.

Longitudinal Tract Crosswalks

- Tract cross-walk weights derived using population overlaps from the Longitudinal Tract Database are available for 1970 to 2010 from Logan et al. (2014) to harmonize tract-level data across decades to 2010 boundaries. Weights for 1950 and 1960 come from Lee and Lin (2017) and are derived from area overlaps.

Table G.34: Cities and Map/Data Availability

Metropolitan Statistical Area	Yellow Book	HOLC	Historical Roads	Census
Albany-Schenectady-Troy, NY	X	X	X	X
Allentown-Bethlehem-Easton, PA-NJ	X	X	X	X
Atlanta-Sandy Springs-Marietta, GA	X	X	X	X
Baltimore-Towson, MD	X	X	X	
Bangor, ME	X			
Baton Rouge, LA	X		X	
Battle Creek, MI	X	X	X	
Birmingham-Hoover, AL	X	X	X	
Boston-Cambridge-Quincy, MA-NH	X	X	X	X
Buffalo-Niagara Falls, NY	X	X	X	
Burlington-South Burlington, VT	X			
Chattanooga, TN-GA	X	X	X	
Chicago-Joliet-Naperville, IL-IN-WI	X	X	X	X
Cincinnati-Middletown, OH-KY-IN	X	X	X	
Cleveland-Elyria-Mentor, OH	X	X	X	X
Columbia, SC	X	X	X	
Columbus, OH	X	X	X	
Dallas-Fort Worth-Arlington, TX	X	X	X	X
Davenport-Moline-Rock Island, IA-IL	X	X		
Denver-Aurora-Broomfield, CO	X	X		
Des Moines-West Des Moines, IA	X	X		
Detroit-Warren-Livonia, MI	X	X	X	X
Erie, PA	X	X	X	
Eugene-Springfield, OR	X			
Flint, MI	X	X	X	
Fort Smith, AR-OK	X		X	
Gadsden, AL	X		X	
Grand Rapids-Wyoming, MI	X	X	X	
Great Falls, MT	X			
Greenville-Mauldin-Easley, SC	X		X	
Harrisburg-Carlisle, PA	X	X	X	
Hartford-West Hartford-East Hartford, CT	X	X	X	X
Houston-Sugar Land-Baytown, TX	X	X	X	X
Indianapolis-Carmel, IN	X	X	X	
Jackson, MS	X	X	X	
Kansas City, MO-KS	X	X	X	X
Kingsport-Bristol-Bristol, TN-VA	X			
Kingston, NY	X		X	
Knoxville, TN	X	X	X	
Lake Charles, LA	X		X	
Lansing-East Lansing, MI	X	X	X	
Lincoln, NE	X	X	X	
Little Rock-North Little Rock-Conway, AR	X	X	X	
Los Angeles-Long Beach-Santa Ana, CA	X	X	X	X
Louisville/Jefferson County, KY-IN	X	X	X	
Macon, GA	X	X		
Manchester-Nashua, NH	X	X		

Notes: The table displays 96 cities because while there are 100 cities in the Yellow Book, not all of them have an associated Metropolitan Statistical Area. Some constitute Micropolitan Statistical Areas, and two cities (Dallas and Fort Worth) are combined into one MSA. The HOLC redlining maps are available for more cities than those listed, but the table is restricted to the sample of Yellow Book maps. Historical road maps are also available for more cities, but for this paper, only 71 are digitized. The Census column indicates which cities are included in the quantitative analysis using Decennial microdata for estimation.

Table G.34: Cities and Map/Data Availability CONTINUED

Metropolitan Statistical Area	Yellow Book	HOLC	Historical Roads	Census
Memphis, TN-MS-AR	X	X	X	
Miami-Fort Lauderdale-Pompano Beach, FL	X	X	X	
Milwaukee-Waukesha-West Allis, WI	X	X	X	
Minneapolis-St. Paul-Bloomington, MN-WI	X	X	X	X
Monroe, LA	X		X	
Montgomery, AL	X	X	X	
Nashville-Davidson–Murfreesboro–Franklin, TN	X	X	X	
New Orleans-Metairie-Kenner, LA	X	X	X	
New York-New Jersey-Long Island, NY-NJ-PA	X	X	X	X
Oklahoma City, OK	X	X	X	
Omaha-Council Bluffs, NE-IA	X	X	X	
Pensacola-Ferry Pass-Brent, FL	X			
Peoria, IL	X	X	X	
Philadelphia-Camden, PA-NJ-DE-MD	X	X	X	X
Phoenix-Mesa-Glendale, AZ	X	X		
Pittsburgh, PA	X	X	X	X
Pocatello, ID	X			
Portland-South Portland-Biddeford, ME	X			
Portland-Vancouver-Hillsboro, OR-WA	X	X	X	X
Providence-New Bedford-Fall River, RI-MA	X	X	X	X
Rapid City, SD	X			
Reading, PA	X		X	X
Richmond, VA	X	X		
Roanoke, VA	X	X		
Rochester, NY	X	X	X	
Saginaw-Saginaw Township North, MI	X	X	X	
St. Joseph, MO-KS	X	X		
St. Louis, MO-IL	X	X	X	X
Salem, OR	X			
San Antonio-New Braunfels, TX	X	X		
San Francisco-Oakland-Fremont, CA	X	X	X	X
Seattle-Tacoma-Bellevue, WA	X	X		
Shreveport-Bossier City, LA	X	X	X	
Sioux Falls, SD	X			
Spartanburg, SC	X		X	
Springfield, MA	X	X	X	X
Syracuse, NY	X	X	X	
Tampa-St. Petersburg-Clearwater, FL	X	X	X	
Toledo, OH	X	X	X	
Topeka, KS	X	X		
Tucson, AZ	X			
Tulsa, OK	X	X	X	
Tuscaloosa, AL	X		X	
Utica-Rome, NY	X	X	X	
Virginia Beach-Norfolk-Newport News, VA-NC	X	X	X	X
Washington-Arlington, DC-VA-MD-WV	X		X	X
Wheeling, WV-OH	X	X	X	
Wichita, KS	X	X		
Worcester, MA	X		X	X

Notes: The table displays 96 cities because while there are 100 cities in the Yellow Book, not all of them have an associated Metropolitan Statistical Area. Some constitute Micropolitan Statistical Areas, and two cities (Dallas and Fort Worth) are combined into one MSA. The HOLC redlining maps are available for more cities than those listed, but the table is restricted to the sample of Yellow Book maps. Historical road maps are also available for more cities, but for this paper, only 71 are digitized. The Census column indicates which cities are included in the quantitative analysis using Decennial microdata for estimation.

7.1 Iterative Matching Procedure

This paper aims to match children and parents by name following an approach that is similar to the iterative process undertaken by Abramitzky et al. (2012, 2014). It employs machine learning algorithms as in Feigenbaum (2016). However, in addition to their methods, it also includes a variety of string comparison functions besides Jaro-Winkler distance that permits more adjustment for misspellings. I present below the steps of the matching algorithm.

Input Datasets – The two main samples that enter into the matching procedure are children from the Numident and potential parents who file IRS 1040 forms. As described in the Data section, I restrict the full universe of individuals with SSNs to those born between 1964 to 1979 since those cohorts are the likely dependents of parents tax filers in the 1974 and 1979 1040 forms. This linkage then allows researchers to determine the economic status of children during their childhood. These two years of tax data are the earliest ones that cover the whole U.S. Linkages start with the 1964 cohort because in 1974, they are aged 10 and most are still living at home with their parents. For later ages, it becomes harder to link children as they are no longer listed as dependents.

Blocking and Matching Variables – The variables used for comparison are name variables and the coarse geographic variable of state of birth. An additional commonly used variable for linkage is year of birth. However, unlike other procedures that link an *individual to themselves* across multiple datasets that may contain year of birth, in this case, parents are matched to children who do not share year of birth. As the main goal of the matching variables is to restrict to the relevant samples, I approximately obtain adult tax filers in the right age range by including only those with dependent children.

Given that the whole population for several birth cohorts is included in the two input datasets, even with available modern computing power, it would be infeasible to evaluate matches between all children and all parent tax filers. Therefore, matches are compared only within specified blocks that are constructed from variables that must exactly match inside the block. No comparisons are made across blocks. One of the main blocking variables is state of birth. For children from the Numident, I observe their state of birth directly. For parent tax filers, the state of birth of their dependents is not listed. Therefore I assume that they filed in the same state as their child was born in and retrieve the state of tax filing. Only native-born children are included in the sample because state of birth is unavailable for the foreign-born, who would thus not match on the variable for state of tax filing.

Subsequent to the blocks being created, pairwise comparisons are then evaluated on matching variables that do not have to exactly match. Most of the linking occurs through comparing the parent names of the children in the Numident and the names of the primary and secondary tax filers on the 1040 forms. With other economists at the Census, we were able to obtain the names of both parents for every person in the Numident from the SSA in a restricted file. Upon filing an application with the SSA, individuals must include both their own name as well as their parents' names. From the IRS, we were also able to obtain the names of all tax filers, and another source of names for tax filers comes from linking the Numident names to the filers directly. As the mother's last name in the tax filing may be different from the name listed in the Numident as a result of name changes upon marriage, I retrieve the mother's maiden name using the parent names from the SSA.

As names are listed imprecisely, I modify and apply the fuzzy matching techniques of Cuffe and Goldschlag (2018) created for business record linkage to this setting for child-parent name matching. Whether the names are considered a match depends on a variety of string comparison functions that output scores for the level of correspondence between the names.

String Comparison Functions – The most commonly used string similarity measure is Jaro-Winkler distance which depends on the length of the string, the number of characters within some distance apart that are the same, and the number of transpositions that need to occur for characters to be in the same position. The matching algorithm contains several additional string comparison functions which are listed below.

- Jaro distance - The same measure as Jaro-Winkler without the Winkler modification
- Q-Gram - Measure of the number of common q-grams between strings
- Positional Q-grams - Measure of common q-grams accounting also for the position
- Skip-grams - Measure using bi-grams and surrounding context
- Edit (Levenshtein) distance - The number of edits (insertions, deletions, substitutions) needed for one word to become the other
- Damerau-Levenshtein distance - Includes a modification of the Levenshtein distance by including transpositions as operations also
- Bag distance - A cheap distance measure that is weakly smaller than edit distance
- Smith-Waterman distance - Compares segments of all possible lengths and optimizes the similarity measure
- Sequence matcher - Finds the longest contiguous matching subsequence
- Soundex - Phonetic measure based on sound of words
- Longest common substring - Measure based on lengths of common substrings
- Permuted Winkler - Winkler comparator on permutations of words
- Character histograms - Cosine similarity measure of histograms of characters

Machine Learning Algorithm – The linkage algorithm includes the above listed string comparison functions into a machine learning random forest model to flexibly distinguish matches. Names of parents enter into the string similarity measures above, and a vector of scores is created for each pairwise comparison. Large vectors of scores for every possible comparison are then entered into the random forest model after its parameters are estimated off a training dataset of comparisons partitioned into and labeled as matches and non-matches.

The training data is constructed using true children-parent matches from IRS 1040 tax forms in 1994, the first year that tax filings included dependent identifiers. With the dependent PIKS, I then obtain names for their parents listed on the Numident and match them to names of tax filers. Because the source of the names data is the same, the training data would exhibit the same types of mis-spellings as the input data that is be matched later on. Therefore the training set is highly representative of the target data and would accurately inform the model.

Iterative Process – I follow an iterative matching approach similar in style to Abramitzky et al. (2012) and successively relax the comparison criteria in order to obtain a larger number of children-parent linkages. Model training is completed for each round of blocking and matching, so the parameters of the machine learning model are different for each round.

<u>Round 1</u> – Match to both parents. IRS sample requires two tax filers on the 1040 form. Numident sample is limited to children born between 1964 and 1974 for the 1974 IRS form and children born between 1969 and 1979 for the 1979 IRS form.

The blocking variables are:

1. Father first and last initials

2. Mother first and last initials
3. State of birth to state of tax filing

The matching variables are:

1. Father first and last name
2. Mother first and last name

<u>Round 2</u> – Match to mother only. IRS sample requires a single tax filer who is female on the 1040 form. Numident sample is limited to children born between 1964 and 1974 for the 1974 IRS form and children born between 1969 and 1979 for the 1979 IRS form, who additionally were not previously matched.

The blocking variables are:

1. Mother first and last initials
2. State of birth to state of tax filing

The matching variables are:

1. Mother first, middle, and last name

<u>Round 3</u> – Match to father only. IRS sample requires a single tax filer who is male on the 1040 form. Numident sample is limited to children born between 1964 and 1974 for the 1974 IRS form and children born between 1969 and 1979 for the 1979 IRS form, who additionally were not previously matched.

The blocking variables are:

1. Father first and last initials
2. State of birth to state of tax filing

The matching variables are:

1. Father first, middle, and last name