Opportunity in Motion: Equilibrium Effects of Infrastructure on Economic Mobility

Laura Weiwu*

University of California, Berkeley

July 2025

Abstract

This paper studies the aggregate impacts of a prominent place-based policy, the Interstate highway system, on the local economic development of neighborhoods and the intergenerational outcomes of children. I employ a newly linked panel of the near-universe of cohorts born between 1964 and 1979. Greater commuting access from Interstate construction raises average income and employment rates of existing residents. Given the large-scale nature of the policy, targeted census tracts also experience equilibrium effects through inflows of higher SES households. Direct income effects and equilibrium migration effects both have causal benefits for children as estimated using a movers design. However, migration leads to spillovers elsewhere as areas with lower access improvements worsen in peer quality, which adversely affects children in those locations. Within a spatial framework that measures intergenerational impacts, population reallocation and changes in neighborhood composition cancel out for children's outcomes, while direct economic access benefits boost economic mobility.

*laurawww@berkeley.edu. I am deeply grateful to my advisors David Autor, Dave Donaldson, and David Atkin for their dedication and guidance. I want to acknowledge Nate Baum-Snow for sharing data on interstate highway routes and Martha Stinson, Katie Genadek, Todd Gardner, and Jonathan Rothbaum for their invaluable assistance with the Census survey and administrative data. The parent-child linkages were constructed by Martha Stinson and Laura Weiwu as part of the internal Census Bureau project "Measuring Intergenerational Mobility Through Use of Linked IRS Data" and are part of a larger effort to improve Census linking methodology and income measurement. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under grant #1745302 and the Institute for Humane Studies under grant IHS018085. The Census Bureau has ensured appropriate access and use of confidential data and has reviewed these results for disclosure avoidance protection (Project 7521691: CBDRB-FY23-CED006-0011, Project 7521692: CBDRB-FY25-CED006-009).

1 Introduction

The U.S. exhibits large disparities in economic opportunity across its cities and persistent gaps in the long-run outcomes of children by 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 greatly across neighborhoods (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 reallocation of economic benefits.

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 that highways increased access to workplaces for suburban neighborhoods, and these improvements in access raised incomes for existing residents to boost economic opportunity. Second, given the massive scale of the policy, households responded by migrating toward areas with greater commuting cost reductions. I document that 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). For example, Derenoncourt (2022) finds Black migration from the South re-shaped economic mobility in Northern urban areas. In this paper, I quantify the net effects of the migration channel for both the destination locations (e.g. Northern cities) and the *origin* locations (e.g. the rural South) but at the spatial scale of neighborhoods rather than regions.

To assess the importance of the indirect effects, I develop a theoretical framework for measuring how large-scale policies affect long-run consequences of children. One 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, one on spatial economics and the other on intergenerational mobility. Notably, it is sufficiently general for investigating how myriad place-based policies impact long-run outcomes, not only transportation infrastructure.

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 income groups.

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 and Hendren (2018a) provides correlational evidence on how segregation and average income relate to causal outcomes of place, 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 *causal impacts of 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 a 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 Medicaids among other ambitious domestic programs to promote economic opportunity. Modern-day measures, such 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).

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. As migration could contribute to the increase in income, I decompose how much of the change is through effects on existing residents. Using the detailed microdata, I find that 80% of the income gains are unexplained by the changing composition of neighborhoods. However, the reallocation/sorting responses to the interstate highway system are significant. Areas with greater increases in connectivity, which tend to be suburban neighborhoods, experienced inflows of higher-education, higher occupational status, and White 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.

Descriptive correlations between average income, educational composition, and racial composition for neighborhoods (tracts) and the average adult income ranks of children have previously been documented to be strong Chetty et al. (2025). 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 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 higher-educated, higher-occupational status, or more White. 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, estimating place effects for children 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 larger magnitudes for lower income children where higher income and better peers lead to improved outcomes later in life. 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 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 (2025).

In the final section of the paper, I combine the previously estimated statistics to assess the aggregate consequences of interstate highways on intergenerational mobility. 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 through its two channels. I find that population reallocation and changes in neighborhood composition cancel out for children's outcomes. Direct economic access benefits raise intergenerational mobility as lower income children benefit more from living in higher-income neighborhoods.

Related Literature – This paper is 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 intergenerational 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 provides historical background on the interstate highway system. Section 4 characterizes the theoretical framework. Section 5 presents the empirical evidence on highway impacts. Section 6 measures how place characteristics affect economic mobility. Section 7 conducts counterfactual analyses. Section 8 concludes with policy implications.

2 Historical Data on Intergenerational Income Mobility

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 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 C.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.[1] 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.1 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 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.1. 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

[1] 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 C.1 for a detailed description.

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 noted in Haider and Solon (2006) and Nybom and Stuhler (2016).

When studying intergenerational dynamics, researchers often use household income to represent parental economic resources available to children. 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 1, 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 2. 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.2. 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.2, I test whether group means are statistically different while including birth year and race fixed effects. While the difference between

Table 1: Summary Statistics on Economic Mobility

	(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.

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 1, 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.2 Column (2) to Table 2 Column (1) for county one-time movers and Table 2 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.

2.1 Place-Level Variation in Children's 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.

Variation Across Counties – Children often move across counties and tracts during their childhood, as was summarized previously in Table 2. 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 per-

Table 2: 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.

centile of the parental income distribution by estimating Equation (**??**) for each county denoted c and pooling all cohorts born between 1964 and 1979.

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

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 3.[2] 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 (1) 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 3, 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

[2]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.

Table 3: 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.

around 0.18 which can arise from true race-specific heterogeneity across 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 point of interest in the estimation of causal impacts of place.

3 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.

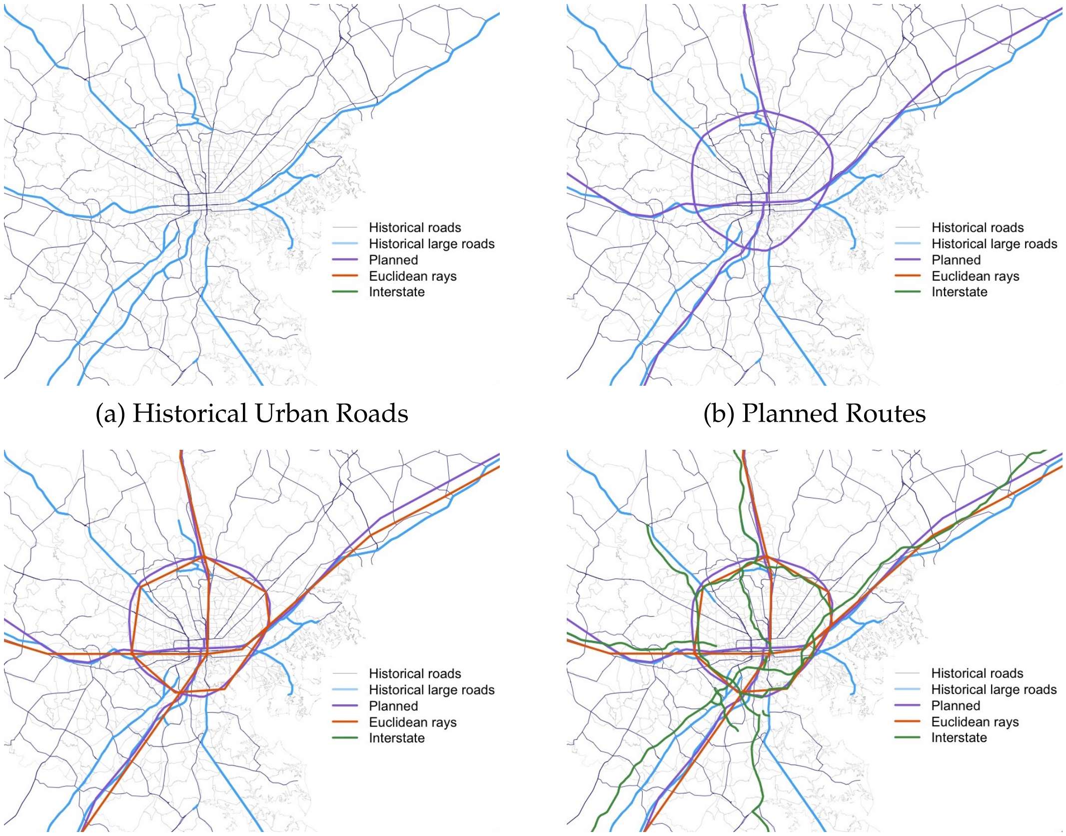
Brief History – When the construction of the interstate network began, suburbanization into peripheral neighborhoods was already well underway. The expansion of the existing road network with high-speed limited access freeways further precipitated migration away from central areas (Jackson, 1985). With 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 (**?**). 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

Figure 1: Historical Roads and Highways for the Baltimore Metro 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 1a compared to Panel 1d, 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.

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 come 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 Baltimore is depicted in Figure 1. 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.

4 Framework for Mapping Policy into Economic Mobility

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.

4.1 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.

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

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

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

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 5, I provide empirical evidence on (1) and (2) in response to the interstate development. In Section 6, I estimate (3) using a movers design for families that move across origins and destinations with different characteristics.

4.2 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.[3]

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

$$u_{nmg}(i) = \frac{z_n(i)\epsilon_m(i)B_{ng}Q_n^{\beta_g - 1}w_{mg}}{d_{nmg}} \tag{3}$$

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.

[3] 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.

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

$$\pi_{mg|n} = \frac{T_{mg}(w_{mg}/d_{nmg})^{\phi}}{\sum_l T_{lg}(w_{lg}/d_{nlg})^{\phi}} = \frac{T_{mg}(w_{mg}/d_{nmg})^{\phi}}{\Phi_{ng}} \tag{4}$$

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 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.

$$\pi_{ng} = \frac{(B_{ng}CMA_{ng}Q_n^{\beta_g-1})^{\theta_g}}{\sum_t(B_{tg}CMA_{tg}Q_t^{\beta_g-1})^{\theta_g}} \tag{5}$$

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 $\mathbf{L}_g$.

$$L_{ng} = \pi_{ng}\mathbb{L}_g \tag{6}$$

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.

$$H_n = \left(\frac{1-\mu}{\mu}\right)^{\frac{1-\mu}{\mu}} K_n Q_n^{\frac{1-\mu}{\mu}} \tag{7}$$

$$Q_n H_n = \sum_g (1-\beta_g)\bar{w}_{ng}L_{ng} \tag{8}$$

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.

$$U_g = \left[\sum_n \left(B_{ng}CMA_{ng}Q_n^{\beta_g-1}\right)^{\theta_g}\right]^{1/\theta_g} = \left[\sum_n \left(B_{ng}\left(\sum_m T_{mg}(w_{mg}/d_{nmg})^{\phi}\right)^{1/\phi} Q_n^{\beta_g-1}\right)^{\theta_g}\right]^{1/\theta_g} \tag{9}$$

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 (6)
2. *Housing supply and demand* (8)
3. *Closed City* where $\sum_i L_{ng} = \mathbb{L}_g$

4.3 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 (5), 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 (4), neighborhood average income aggregates across workplaces and the wages received in those locations.

$$\bar{w}_{ng} = E[w_{mg}|n] = \sum_m \pi_{mg|n} w_{mg} = \sum_m \frac{T_{mg}(w_{mg}/d_{nmg})^\phi}{\sum_s T_{sg}(w_{sg}/d_{nsg})^\phi} w_{mg} \tag{10}$$

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.

5 Empirical Evidence of Highway Impacts on Neighborhoods

5.1 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.5, 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).

5.2 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 $m \in \{1, \dots, M\}$ 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_n^{YB} and JMA_n^{Rays} 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).

5.3 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.

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

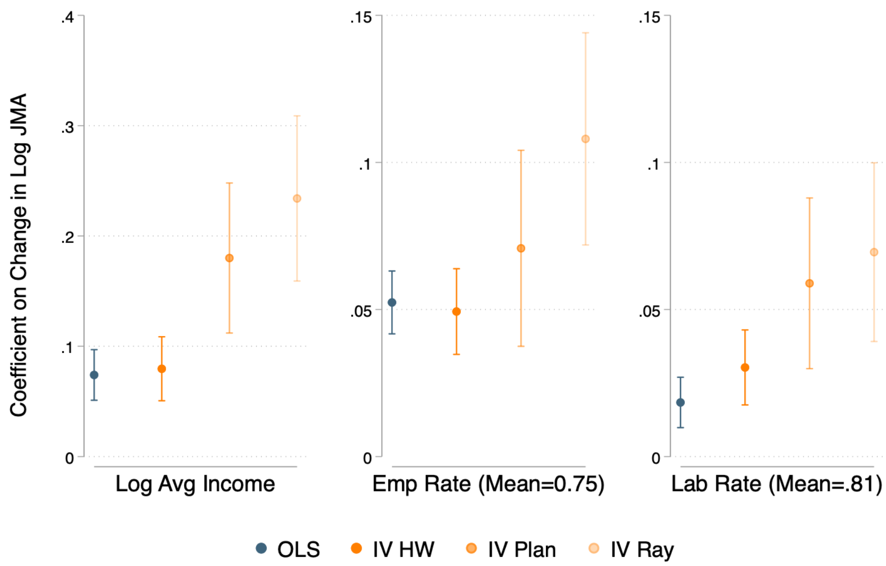
The geographic controls in X_n account for the non-random placement of the interstate highway system and are distance to the central business district, large historical urban roads, rivers, lakes, shores, ports, historical railroads and canals. With fixed effects at the city-level, the empirical variation is only across tracts within metropolitan areas. I present OLS results in Table A.6 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.7 and plotted in Figure 2 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.

Figure 2: 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.7

5.4 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 5 and 6, I obtain the estimation equation below.

$$\Delta \log L_{ng} = \alpha_g + \theta_g \Delta \log JMA_{ng} + \mathbf{X}_n \eta + \epsilon_{ng} \tag{11}$$

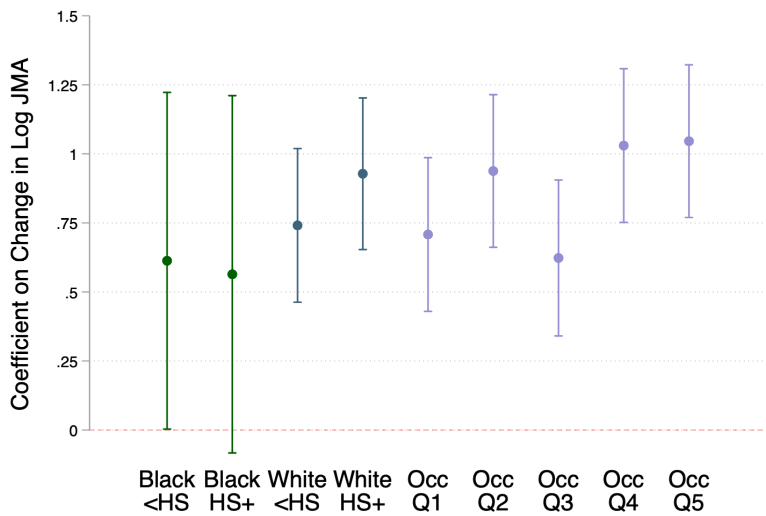
Additionally, the specification above includes controls in $\mathbf{X}_n$ for the previously discussed geographic features as well as city fixed effects. The first difference is over 1960 and 1970 using Decennial restricted data as with the earlier estimating equation.

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 3. 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.

Figure 3: 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.

5.5 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 (2), 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.

6 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.

6.1 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.2, 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. [4]

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 responses to interstate highways, respond more to differences across neighborhoods. The higher spatial mobility suggests they select more 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

[4] Additional results at the county-level are also available in Appendix Figure B.3. 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).

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.

6.2 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, which complicates 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 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.[5] These county-level average predicted income ranks serve as a measure 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.

$$y_i = \sum_{s=1964}^{1979} I\{s_i = s\}(\lambda_s^1 + \lambda_s^2 \bar{y}_{por}) + \sum_{m=1}^{28} I\{m_i = m\}\phi_m + \sum_{m=1}^{28} b_m I\{m_i = m\}\Delta_{odpr} + \epsilon_{1i} \tag{12}$$

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

$$\begin{aligned} y_i = &\sum_{s=1964}^{1979} I\{s_i = s\}(\lambda_s^1 + \lambda_s^2 \bar{y}_{por}) + \sum_{m=1}^{28} I\{m_i = m\}\phi_m \\ &+ I\{m_i \leq 23\}(\rho + (23 - m_i)\gamma)\Delta_{odpr} + I\{m_i > 23\}(\delta^1 + (23 - m_i)\delta^2)\Delta_{odpr} + \epsilon_{2i} \end{aligned} \tag{13}$$

where as in the above specification, age at move fixed effects and cohort fixed effects interacted with origin predicted income rank are included.[6] 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 esti-

[5]These predicted child income ranks do not include one-time movers to ensure that a child's own outcome does not enter the definition of neighborhood quality. These exposure-weighted income ranks are estimated following Equation 1. The exposure weights that are used for predicting income rank are constructed from residential location up until age 23.

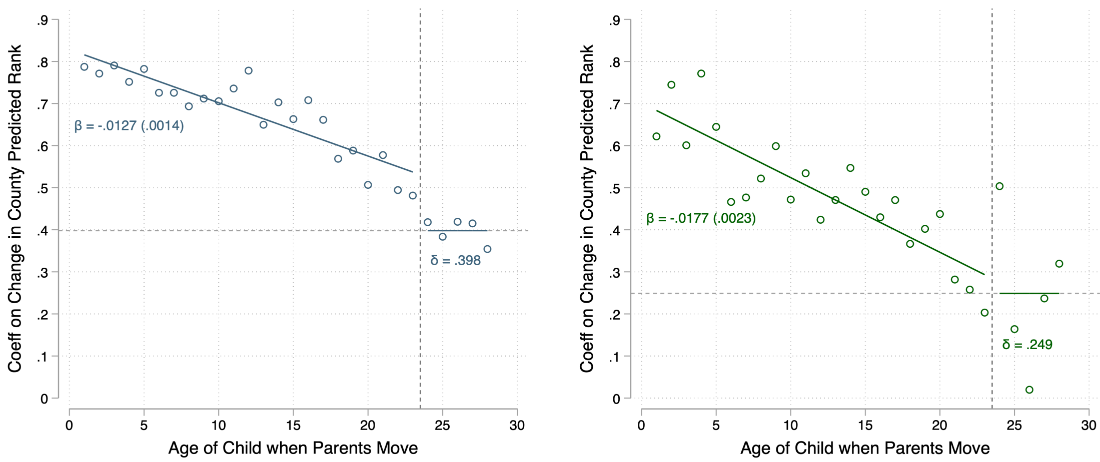
[6]I further include origin and destination fixed effects and family fixed effects to test additional selection in robustness checks.

mation. 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 4 for White and Black boys where the age at move coefficients are presented in the scatter plot. In Figure 4, 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 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.

Figure 4: 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.

Estimates for the linear parametric specification are presented in Table 4 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.

6.3 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 neigh-

Table 4: Movers Exposure Effects By Race and Gender

	(1)	(2)	(3)	(4)
	White		Black	
Variables	Male	Female	Male	Female
≤ 23 Exposure Slope	0.0128	0.0104	0.0171	0.0116
	(0.0017)	(0.0016)	(0.0029)	(0.0026)
≤ 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.

borhood 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 2) for the 40,000 tracts in the U.S.

The specification I estimate for the extension is similar in form to Equation 13. 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 5 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.[7] 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 5, 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.

7 Aggregate Effects of Highways on Children's Outcomes

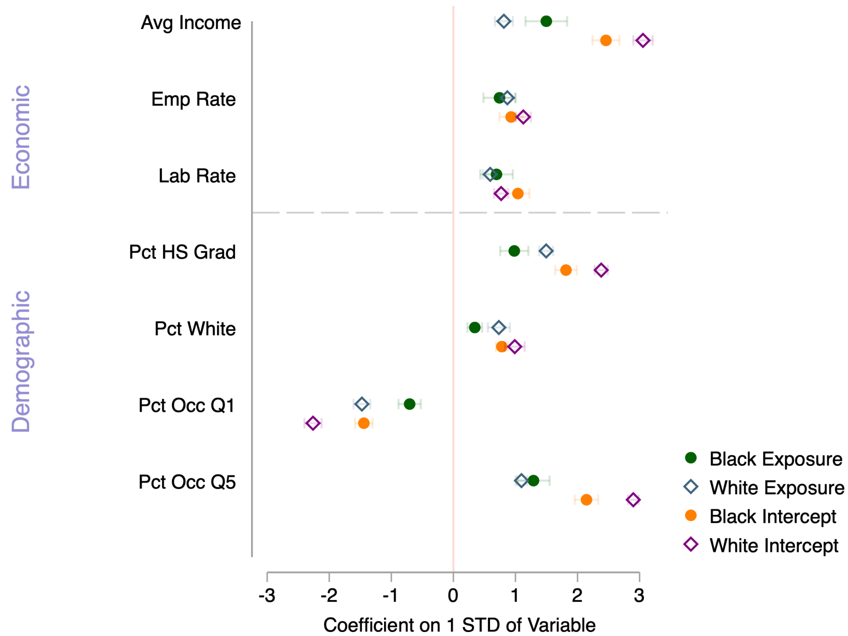
In future work, I will solve for the full predicted migration responses and change in neighborhood characteristics from the equilibrium model in response to interstate development. Returning to the equation for aggregate consequences on children's outcomes

$$\frac{d\bar{y}_g}{d\delta} = \sum_{n=1}^{N} \underbrace{\frac{d\pi_{ng}}{d\delta}}_{(1)} \underbrace{(\mathbf{x}_n \beta_g)}_{(3)} + \pi_{ng} \sum_{k=1}^{K} \underbrace{\frac{d\mathbf{x}_{n,k}}{d\delta}}_{(2)} \tag{14}$$

each piece for the change in aggregate income is now defined. In Equation 5, 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 11 then translates into changes in peer composition in the term (2). Average income is predicted by the model in Equation 10 and is another characteristic of locations altered by the interstate highway

[7] Additional results at the county-level are provided in Figure B.4.

Figure 5: 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.

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.

8 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.

References

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

Chetty, R., Friedman, J., Hendren, N., Jones, M. R., and Porter, S. The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility. 2025.

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.

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.

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 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.

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.

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.

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.

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.

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.

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.

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.

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.

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

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

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.

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.

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.

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.

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).

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.

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.

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.

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

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.

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.

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.

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

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

Weiwu, L. Unequal Access: Racial Segregation and the Distributional Impacts of Interstate Highways in Citie. 2025.

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.

Appendices

A Tables

Table A.1: 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.2: 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.3: 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.4: 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.5: 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.6: 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.7: 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. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

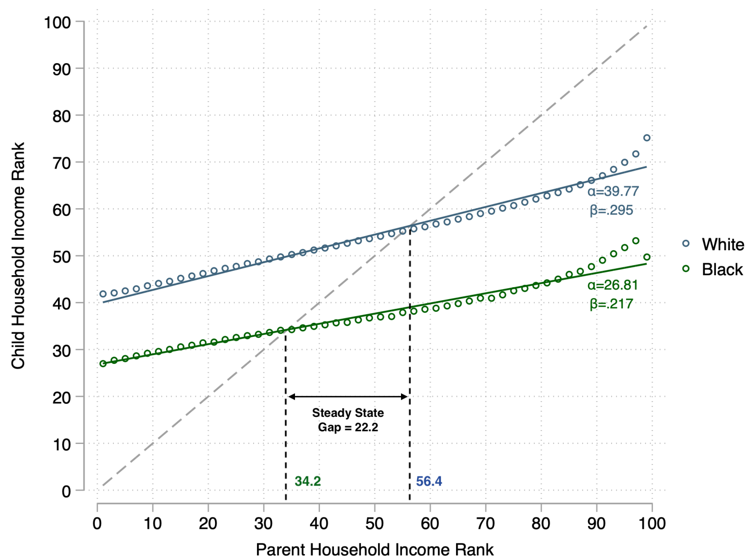
Table A.8: Parent Movers Panel - Two-way FE Income Changes in JMA

	(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.

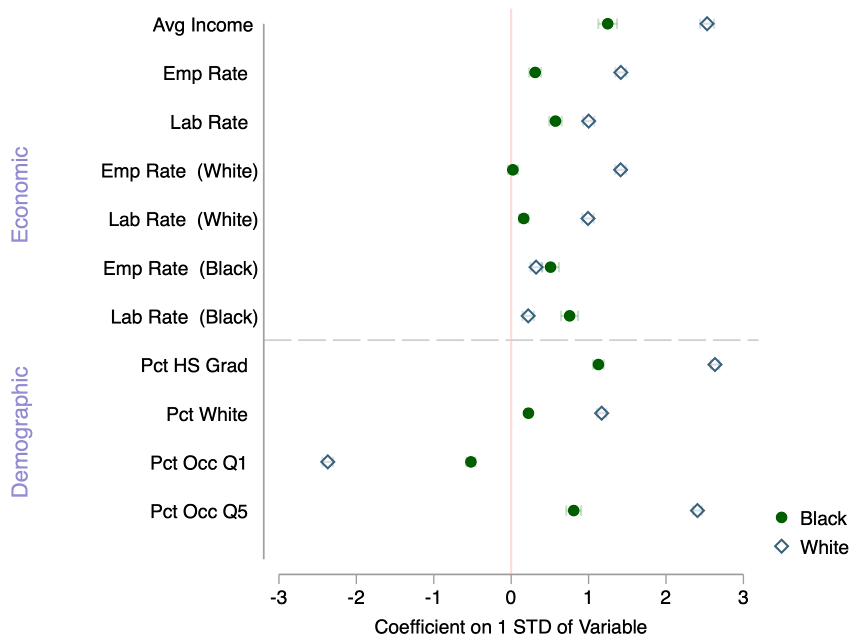
B Figures

Figure B.1: Child Household Income Rank by Race - Steady State



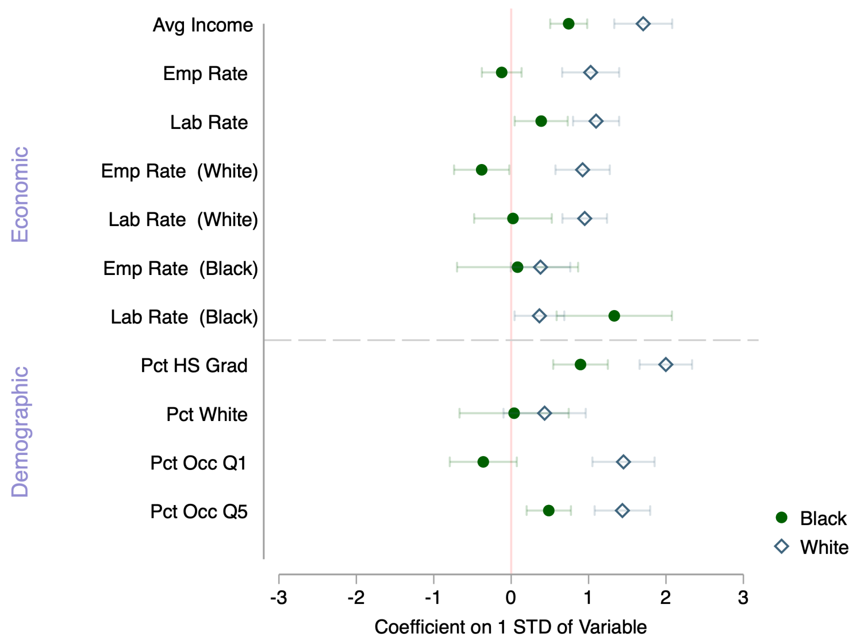
Note: Average child household income rank is measured over 50 bins across parental household income rank separately for White and Black individuals using the 1040 tax data. Parental household income comes from the 1040 forms and is the average of the first four years of tax data available post-birth 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 **??**. 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.2: Descriptive Correlations between Predicted Child Income Rank (P25) and Tract-Level Characteristics



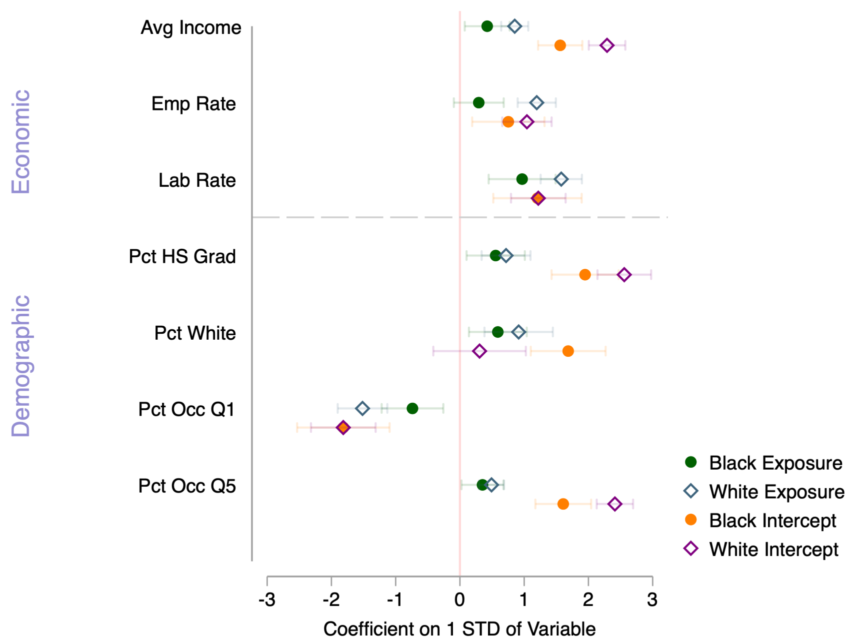
Note: Predicted income rank is computed by estimating linear rank-rank correlations for each racial group in each tract and then predicting the rank of children from the 25th percentile of the parent income distribution. Tract characteristics are calculated using the Decennial Census in 1970. Percentage 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.3: 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.4: 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.

C Data

C.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.

Round 1 – 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

Round 2 – 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

Round 3 – 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