Opportunity in Motion: Infrastructure, Job Access, and Intergenerational Mobility

Laura Weiwu*

University of California, Berkeley

March 2026

Abstract

This paper studies how transportation infrastructure improves job access and, in turn, shapes neighborhood economic development and the intergenerational outcomes of children. I use a newly linked administrative dataset covering the near-universe of families with children born between 1964 and 1979. Increased commuting access from Interstate highway construction raises average neighborhood income through both direct income gains for existing residents and compositional changes driven by selective migration. Using a movers design, I show that increases in neighborhood income benefit children and reduce intergenerational inequality, with larger gains for children from lower-income families. However, migration responses generate negative spillovers: areas with smaller access improvements experience declines in peer quality, adversely affecting children in those locations. I develop an accounting framework to quantify the aggregate effects on intergenerational mobility and find that the direct economic benefits of improved access outweigh these negative spillovers. Nevertheless, changes in neighborhood composition increase the spatial inequality of opportunity.

*lauraww@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 cities and persistent gaps in the long-run outcomes of children by socioeconomic background (Chetty et al., 2020). Place is widely viewed as a key determinant of intergenerational inequality, as employment access, educational quality, and peer environments vary substantially across locations (Wilson, 1987; Reardon and Bischoff, 2011). This raises a central question: can place-based policies improve local conditions and reduce these disparities? Existing work primarily studies shocks to local labor demand (e.g., Baum-Snow et al. (2019); Chetty et al. (2025)), making it difficult to assess how policy can alter opportunity across locations. This paper studies how a major place-based policy—the construction of the Interstate highway system—reshapes intergenerational mobility through changes in job access.

Transportation infrastructure is one of the most pervasive and durable forms of place-based policy. Governments have long used roads and railways to stimulate local development and integrate lagging areas into broader labor markets, from highways built under the Appalachian Regional Commission to infrastructure investments such as China's Belt and Road Initiative.

I focus on *two channels* through which transportation infrastructure shapes intergenerational opportunity. First, highways increase access to workplaces, particularly for suburban neighborhoods, raising the incomes of existing residents. Classic work by Kain (1968) and Wilson (1987) emphasizes spatial mismatch as a driver of inequality, yet correlational evidence from Chetty and Hendren (2018) finds no cross-sectional relationship between job access and mobility.[1] These findings do not rule out that policies that *increase* commuting access can improve long-run outcomes. Leveraging the natural experiment of Interstate construction and new intergenerational linkages, this paper provides such evidence.

Second, place-based policies may generate trade-offs across locations through migration and reallocation. The large-scale expansion of the Interstate system induced households, particularly more-educated, higher occupational status, and White families, to move toward areas with greater commuting improvements. As a result, central neighborhoods lost higher-SES residents and experienced declines in peer quality. These migration responses, which are part of broader suburbanization trends, generate general equilibrium effects that reshape the distribution of opportunity across space.

While infrastructure generates local gains, it can also create spillovers that impose costs elsewhere (Glaeser and Gottlieb, 2008; Kline and Moretti, 2014). Much of the existing literature studies these spillovers through agglomeration and their effects on productivity and local economic activity (Duranton and Puga, 2004; Greenstone et al., 2010). In contrast, this paper shifts the focus

[1] Relatedly, Card et al. (2024) finds that geographic proximity to jobs does not differ by race.

to intergenerational mobility and highlights sorting by socioeconomic status as a key source of externalities for children's outcomes (Massey and Denton, 1993; Sharkey, 2008; Diamond, 2016; Chetty and Hendren, 2018; Fajgelbaum and Gaubert, 2020). Closely related, Derenoncourt (2022) shows how Black migration reshaped economic mobility in Northern urban areas. Building on this work, I quantify the net effects of migration for both destination and *origin* locations, allowing me to assess how place-based policies redistribute opportunity across space.

Notably, the framework is sufficiently general to study how a wide range of place-based policies affect long-run outcomes. Because such policies create both winners and losers, it also enables an assessment of whether, in *aggregate*, children's outcomes improved—or whether gains in some locations were offset by losses in others—and whether these gains were distributed equally across low- and high-income children.

To quantify the net impact of place-based policies on children's long-run outcomes, I estimate how neighborhood characteristics respond to the Interstate highway system and how these changes translate into children's outcomes. I focus on changes in average income and peer composition, where 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 neighborhood characteristics but also shapes which children are exposed to these characteristics. As families migrate outward, they become less exposed to central city conditions and more exposed to suburban environments. A key empirical object is therefore the extent of residential re-sorting across locations by different groups (low- and high-income) of children, as these migration responses can amplify inequality if higher-income households suburbanize more and the suburbs offer better opportunities.

To translate these place characteristics into children's outcomes, I estimate the treatment effects of exposure to higher average neighborhood income and higher status peers for different groups of children (e.g., by income background). I measure long-run outcomes using 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, achieving a match rate of 67%. These newly linked cohorts fill a gap for large-scale measures of intergenerational mobility: modern-day measures, such as in Chetty et al. (2025), begin with cohorts in 1978, while earlier measures using full-count Censuses end with cohorts in the 1950 Census (Abramitzky et al., 2014).

I represent the changes from the Interstate highway shock as an increase in Commuting Market Access (CMA), which is micro-founded using quantitative spatial frameworks from Ahlfeldt et al. (2015), Donaldson and Hornbeck (2016), and Tsivanidis (2022). CMA is a neighborhood-level measure that aggregates commute costs across all workplaces and the employment-weighted wages at those workplaces. It is calculated using microdata from the Decennial Censuses on the Journey to Work from 1960 onwards and changes in travel time derived from digitized historical road maps, datasets constructed and previously used in Weiwu (2025). For the scope of this paper, all effects of Interstate development are routed through CMA. To eliminate local costs such as pollution or displacement, I drop areas immediately by highways from the sample and control for distance to highways.[2]

Using changes in CMA at the tract level, I provide empirical evidence on the impacts of Interstate highways on neighborhood characteristics, leveraging quasi-random placement through an instrumental variables strategy. With planned maps that I digitized for 100 cities and straight-line rays that connect intermediate cities along planned routes, I instrument for highway location, following Baum-Snow (2007). I show that average neighborhood income rises significantly with increases in commuting access. To account for the role of migration and sorting, I decompose the sources of this income growth and find that 80% of the increase is not explained by compositional changes, but instead reflects gains among existing residents.

However, sorting responses to the Interstate highway system are substantial. Areas with greater increases in connectivity, typically suburban neighborhoods, experience inflows of higher-education, higher occupational status, and White households, with corresponding outflows from central neighborhoods. These mobility responses mechanically alter peer composition across locations. Splitting neighborhoods into those with above- and below-median changes in CMA, I find that areas with high CMA shocks experience a 0.15 SD increase in the share of top-quintile income peers, while those with low CMA shocks experience a 0.12 SD decrease.

Given these changes in neighborhood characteristics, I examine how they translate into children's outcomes by estimating the treatment effects of exposure. Descriptive correlations between neighborhood characteristics and adult outcomes may reflect selection and omitted variable bias. To address selection, I extend the movers design of Chetty and Hendren (2018), which exploits differences in children's age at move to generate quasi-random variation in exposure. While Chetty and Hendren (2018) estimate place effects at broader geographic units, applying this approach at the tract level is infeasible given limited observations per location. I therefore adapt the design to focus directly on neighborhood characteristics, measuring how adult outcomes vary with exposure to tracts with higher average income or higher socioeconomic status peers. By concentrating variation along these characteristics rather than estimating place effects for each tract, this approach reduces dimensionality and enables tract-level analysis.

To address omitted variables bias, I instrument neighborhood characteristics using the CMA shock from Interstate development. As documented above, commuting improvements raise neigh-

[2]Recent research by Brinkman and Lin (2022), Weiwu (2025), and Valenzuela-Casasempere (2025) studies the disamenities and displacement caused by Interstate routes.

borhood income, providing a first stage. To generate additional variation, I construct group-specific CMA measures that exploit heterogeneity in workplaces across top- versus bottom-quintile occupations. Differential migration responses to these group-specific shocks provide additional identifying variation for peer composition. I implement this IV strategy using a control function approach following Wooldridge (2015).

Using this design, I find that neighborhoods with higher average income and better peers improve children's outcomes, with larger effects for lower-income children. For example, a one standard deviation increase in the share of top-quintile households raises children's income ranks by 0.5–1.5. These effects are about one-third the size of the corresponding descriptive correlations between neighborhood characteristics and outcomes, implying that selection accounts for roughly two-thirds of the observed associations.

In the final section, I combine these estimates to assess the aggregate consequences of Interstate highways on intergenerational mobility. Direct improvements in job access raise mobility, particularly for lower-income children. However, migration-induced changes in peer composition increase spatial inequality of opportunity, as children in central neighborhoods are adversely affected by the out-migration of higher-SES families. In aggregate, these peer composition effects largely offset one another across locations, while the direct economic benefits remain positive on net.

Taken together, the results show that transportation infrastructure can substantially improve economic opportunity by increasing access to jobs and raising the incomes of existing residents, with particularly large benefits for children from lower-income families. At the same time, these gains are accompanied by migration responses that reshape neighborhood composition and generate offsetting spillovers across locations. While the direct effects of improved commuting access increase intergenerational mobility on average, the resulting sorting of households amplifies spatial inequality of opportunity. These findings highlight that place-based policies can both expand opportunity and redistribute it across space, underscoring the importance of accounting for equilibrium responses when evaluating their aggregate impacts.

Related Literature – This paper relates to a body of work on the long-run impacts of transportation infrastructure. Prior studies examine how roads affect migration (Black et al., 2015) and local labor market opportunities (Duranton and Turner, 2012; Adukia et al., 2020; Asher and Novosad, 2020), but relatively little work studies intergenerational impacts. Exceptions include Costas-Fernandez et al. (2023), which examines how 19th-century railroads in England affected occupational mobility, and Heath Milsom (2023), which shows how market access in West Africa improves educational mobility. This paper instead studies a U.S. setting where commuting access from highway construction affects long-run income mobility and highlights a distinct *mechanism*: improved commuting enables better job matching and wage gains without requiring residential relocation. I further incorporate the role of migration and sorting, decomposing the total impact into direct effects from job access and indirect equilibrium effects from changes in neighborhood composition.

This paper is also connected to a large literature on the geographic determinants of children's outcomes. Kain (1968), Wilson (1987), and Haltiwanger et al. (2020) emphasize spatial mismatch as a barrier to economic opportunity for low-income and minority families. I show that reducing commuting distance improves outcomes for both adults and their children. Related work documents the adverse effects of concentrated poverty and segregation (Massey and Denton, 1993; Sampson et al., 2002; Sharkey, 2008; Andrews et al., 2017; Chyn, 2018), while Chetty and Hendren (2018); Chetty et al. (2020) map the geography of opportunity using administrative data. I contribute by showing that opportunity is not fixed across space but evolves in response to place-based policies, implying that policymakers can change mobility by targeting places rather than relocating families.

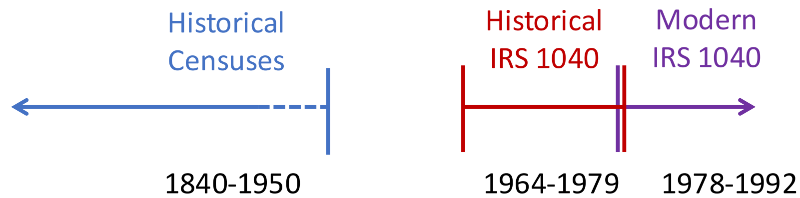
Finally, this paper builds on the quantitative spatial economics literature (Allen and Arkolakis, 2014; Ahlfeldt et al., 2015; Tsivanidis, 2022). Busso et al. (2013) and Diamond and McQuade (2018) study how local interventions interact with mobility and housing markets, while Gaubert et al. (2021) analyzes optimal place-based policy design under sorting. I similarly emphasize equilibrium population responses, but focus on intergenerational mobility. In this context, migration and sorting are unintended consequences of place-based policies that reshape the distribution of opportunity across locations. Closely related, Chyn and Daruich (2022) and Fogli et al. (2025) develop macro frameworks to study policies such as Moving to Opportunity (MTO), calibrated to cross-sectional estimates from Chetty et al. (2025). In contrast, I leverage panel variation from a natural experiment to estimate causal spillover effects and evaluate the realized impacts of an implemented place-based policy.

2 Historical Data on Intergenerational Income Mobility

To measure intergenerational mobility for the mid-20th century, I use a new panel dataset of children born in the years of 1964 to 1979 constructed as part of a joint effort with the Census Bureau, described in 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 – We focus on children in the cohorts of 1964 to 1979, which as shown in Figure 1, previously occupied a gap in large-scale measures of intergenerational mobility. Administrative parent-child linkages used by Chetty et al. (2014) come from the 1994 IRS tax form, which records children born up to 16 years earlier in the cohort of 1978. Earlier measures from the historical Decennial Censuses end in 1950. New estimates of intergenerational mobility for these cohorts, by race and gender, are included in

Figure 1: Timeline of Intergenerational Mobility Measures



Note: Historical Censuses refers to the person-level full count Decennial Census surveys from IPUMS for 1850-1950 (including children born as early as 1840). Earlier household-level data are also available. The modern 1040 linkages used in Chetty et al. (2025) refer to the children dependents recorded in the 1994 tax form, which includes children born as early as 1978 who are aged 16 in 1994.

Appendix Tables A.1 and A.2.

The sample is constructed using 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 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.[3] 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. This parallelization reduces the computation time from 2 years to 2 weeks.

[3]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.

Match Rates – With these linkages, I calculate match rates listed by year of birth in Table A.3 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. The names comes 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 the corresponding 1994 IRS 1040 form. The flexibility of the random forest model further captures additional matches. Lastly, name matching previously used only one first and last name, leading to many non-unique names (e.g. John Smith) 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.

Match rates by gender and race are displayed at the bottom of Table A.3. Rates are essentially the same across men and women because matching on parent names addresses the complication of name changes upon marriage for women. As in other studies, it is challenging to attain match rates for the Black population that are as high as that for the White population due to their lower coverage in survey and administrative sources. While the match rate for the White population is exceptionally high at 72%, the match rate for the Black population of 60% is still notable, reaching the highest match rates in other datasets for the White population.

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

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.[4] For individual income (to isolate the role of marriage), data comes from W-2 earnings records.[5]

[4]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).

[5]For the 1970 to 1979 cohorts, average individual income is calculated 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

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.[6] Hispanic is separated out from White and Black throughout. Summary statistics by race are provided in Appendix Table A.4.

Geographic Variables – Moves are observable in the 1040 forms at detailed geographies through the address of filing variable.[7] 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 (2018). Geographic variables are available at the tract-level for the large majority, as shown in Table A.6. 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 tract-level. However, comparing Appendix Table A.5 Column (2) to Appendix Table A.6 Column (1) for tract one-time movers, I find that children whose location is observed tend to be of higher economic status. The movers design then provides a LATE on this selected sample.

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.

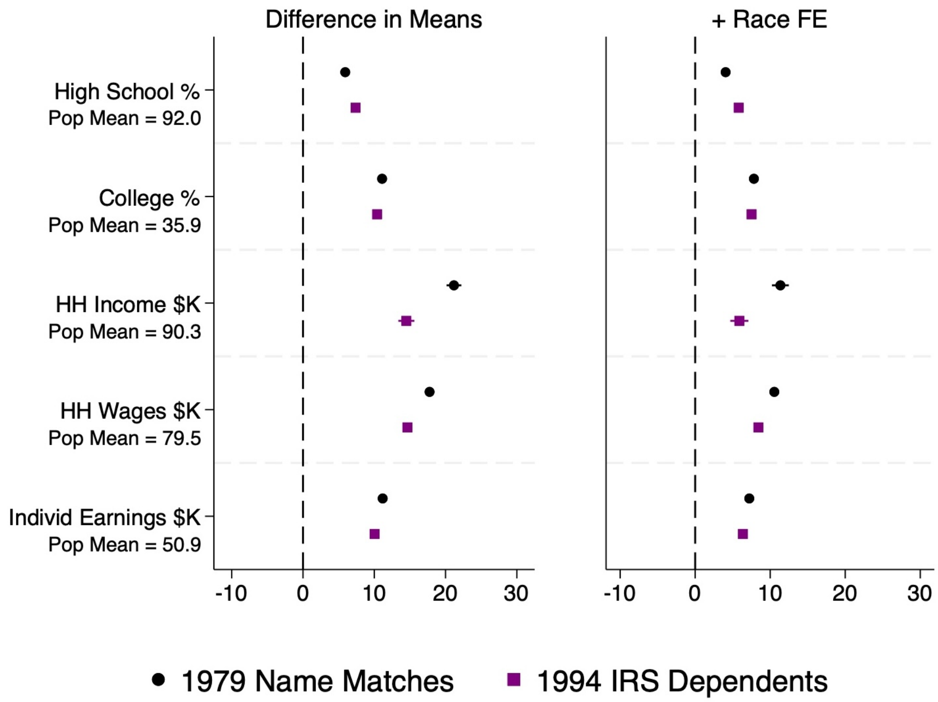
Representativeness – To validate the quality of the matches, I compare our matches to the administrative recorded linkages used by Chetty et al. (2014) for the cohort of 1979 which, as shown in Figure 1, is one of the few years that appears in both datasets. I examine how representative the matched children are of the overall population of children in Figure 2 for both sets of linkages. Comparing the unmatched Numident children to the matched children, matched ones tend to fare better later in life in terms of both educational attainment and adult income, including when race fixed effects are added. Importantly, our matches are as representative as the linkages used by Chetty et al. (2014). The relative similarity in representativeness suggests most of the selection into the sample comes from tax filing being non-random (tax filers are positively selected) rather than bias produced from our matching algorithm.

age range of 41 to 45 since W-2 earnings files are available starting in 2005 when the 1964 cohort is aged 41.

[6] In Panel C of Table A.4, 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.

[7] 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.

Figure 2: Representativeness of Matches for the 1979 Cohort:
Characteristics of Matched vs Non-matched Children



Note: The difference in means compares children born in 1979 who are matched to those who are not matched. Matching occurs either through name-matching in the 1979 tax filing or through being directly recorded in the 1994 tax filing. High school and college graduation rates come from the ACS surveys. Income is Adjusted Gross Income and Wages is Wage & Salary income 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. All income and earnings are in 2018 dollars. Race fixed effects are included in the right panel. CBDRB-FY23-CED006-0011, CBDRB-FY25-CES022-002

3 Framework to Map Policy into Intergenerational Mobility

In this section, I provide a general framework for mapping place-specific shocks into aggregate outcomes of children. This framework accounts for the changes not only in the neighborhoods that are targeted by the place-based intervention, but changes in all neighborhoods. Reallocation in response to shocks can endogenously affect the characteristics of a broad set of neighborhoods, with some spillovers being negative if economic benefits are shifted away. This reallocation further changes which individuals are exposed to the neighborhood-level impacts of place-based policy. Altogether, it is unclear if, in aggregate, children's outcomes are improved. The net impact of these various forces is an empirical question.

Below, I lay out a simple model that delineates the set of parameters needed for an aggregate quantification of intergenerational impacts. With its general structure, it can be applied to other place-based policies beyond transportation infrastructure.

3.1 Aggregate Consequences of Place-Based Policies on Income Mobility

Children's long-run incomes as adults are the main outcome of interest. I derive an expression that can be populated with purely empirical estimates and does not require further structure on the general equilibrium system.

I consider the distributional impacts across heterogeneous groups of children. Groups are denoted with the subscript g and represent categories such as parental income quintile e.g. $g \in (Q_1, \ldots, Q_5)$. Individual children's long-run 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 including parental income p_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

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

The vector of neighborhood characteristics $\mathbf{x}_n$ is of length K. In the setting of Interstate development, it includes the 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 of the population that is high-SES (based on parental background), which can change if there is differential sorting in response to policy shocks.

For estimation, I specify a linear function for $f_g(\cdot)$, as is typically done in the literature, where children's long-run incomes are determined as follows:

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

In the above equation, α_g is a group-specific factor that is a level shifter of children's long-run incomes. Importantly, $\beta_{g,k}$ is the causal impact of neighborhood characteristic $x_{n,k}$ on children.

While children's outcomes are determined at the individual level, I focus on features at the neighborhood level to clarify how place-specific shocks affect outcomes. To do so, I partition the set of children in S_g into the neighborhoods they live in for $n = 1, \ldots, N$ such that $S_g = \{S_{g1}, \ldots, S_{gN}\}$. The aggregate long-run income of children (which averages across all children) can then be re-formulated as an aggregator of neighborhood characteristics with neighborhood shares as exposure weights.

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

Specifically, π_{ng} is the share of children from group g living in n.

Through the lens of this expression for aggregate children's incomes, 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 households of different SES groups live (i.e. peer composition).

A general shock represented by δ transmits into aggregate children's long-run outcomes with the following first-order approximation that sums up across all neighborhoods.

$$
\boldsymbol{\Delta}\overline{y}_g = \underbrace{\sum_n \pi_{ng}\,\beta_g^\top\,\boldsymbol{\Delta}\,\mathbf{x}_n(\delta)}_{\text{feature changes}} + \underbrace{\sum_n \boldsymbol{\Delta}\,\pi_{ng}(\delta)\,\beta_g^\top\,\mathbf{x}_n}_{\text{relocation/exposure changes}}
$$

The first component is the change in neighborhood features due to the shock $\left(\boldsymbol{\Delta}\,\mathbf{x}_n(\delta) = \frac{\partial x}{\partial \delta}\delta\right)$. The second component is the change in exposure/neighborhood shares as a result of migration responses to the shock $\left(\boldsymbol{\Delta}\,\pi_{ng}(\delta) = \frac{\partial \pi_{ng}}{\partial \delta}\delta\right)$.

Returning to the context of this paper, I then specify that the shock is due to the Interstates raising Commuting Market Access (CMA), an aggregator that will be defined in the empirical section. The two neighborhood characteristics of interest that change with highway development are neighborhood average income ($\bar{p}_n$) and peer composition represented by percentage top-quintile ($pctQ_{5n}$). The first-order approximation is then the expression

$$
\boldsymbol{\Delta}\overline{y}_g = \underbrace{\sum_n \pi_{ng}\beta_{g,\bar{p}}\,\boldsymbol{\Delta}\,\bar{p}_n}_{\text{avg income changes}} + \underbrace{\sum_n \pi_{ng}\beta_{g,pctQ_5}\,\boldsymbol{\Delta}\,pctQ_{5n}}_{\text{peer composition changes}} + \underbrace{\sum_n \boldsymbol{\Delta}\,\pi_{ng}\,\beta_g^\top\,\mathbf{x}_n}_{\text{relocation/exposure changes}}
$$

The neighborhood-level changes come from the improvement of economic conditions because of the direct effects of commuting access ($\boldsymbol{\Delta}\,\overline{p}_n = \frac{\partial \bar{p}_n}{\partial CMA}\Delta CMA$), the migration responses to the improvement in economic conditions ($\boldsymbol{\Delta}\,\pi_{ng} = \frac{\partial \pi_{ng}}{\partial CMA}\Delta CMA$), and the changes in peers in the origin and destination locations ($\boldsymbol{\Delta}\, pctQ_{5n} = \frac{\partial pctQ_{5n}}{\partial CMA}\Delta CMA$).

For quantification of this expression, estimates are required for the changes in neighborhood features ($\boldsymbol{\Delta}\,\bar{p}_n$, $\boldsymbol{\Delta}\, pctQ_{5n}$) and migration ($\boldsymbol{\Delta}\,\pi_{ng}$) in response to the Interstate development. In addition, the expression relies on estimates of the β_g parameters for the treatment effects of neighborhood characteristics on children, which do not yet exist in the literature. In Section 6, I estimate β_g using a movers design for families that move across origin and destination tracts with different characteristics. These β_g parameters are not specific to Interstate highways and are of general interest.

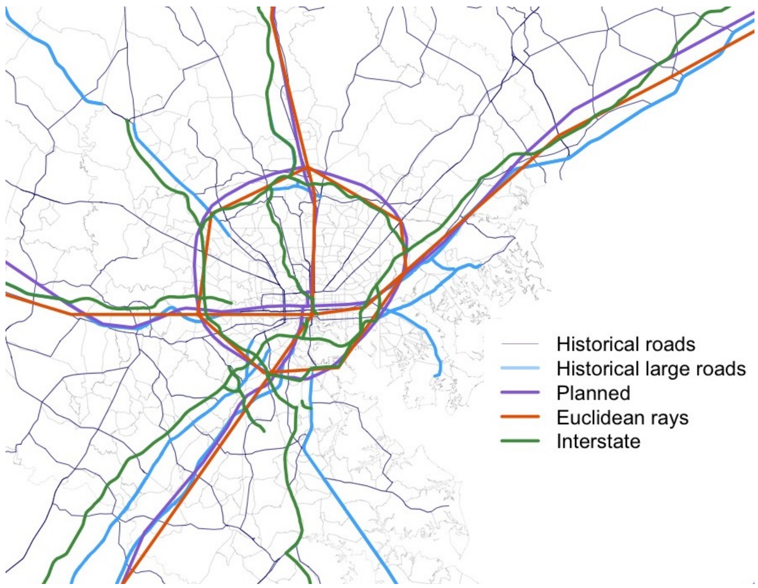
4 Historical Background on the Interstate Highway System

The Interstate highway system is 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 children in the newly linked administrative data, who were 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 out of consideration.

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 expansions, such as through the Infrastructure Investment and Jobs Act of 2021, Interstate development continues to today.

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 (Weiwu, 2025). Neighborhoods in the center of city were left behind in the wake of progress in suburban areas.

Figure 3: Historical Roads, Instruments, and Highways for the Baltimore Metro Area



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 shown, 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 plans. Interstate routes are the constructed Interstate network.

In contrast to the benefits, Interstate routes displaced hundreds of thousands of families and polluted the nearby environment. 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 and environmental activism.

While the displacement and environmental pollution caused by Interstate highways may have affected intergenerational mobility, this study focuses on the economic benefits of transportation, which generalizes to other place-based policies that aim to stimulate economic activity. In estimation, the data sample is limited to areas away from highways to reduce the influence of these local harms on the empirical findings.

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. Some historical roads were converted to Interstate, and the ones that were not are control areas to compare against. 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). These geographic features are also included as controls.

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). Neighborhoods coincidentally between major cities are more likely to have an Interstate highway built through them.

An example of the various networks for Baltimore is depicted in Figure 3. As is visible in these maps of Baltimore, planned routes and Euclidean rays follow the general direction of Interstate highways, and highway routes replaced existing large roads in many cases. Additional details on pre-trends and balance for the route instruments are located in Weiwu (2025).

5 Empirical Evidence on Highway Impacts

While estimation in this paper occurs through reduced-form, empirical specifications, the estimating equations can be micro-founded using a quantitative spatial model. I begin with the model to illustrate where the empirical equations and the commuting market access (CMA) aggregator originate from.

I then provide cross-sectional evidence on how children's long-run income is related to commuting access improvements in their location of birth, without yet decomposing the channels through which CMA affects children. These comparisons between more and less-treated locations are informative but insufficient for understanding aggregate impacts. Subsequently, I return to the expression laid out in Section 3.

To quantify the components in the aggregate impacts expression, I next estimate how neighborhood average income rises with CMA improvements and measure how reallocation from Interstate development affects all neighborhoods, not only those directly targeted by highways, through migration responses. These neighborhood changes then feed into the aggregate impacts on children.

5.1 Microfoundations using a Spatial Model of Neighborhoods

This model builds on existing quantitative spatial frameworks with commuting networks as described in Ahlfeldt et al. (2015) and Tsivanidis (2022).

Agents in the model who choose locations are the parents of children. Households are heterogeneous by SES 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 ($d_{nmg} = t^{\kappa}_{nmg}$) after receiving an idiosyncratic shock for residential locations and an idiosyncratic shock for workplaces.

A perfectly elastic housing construction sector responds to changing housing demand across neighborhoods. In equilibrium, housing markets clear to determine residential populations.[8]

Parent i's indirect utility is expressed following

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

They 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.

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

$$\pi_{mg|n} = \frac{T_{mg}(w_{mg}/d_{nmg})^{\phi}}{\sum_l T_{lg}(w_{lg}/d_{nlg})^{\phi}} = \frac{T_{mg}(w_{mg}/d_{nmg})^{\phi}}{\Phi_{ng}} \tag{2}$$

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

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

d_{nmg} from m increase CMA.

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

$$\pi_{ng} = \frac{(B_{ng} CMA_{ng} Q_n^{\beta_g - 1})^{\theta_g}}{\sum_t (B_{tg} CMA_{tg} Q_t^{\beta_g - 1})^{\theta_g}} \tag{3}$$

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 $\mathbf{L}_g$.

$$L_{ng} = \pi_{ng} \mathbb{L}_g \tag{4}$$

These expressions are next used to predict the changes in neighborhood features and residential shares that enter into the aggregate impacts expression and determine children's aggregate outcomes.

5.2 Model Predictions for Neighborhood Characteristics

Neighborhood Average Income – Expected income of tracts is a characteristic of neighborhoods that corresponds to a prediction of the model. Assuming that parental income is equal to the wages they receive, $p_i = w_i$, the spatial model provides an expression for neighborhood average income.

Using the equation for conditional commuting shares in Equation (2), neighborhood average income aggregates across workplaces and the wages received in those locations.

$$\overline{w}_{ng} = E[w_{mg}|n] = \sum_m \pi_{mg|n} w_{mg} = \sum_m \frac{T_{mg}(w_{mg}/d_{nmg})^{\phi}}{\sum_s T_{sg}(w_{sg}/d_{nsg})^{\phi}} w_{mg} \tag{5}$$

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. This expression illustrates why greater CMA is expected to change neighborhood average income. As commute costs decline, parents are able to commute to workplaces with higher wages (commuting shares $\pi_{mg|n}$ change), and average income increases ($\Delta \overline{p}_n > 0$).

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 (3), the model provides a means for obtaining the change in residential shares/exposure to neighborhoods 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, peer composition represented as the percentage of households who are in the top-quintile can be easily calculated as

$$pctQ_{5_n} = \frac{L_{n,Q_5}}{L_n} \tag{6}$$

where $L_n = \sum_g L_{ng}$. This expression for sorting of peers across neighborhoods determines one component of the change in place characteristics from the Interstate highway system.

In later sections, after estimating θ_g, I use the above residential choice equations to quantify $(\Delta \pi_{ng})$ and $(\Delta \, pct \, Q_{5_n})$.

5.3 Decennial Census and Commuting Data

To measure commuting access, I use microdata from the Decennial Censuses from 1960 onwards 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 group. Since tracts are constantly being re-defined over time, I create consistent tract definitions using the Longitudinal Tract Database (Logan et al., 2014). 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 group.

Commuting market 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 71 of the largest U.S. cities 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. Summary statistics on characteristics of census tracts using the 1970 Decennial microdata are shown in Appendix Table A.7.

5.4 Measurement of Commuting Market Access

Commuting market access empirically is characterized as a specific case of the CMA measure micro-founded using the quantitative spatial model. I define the empirical CMA aggregator below.

Let n be the residential neighborhood at the tract level and m be the workplace location. Commuting market access from a neighborhood n aggregates over all workplaces $m \in \{1, \ldots, M\}$ with the two connected by commute costs d_{nm}.

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

In the summation, wages w_m at workplace m are discounted by the commute costs d_{nm} which follow the functional form $d_{nm} = \exp(t_{nm}^{\kappa})$ 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 the empirical CMA under the definition of CMA from the model section, the labor supply elasticity $\phi = 1$ and the workplace scaling factor $T_m = \frac{L_m}{\mathbf{L}}$. Commuting parameter $\kappa = 4$ following estimates from Weiwu (2025). Wages and employment levels come from the Decennial microdata, and commute times come from the ArcGIS network calculations.

Exogeneity of Commuting Access Induced by Interstate Highway Shock – The empirical results measure neighborhood-level changes after an improvement in CMA (in most specifications, from 1960 to 1970). Changes in neighborhood outcomes to the CMA shock are measured in log-log form, where $\Delta_{70-60} \log CMA_n$ is defined as

$$\Delta_{70-60} \log CMA_n = \log \sum_m \frac{w_{m,1970} L_{m,1970}}{d^{HW}_{nm,1970} \mathbf{L_{1970}}} - \log \sum_m \frac{w_{m,1960} L_{m,1960}}{d^{HW}_{nm,1960} \mathbf{L_{1960}}}$$

To obtain a more exogenous form of the shock to commuting access, the instrument for changes in the CMA measure only includes changes in the commute costs from Interstate development. Wages and employment are fixed at baseline levels. Specifically, I set wages and employment for workplaces m to their values in 1960 and commute costs d_{nm} to values in later decades (e.g. 1970, 1980) after the construction of the Interstate highway system. Consequently, $\Delta_{70-60} \log CMA^{IV}_{n,HW}$ is the following

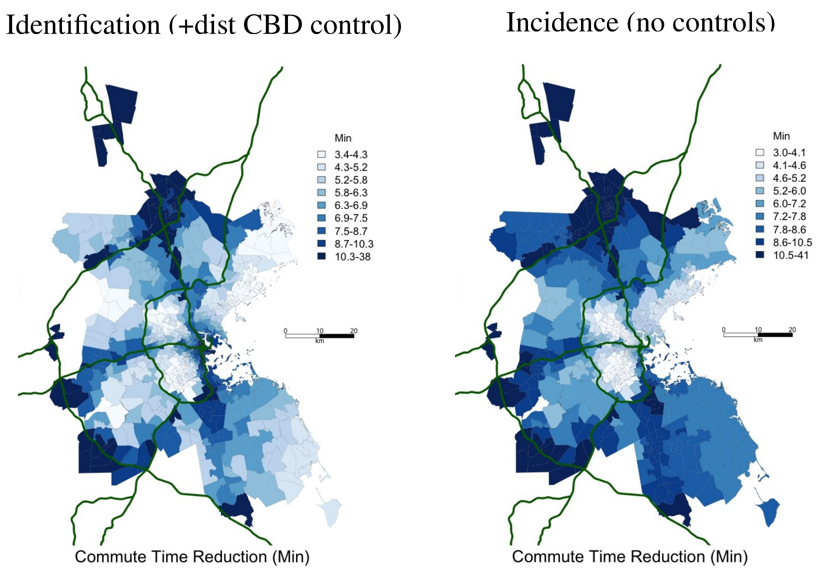
$$\Delta_{70-60} \log CMA^{IV}_{n,HW} = \log \sum_m \frac{w_{m,1960} L_{m,1960}}{d^{HW}_{nm,1970} \mathbf{L_{1960}}} - \log \sum_m \frac{w_{m,1960} L_{m,1960}}{d^{HW}_{nm,1960} \mathbf{L_{1960}}}$$

As highway placement is not completely random, I modify the CMA instrument with the two alternative instruments for route placement where the change in commute costs comes from the construction of the planned network or the Euclidean rays. For the instruments $\Delta_{70-60} \log CMA^{Plans}_{n,HW}$ and $\Delta_{70-60} \log CMA^{Rays}_{n,HW}$ respectively, I replace $d^{HW}_{nm,1970}$ with d^{Plans}_{nm} and d^{Rays}_{nm}.

5.5 Cross-sectional Evidence on Children's Income after CMA shock

With the empirical CMA measure, I estimate in a cross-sectional specification how children's income in adulthood y_i is related to increases in CMA during their childhood. I limit the sample to children born between 1964 and 1970 and measure the CMA change from 1960 to 1970 at their location of birth n, which is recorded in the Numident. The specification is the following:

Figure 4: Spatial Variation in Interstate Commuting Shock for Identification vs. Incidence



Note: The changes in commute times come from network calculations in ArcGIS. For each census tract, the plotted value takes an average across all workplaces of travel time to that tract. The left plot residualizes these tract-level values on distance from the CBD. The right plot depicts the raw average of commute times.

$$\log y_i = \phi_{m(i)} + \lambda \Delta \log CMA_n + \mathbf{X}_i \gamma + \mathbf{X}_n \zeta + \nu_i$$

In the rest of the equation, the individual-level controls $\mathbf{X}_i$ include parental income p_i, which is calculated as the mean of parental income during childhood. The effects of CMA then come through changes to the place, such as through other households having more income and thereby improving public goods, rather than changes in the resources of the child's household.

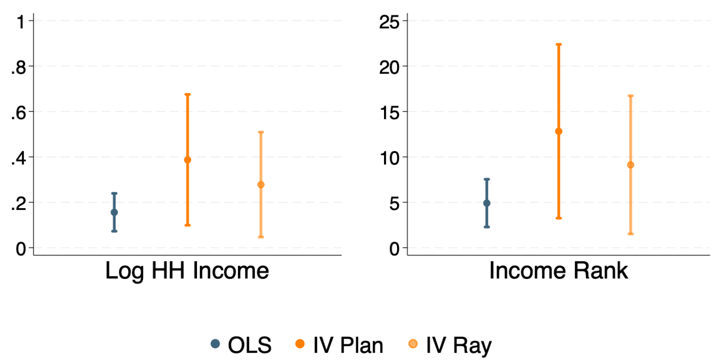
As CMA is selectively rising across locations, for proper identification of the causal effects of the Interstate shock, geographic controls for distance to historical large roads (Interstate candidates), railroads, rivers, canals, and ports are included. I control for distance from highways to absorb effects related to pollution. The sample is already limited to areas away from highways.

Notably, distance from the CBD ($distCBD_n$) is also controlled for. As shown in Figure 4, CMA is increasing the most in the suburbs. Without the $distCBD_n$ control, the cross-sectional comparison would be biased since the suburbs are substantially different from the central city along many dimensions. Identification, after controlling for distance from the CBD, then relies on variation within the suburbs, comparing one suburb that received an Interstate highway to one that is along a large road that was not converted to an Interstate.

Metro area (CBSA) fixed effects absorb variation across metros, so the comparison is across neighborhoods, within metros. For further robustness related to exogeneity of the CMA shock, the reduction in commute costs in CMA is instrumented using the construction of the planned or ray routes rather than the Interstate routes.

In the results shown in Figure 5, the outcome is either the log of the child's income or the child's income rank. The estimated elasticities between changes in children's income as adults and changes to CMA during their childhood are large and range from around 0.15 using OLS to 0.4 using the IV strategy. The IV estimates are more positive, suggesting there is negative selection in the placement of Interstate roads. The OLS estimates already include the geographic controls and thus are purged of positive selection in highway location. When income ranks are used as the outcome, the estimated values for the coefficient on $\Delta \log CMA_n$ range from 5 to 13. These numbers are substantial in size and indicate that commuting access is an influential factor for how places shape economic opportunity. To benchmark the magnitude of the coefficient, the Black-White income rank gap is around 12 income ranks, so doubling CMA in Black neighborhoods could potentially close this gap, assuming the estimated coefficient is unbiased and constant across groups.

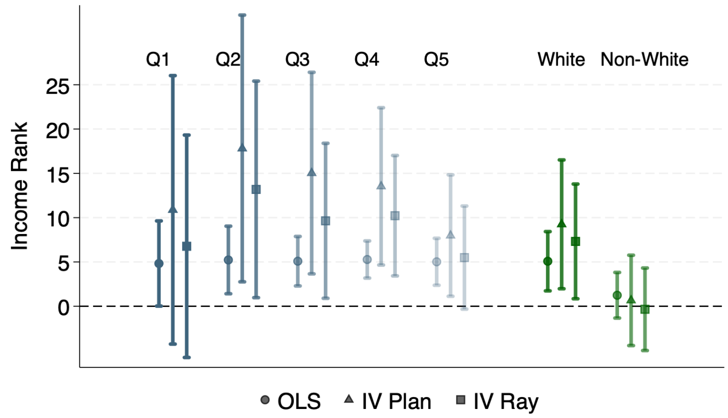
Figure 5: Children's Adult Income to CMA Shock at Place of Birth



Notes: 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. CMA is measured using the Decennial microdata in 1960 and 1970.

However, the coefficient may not be constant and could be highly different across groups. In the next set of results, I estimate whether the impact of improvements in CMA at an individual's childhood residence is heterogeneous across parental income background or the race of the child. Displayed in Figure 6, we see that the effects are strongest for children from parental income backgrounds in the middle of the distribution. Additionally, only White children, not Black children appear to benefit from increases in CMA. These findings by race may be due to the mechanism through which CMA is increasing. As interstate development only reduces commute times by car, low car ownership by Black families may lead them to gain less.

Figure 6: Heterogeneity Across Income and Race



Notes: 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. CMA is measured using the Decennial microdata in 1960 and 1970. Race comes from the 2000 Census/ACS surveys. The quintiles are categorized using parental income rank, which

These reduced form comparisons between more and less treated locations are valuable and new to the literature, since no existing paper has examined the long-run effects of the Interstate highway system. These findings indicate that the large-scale investment undertaken by the U.S. government generated both short-term and long-term, dynamic effects for the economy. However, these comparisons are insufficient for understanding the total impact of interstate development for multiple reasons.

First, there are two dimensions through which highways impact places. There are positive benefits for targeted areas and potentially, also negative spillovers to other locations. The reduced form comparison bundles these two channels. In later sections, I disentangle the multiple dimensions of the highway shock into the economic benefits and peer spillovers from migration responses.

Additionally, place is measured at the location of birth for these children. With residential mobility, only stayers receive the full treatment effect from improvements in job access (corresponding to the treatment on the treated) versus the current estimate is an intent-to-treatment estimate that does not account for movers. However, moving itself is an endogenous choice, so limiting the sample to stayers would create bias in the estimates. Using a movers design in which moving is allowed to be endogenous but the *timing* of the move is not, I estimate the treatment effects of greater dosage of locations on children.

Lastly, the reduced form comparisons ignore the differences in spatial exposure to the treatment such as which groups are living more in areas that receive a large shock. The aggregate impacts expression includes the residential shares and weights the neighborhood changes with the share of the children that are located in each place.

5.6 Economic Impacts of Improved Commuting Access for Neighborhoods

In this section, I return to estimating neighborhood-level changes from the Interstate shock to quantify the aggregate impacts expression. Rather than using a cross-sectional comparison, I employ a difference-in-differences strategy where the neighborhood-level outcome is changing between 1960 to 1970 and is regressed on the change in CMA between 1960 and 1970.

The specification shown below is a long-difference equation from 1960 to 1970.

$$\Delta \log (avginc_n) = \psi_{m(n)} + \gamma \Delta \log CMA_n + \mathbf{X}_n \zeta + \nu_n$$

This equation is similar to the equation where children's adult income in levels was the outcome, but here the outcome is neighborhood average income in changes. The geographic controls in X_n, accounting for the non-random placement of the Interstate highway system, are distance to the CBD, highways, large roads, railways, canals, rivers, etc. The sample drops areas next to highways. CBSA fixed effects are included to exploit variation within metro areas in where CMA is rising the most. Changes in commuting are also instrumented using the planned or ray routes.

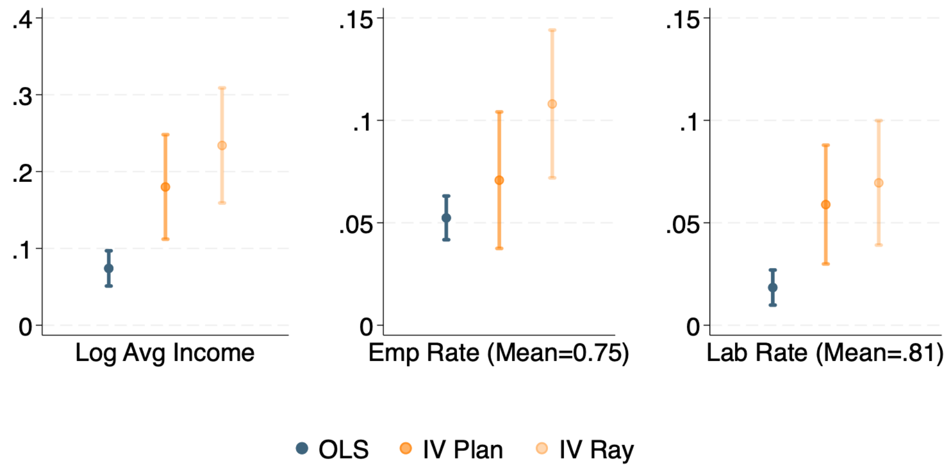
The OLS results are plotted in Figure 7 where I find that greater CMA results in increases in tract-level average income. The estimates instrumented using the changes in commute times coming from the planned and ray networks are larger, likely due to negative selection on trend in where highways were located. Going from OLS to IV, the coefficient increases in magnitude to about twice the size, from an elasticity of 0.8 to an elasticity of 0.24. Overall, the estimates suggest that commuting access enables residents of a neighborhood to obtain better jobs and increase their incomes. These place-specific effects can then shape economic opportunity for children since neighborhood residents contribute to public services through tax payments.

In two additional panels, I present results for changes in the employment rate and labor force participation rate, since earlier studies by Kain (1968) and Wilson (1987) discuss how unemployment is related to spatial mismatch. I find positive relationships with changes in CMA as well. For the purposes of how CMA affects children, I route the effects entirely through neighborhood average income, rather than unemployment. Average income is calculated including unemployed individuals, so it includes both the intensive and extensive margins in the labor market.

5.7 Population Sorting to Improvements in Commuting 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 economic status. With differential migration, there would mechanically be changes in peer composition in response to changes in commuting access from the Interstate network, as is indicated by equation (6) from the model section. I follow a difference-in-differences strategy similar to the one used for measuring

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



Note: Tract characteristics are calculated using the Decennial Census in 1960 and 1970. Average income, employment rate, and labor force participation rate is calculated among men aged 16 and up. Employment rate has men aged 16+ as the denominator and employment among men aged 16+ as the numerator. Coefficient estimates, standard errors, dependent variable means and F-stats are reported in Table A.8

neighborhood income changes to CMA improvements.

Taking logs and first differences of Equations (3) and (4), I obtain the estimation equation below.

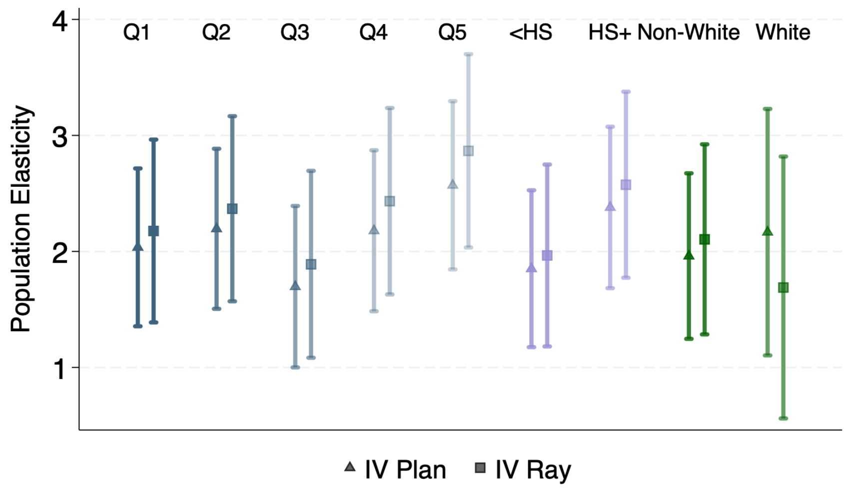
$$\Delta \log L_{ng} = \phi_{gm(n)} + \theta_g \Delta \log CMA_n + \mathbf{X}_n \eta_g + \epsilon_{ng} \qquad \text{[second-stage]}$$

The specification includes controls in $\mathbf{X}_n$ for the previously discussed geographic features and CBSA fixed effects. The first difference is over 1960 and 1970 as was the case with the earlier estimating equation.

I split the microdata into several groups to document who responds to improvements in CMA. Groups g are separated by occupation quintile, race, education, and race. 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 CMA, 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 persistent within an individual, population responses by occupational quintiles more closely reflects sorting.

I present the elasticities for each group in Figures 8. Higher occupational status households (Q_5) are the most responsive to job access improvements with a population elasticity to CMA around 2.7. The more educated (more than a high-school degree) are also more mobile with an elasticity of 2.5 compared to an elasticity of 2 for the less educated. Although the standard errors are large for the

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



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

Black population, I find they migrate less than the White population. These results all point to more advantaged populations responding to a greater extent to higher access to employment. Because the Interstate highway system was a large shock that induced substantial migration, equilibrium effects are likely important for the overall impact on economic opportunity. Referring to Figure 4, the reduction in commute costs is greatest in the suburbs, so the population flows are moving outwards to suburban neighborhoods. The heterogeneity in migration leads those suburban areas to become populated by more advantaged families, while the households who stay behind are lower in socio-economic status. In the aggregate impacts expression, I focus on the highest income peers (those in Q_5), and calculate the change in composition $(\Delta pctQ_{5_n})$ given these empirical elasticities.

In this empirical specification, unlike in a discrete choice model, changes in rents or endogenous amenities do not have add-on effects for the sorting response. Instead, all of those channels are combined into one empirical elasticity. For example, low-occupational status households may migrate less because neighborhoods with greater CMA increases are becoming more expensive. The estimated elasticity captures that lower migration response but does not take a stance on why it is lower. In other settings with different types of equilibrium effects that shape sorting, these elasticities would need to be re-estimated, as put forth in the Lucas critique.

Lastly, in Table 1, I infer how much of the income change documented in Figure 7 is a result of sorting. Since neighborhood average income is measured using repeated cross-sections, the composition of households is changing and could be driving the average income rises, rather than existing residents obtaining higher-wage jobs. Using the microdata, I infer how much of the income change can be predicted using the change in educational or occupational composition of the

Table 1: Contribution of Migration to Neighborhood Income Changes

	OLS		IV Rays	
	log Income (Occ)	log Income (Educ)	log Income (Occ)	log Income (Educ)
$\Delta \log CMA$	0.0144*	0.00462	0.0399*	0.0361***
	(0.00581)	(0.00271)	(0.0181)	(0.00979)
R-squared	0.138	0.125	0.0681	0.0585
CBSA FE	Yes	Yes	Yes	Yes
Rounded Obs	20500	20500	20500	20500
KP F-Stat			562	562

residents. As shown in Table 1, the sorting only explains 20% of the increase in average income. This finding should not be surprising given the results on population elasticities to CMA increases. While there is some heterogeneity across groups, *all* groups are moving outwards since even for the lowest-quintile households, the population elasticity is around 2. In the aggregate impacts expression, I re-scale the change in neighborhood outcome to be lower since some of it should be attributed to the migration response.

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 (2018) find much of county level place effects can be explained by observable characteristics. I focus on the characteristics previously documented to have changed with the Interstate highway system—average income and peer composition.

Much of the associations between characteristics and children's outcomes 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 households who are more mobile in response 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 high-SES families tend to vary along many dimensions such as crime levels, local governance, 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 neighborhood income or peers is not driving the differences in outcomes for children but rather differences in local governance correlated with income or peers, the correlations would not be informative for the treatment effect of increasing neighborhood income or peer quality.

In light of these identification challenges, in later sections I develop a research design 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 shock from the Interstate highway system to construct instruments for neighborhood characteristics within the movers design. The movers-IV design simultaneously addresses the two sources of bias.

6.2 Movers Exposure Design to Address Selection in Place Effects

In the movers exposure design developed 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.

I first present the basic mechanics behind the movers design before extending it to the particular application of this paper. The sample focuses on the set of children who move once during their childhood until up the age of 28. Let m_i be the age at which child i moves from origin neighborhood o to destination neighborhood d. In this specification, I examine moves across census tracts. Let $\bar{y}_{pn}$ be the the exposure-weighted outcome of y_i (child household income rank) for children who grew up in location n with parental household income rank p.[9] These tract-level average predicted

[9] 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 **??**. The exposure weights that are used for predicting income rank are constructed from residential location up until age 23.

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 census tracts where the average child fares better in adulthood. Let $\Delta_{odp} = \bar{y}_{pd} - \bar{y}_{po}$ be the predicted difference in children's income ranks in the destination versus origin county. I regress the income rank of children who move on the change in origin and destination quality interacted with age-at-move fixed effects.

$$y_i = \sum_{s=1964}^{1979} I\{s_i = s\}(\lambda_s^1 + \lambda_s^2 \bar{y}_{po}) + \sum_{m=1}^{28} I\{m_i = m\}\phi_m + \sum_{m=1}^{28} b_m I\{m_i = m\}\Delta_{odp} + \epsilon_{1i} \quad (7)$$

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 $(\lambda_s^1 + \lambda_s^2 \bar{y}_{po})$ 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

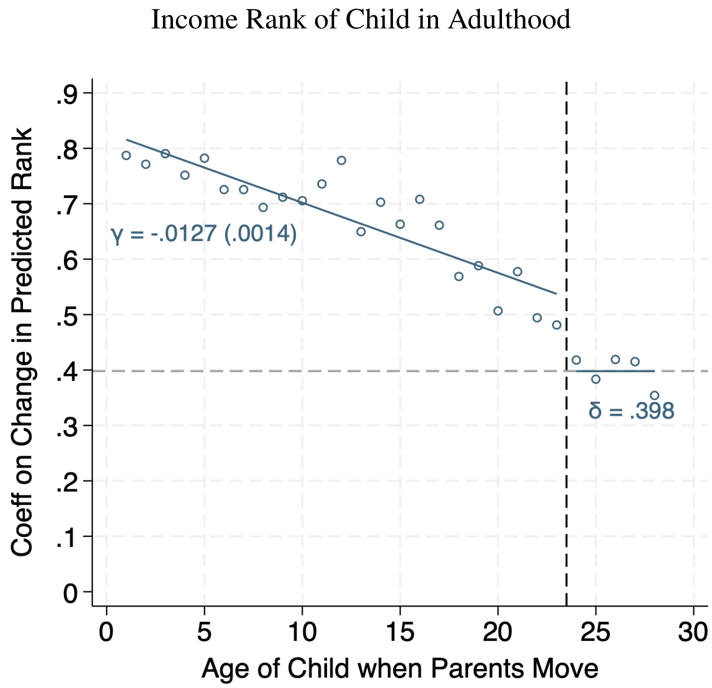
$$\begin{aligned} y_i = &\sum_{s=1964}^{1979} I\{s_i = s\}(\lambda_s^1 + \lambda_s^2 \bar{y}_{po}) + \sum_{m=1}^{28} I\{m_i = m\}\phi_m \\ &+ I\{m_i \leq 23\}(\rho + (23 - m_i)\gamma)\Delta_{odp} + I\{m_i > 23\}(\delta^1 + (23 - m_i)\delta^2)\Delta_{odp} + \epsilon_{2i} \end{aligned} \quad (8)$$

Age at move fixed effects and cohort fixed effects interacted with origin predicted income rank are included.[10] The coefficient of interest is γ for the exposure effect for each year spent in the destination location up until age 23. Exposure to the treatment after age 23 δ^2 is presumed to be zero, and I test for this result in the estimation. The intercept term δ^1 is the correlation between the difference in quality of origin versus destination locations and children outcomes who move to the destination at age 23. Because it is assumed that treatment effects end at this point, any correlation would signal selection in the choice of destination neighborhood relative to the origin neighborhood.

I present the age at move coefficients in Figure 9 as a scatter plot. In Figure 9, linear lines fit the estimated coefficients for the age at move bins before and after age 23 and indicate that the

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

Figure 9: Exposure to Place Effect over Age at Move for Movers

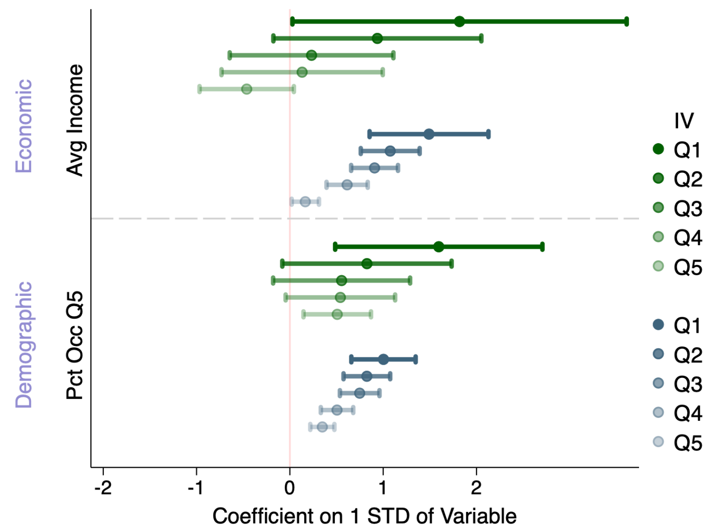


Note: Household income of the child comes from the 1040 forms and is the average of the five years during which the child is aged 35-39. Children are assigned percentile ranks relative to all other children in their birth cohort. The specification calculates the coefficients for child income rank in 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 place with a 1 percentile higher predicted individual income rank in adulthood for children. 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.

exposure effects per year are statistically significant. For children who move to a better location at age 0 and spend their whole childhood in that place, their income rank is substantially higher than the income rank of children who move at age 10 and spend less time in the new location.

If parental income changes, then assessing the impacts of moving to a better location becomes contaminated by the change in household resources rather than being solely the effect of place. In the earlier cross-sectional evidence on how the CMA shock affected children, I controlled for parental income to eliminate this channel from impacting children. For the movers design, 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. In Appendix Table A.9, I find no effect for parents, suggesting the impacts on children are coming from changes in the place.

Figure 10: Causal Impacts of Tract Characteristics



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 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 is calculated among men aged 16 and up.

6.3 Extending the Movers Design to Study Neighborhood Characteristics

The movers design presented above can be extended to study the impact of neighborhood characteristics on children. Previous studies often estimate place effects for each location and then project these place effects on neighborhood characteristics to understand how much of the variation can be explained by any one feature. At the tract-level, this approach becomes intractable given the small sample of movers and the tremendous number of tracts in the U.S. (40,000).

The specification I estimate for the extension is similar in form to Equation 8. 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 use 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 $\mathbf{X}_i = \sum_{s=1964}^{1979} I\{s_i = s\}(\lambda^1_s + \lambda^2_s x_o) + \sum_{m=1}^{23} I\{m_i = m\}\phi_m$.[11] 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 $\mathbf{X}_i$ includes additional location-specific controls to remove omitted

[11] I use this extended movers strategy to estimate how the causal impacts of tracts are related to a larger set of tract characteristics. These results are shown in Appendix Figure B.1.

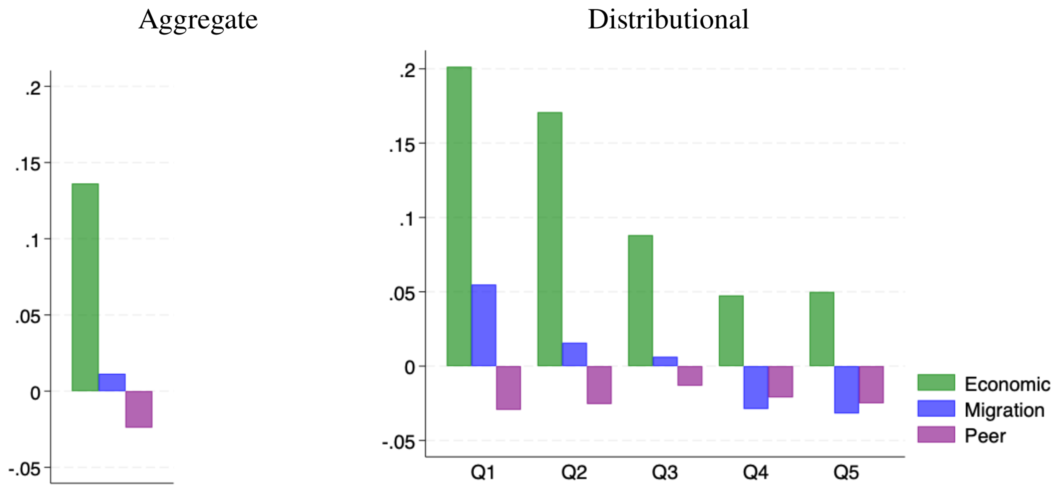
variables bias from factors correlated with neighborhood characteristics.

I implement an instrumental variables strategy using a control-function approach (Wooldridge, 2015). To isolate exogenous variation, I exploit shocks from the interstate highway network, which serve as shifters of neighborhood features. Specifically, I instrument for changes in average neighborhood income ($\bar{p}_n$) using changes in CMA (pooled across the whole population). For the share of high-SES households ($pctQ_{5_n}$), I use a group-specific version of CMA that captures heterogeneity across workplaces in the wages and employment levels by occupational groups. The control-function framework involves first estimating the first-stage regressions for each neighborhood characteristic and retrieving the residuals, which represent the endogenous components unexplained by the instruments. These residuals are then included in the second-stage regression for individual outcomes to correct for omitted variables bias. This approach yields consistent estimates of the causal impact of neighborhood characteristics while allowing for flexible inclusion of controls for both origin and destination characteristics.

With the equation above, I estimate the coefficient ψ^x for a one standard deviation difference in the neighborhood characteristics of average income ($\bar{p}_n$) and percentage top occupational quintile ($pctQ_{5_n}$). To consider the distributional impacts across different groups of children, I separately conduct the estimation for children from each parental income quintile (low/high-income children). I present the results in Figure 10. I find that there are significant treatment effects for the causal impacts of these tract characteristics. Children experience better adult outcomes in higher-income neighborhoods with higher occupation status households. Whether I use the instrumental variables for tract characteristics does not greatly change the results, suggesting that most of the relationship between the causal impacts of place and income/peers is directly due to those features, rather than some unobserved neighborhood trait.

Importantly, there is substantial heterogeneity in the treatment effects. For children from higher-income families (Q4–Q5), the coefficients on average neighborhood income and the share of high-income households are close to zero or slightly negative, indicating limited effects of exposure to higher-income environments. By contrast, for children from lower-income families (Q1–Q2), both variables exhibit large and positive causal impacts. A one–standard-deviation increase in neighborhood average income or in the share of high-income residents substantially raises their adult incomes by 1-2 income ranks. This pattern suggests that while poorer children benefit from access to economically advantaged tracts, potentially due to better schools and influences from their peers, affluent children may be less sensitive to their environment. Overall, the heterogeneity underscores that neighborhood opportunity is not uniformly increasing. Identical improvements in local characteristics can have markedly different consequences depending on children's family backgrounds.

Figure 11: Decomposition of Changes in Children's Income from Interstates



7 Aggregate Effects of Highways on Children's Outcomes

Having estimated the key parameters and neighborhood responses, namely the effects of neighborhood characteristics on children's outcomes (β_g) and the shifts in local economic conditions, peer composition, and residential shares from ($\Delta\overline{p}_n$, $\Delta pctQ_{5n}$, $\Delta\pi_{ng}$), I can now quantify the aggregate impact of the Interstate Highway System on the long-run income of children.

Returning to the expression from the simple framework, we can decompose how the changes in each component contribute to the total effect on children

$$\boldsymbol{\Delta}\overline{y}_g = \underbrace{\sum_n \pi_{ng}\beta_{g,\overline{p}}\,\boldsymbol{\Delta}\,\overline{p}_n}_{\text{avg income changes}} + \underbrace{\sum_n \pi_{ng}\beta_{g,pctQ_5}\,\boldsymbol{\Delta}\,pctQ_{5n}}_{\text{peer composition changes}} + \underbrace{\sum_n \boldsymbol{\Delta}\,\pi_{ng}\,\beta_g^\top\,\mathbf{x}_n}_{\text{relocation/exposure changes}}$$

The first captures direct improvements in local economic conditions, reflecting the local wage gains from access to workplaces induced by highways. The second represents changes in peer composition in both origins and destinations. When assessing aggregate impacts, changes in peer composition might appear to cancel out across locations as some areas gain high-income peers while others lose them. However, they do not necessarily net to zero if two forms of heterogeneity are present. First, if the effect of peer exposure ($\beta_{g,k}$) differs sufficiently across children's income groups, and second, if the distribution of those groups across neighborhoods (π_{ng}) is sufficiently uneven. In that case, peer composition changes can produce nontrivial aggregate effects because improvements and deteriorations are weighted differently across groups and locations. Finally, the last component reflects migration responses as families relocate toward newly connected or economically improved areas.

In Figure 11, I find that the primary channel through which the Interstate highway system affected children's long-run incomes is the direct improvement in local economic conditions. Areas that gained commuting access experienced income growth, which translated into higher earnings for children growing up there. These direct economic effects account for nearly all of the aggregate gains and are especially pronounced for children from the lowest-income families (Q1), indicating that infrastructure-driven growth disproportionately benefited those starting from the poorest backgrounds. These distributional consequences lead infrastructure to reduce the persistence of intergenerational inequality since poor children are more like to rise above their economic circumstances.

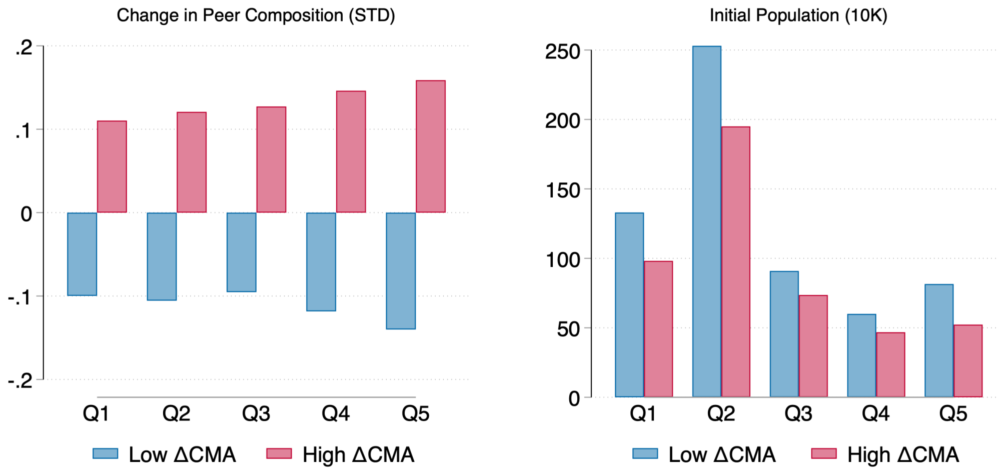
In contrast, the contributions of migration responses and changes in peer composition are comparatively small. This limited net effect of peer spillovers arises because shifts in peer quality tend to balance out (some areas gain higher-income peers while others lose them, as shown in Figure 12) and because the estimated treatment effects of peer exposure and initial location shares are not sufficiently heterogeneous to generate large overall changes. Taken together, these results suggest that the highways' main contribution to intergenerational mobility came from expanding local economic opportunity.

However, Figure 12 does indicate an increase spatial inequality across neighborhoods. Areas that experienced small or negative changes in connectivity (ΔCMA), typically central-city tracts, saw declines in the share of high-income households and thus deteriorations in peer composition. In contrast, suburban tracts with large improvements in connectivity attracted more advantaged families. These divergent trends imply that as highways facilitated suburban access to employment centers, they simultaneously drew higher-income households outward, leaving behind central neighborhoods with fewer high-income peers. Consequently, while aggregate opportunity improved, the geography of opportunity became more unequal: children in newly connected suburban areas were increasingly surrounded by affluent peers, while those in urban cores faced declining environments and potentially weaker social and institutional resources.

8 Conclusion

In this paper, I study the impacts of a far-reaching infrastructure project on economic mobility using rich historical linkages built at the Census Bureau. I explore the mechanisms behind how this particular policy impacted the features of locations, both those directly targeted by highway infrastructure and those indirectly affected through general equilibrium effects. The gains in intergenerational mobility were driven primarily by economic improvements in connected areas, with the largest benefits accruing to children from low-income families. In contrast, migration responses and peer-composition changes contributed little to the aggregate effects.

Figure 12: Change in Peer Quality and Intital Shares Across Tracts



Nevertheless, the redistribution of higher-SES households toward newly connected suburbs increased spatial inequality in opportunity. Central-city neighborhoods with limited access improvements lost advantaged peers, offsetting some of the aggregate gains in mobility. The results underscore that large-scale place-based interventions can foster upward mobility through stimulating economic activity, yet their equilibrium consequences can also reshape the geography of inequality. By developing a framework that includes the indirect equilibrium channels, this paper provides a blueprint for evaluating how the impacts of place-based policies transmit to the next generation.

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.

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.

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.

Asher, S. and Novosad, P. Rural Roads and Local Economic Development. *American Economic Review*, 110(3):797–823, 2020. doi: 10.1257/aer.20180268. URL https://www.aeaweb.org/articles?id=10.1257/aer.20180268.

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.

Baum-Snow, N., Hartley, D., and Lee, K. O. The long-run effects of neighborhood change on incumbent families. 2019. URL https://drive.google.com/file/d/1ZEL_G27IuKqOft_o1Vuxgg63XwBe6q01/view.

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., Rothstein, J., and Yi, M. Reassessing the Spatial Mismatch Hypothesis. *AEA Papers and Proceedings*, 114:221–225, 2024. doi: 10.1257/pandp.20241112. URL https://www.aeaweb.org/articles?id=10.1257/pandp.20241112.

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 2018. ISSN 0033-5533. doi: 10.1093/qje/qjy006. URL https://doi.org/10.1093/qje/qjy006.

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.

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

Duranton, G. and Puga, D. Chapter 48 - Micro-Foundations of Urban Agglomeration Economies. In Henderson, J. V., Thisse, J.-F. B. T. H. o. R., and Economics, U., editors, *Cities and Geography*, volume 4, pages 2063–2117. Elsevier, 2004. ISBN 1574-0080. doi: https://doi.org/10.1016/S1574-0080(04)80005-1. URL https://www.sciencedirect.com/science/article/pii/S1574008004800051.

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

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.

Fogli, A., Guerrieri, V., Ponder, M., and Prato, M. The End of the American Dream? Inequality and Segregation in US Cities. 2025. URL https://ideas.repec.org/p/fip/fedmoi/99487.html.

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.

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

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.

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.

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.

Valenzuela-Casasempere, P. Infrastructure provision and displacement: Evidence from the interstate highway system. 2025.

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

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

Appendices

A Tables

Table A.1: 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.2: 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.3: 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
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.4: Summary Statistics on Individual Characteristics

	(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

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. 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.5: 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.6: Summary Statistics for Movers

	(1)	(2)
Variable	1 Move	0 or 2+ Moves
High School Graduation Rate	0.955	0.953
SD	(0.207)	(0.212)
Rounded N	1501000	3183000
College Graduation Rate	0.421	0.406
SD	(0.494)	(0.491)
Rounded N	1501000	3183000
Adjusted Gross Income (2018 $K)	104.6	100.3
SD	(378.5)	(333.1)
Rounded N	6672000	14050000
Wage & Salary Income (2018 $K)	90.38	87.85
SD	(166.8)	(169.1)
Rounded N	6539000	13770000
Individual Earnings (2018 $K)	59.75	58.44
SD	(126.4)	(255.4)
Rounded N	6224000	13130000

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. Moves are calculated starting when the 1040 data is first available in 1974 up until age 23. Counts are rounded in the following manner: numbers between 10,000 and 99,999 are rounded to the nearest 500; between 100,000 and 9,999,999 to the nearest 1,000; and above 10,000,000 to the nearest 10,000 to meet Census disclosure rules.

Table A.7: Summary Statistics of Tract Characteristics in 1970

	(1)	(2)	(3)
	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.8: 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 CMA	0.0740***	0.0524***	0.0184***
	(0.0117)	(0.00545)	(0.00437)
R-squared	0.129	0.100	0.0978
		Panel C – IV Plans [KP Wald F-Stat = 621]	
Δ Log CMA	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 CMA	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

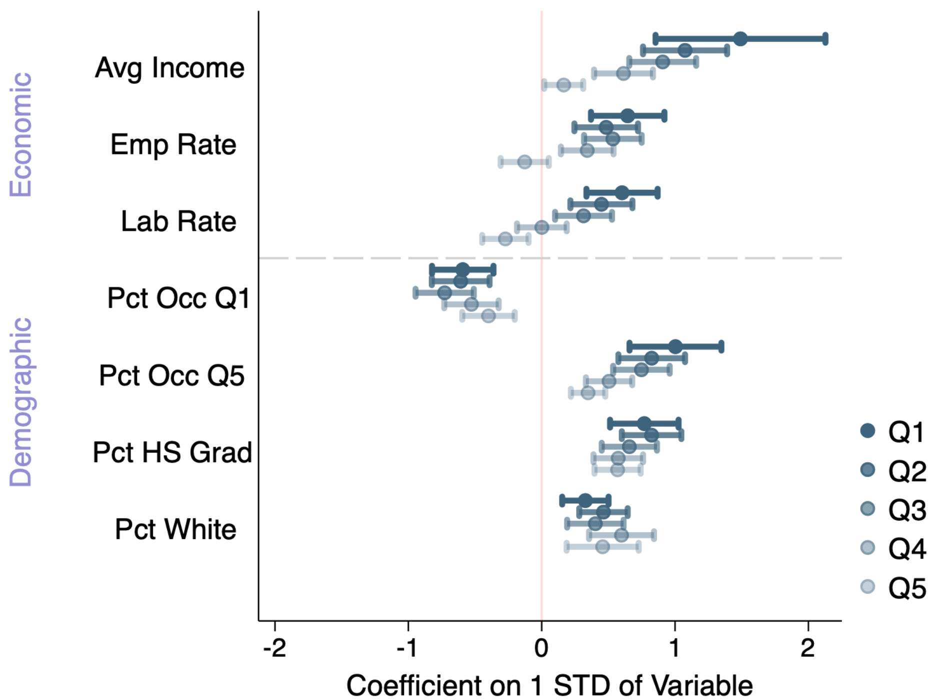
Table A.9: Parent Movers Panel - Two-way FE Income Changes in JMA

	(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: Causal impacts of tracts related to 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.

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