

Estimating Trends in Intergenerational Mobility by Race Using Multiple Data Sources

Jeremy Majerovitz (MIT)

May 4th, 2021

Introduction

Chetty, Hendren, Jones, and Porter (2019):

[B]lacks and whites are in a steady state in which the black-white income gap is due almost entirely to differences in rates of intergenerational mobility rather than transitory or historical factors... [R]educing racial disparities will require reducing intergenerational gaps—that is, disparities in children’s outcomes conditional on parental income...

So we are already in the steady state, and to shrink the black-white income gap we must shrink the black-white mobility gap.

Begs the question: has the racial mobility gap been changing over time?

Prior Literature

Chetty, Hendren, Jones and Porter (2019) - Estimate intergenerational mobility by race. Big black-white gaps.

Chetty, Hendren, Kline, Saez, and Turner (2014) - Estimate trends in intergenerational mobility, not by race. Can only go back to 1971 cohort (for 1987-1993, based on their college attendance).

Davis and Mazumder (2020a) - Estimate trends in intergenerational mobility using NLS and NLSY, show decline in IGM from 1950s cohorts to 1960s cohorts.

Davis and Mazumder (2020b) - Use same NLSY data to estimate IGM by race; not well-powered to do race x time

This Project

We want to estimate the effect of race and parent income on child income, and look at how those estimates have changed over time for different cohorts.

However, getting usefully precise estimates requires having a large data set, and there is no large data set linking parent income, child income, and race that goes far back in time.

Solution: We can combine the NLSY data used by Davis and Mazumder with moments from the census cross-sections

Today: Show how this is done, discuss some more subtle data decisions, and show initial results

Estimation

We will estimate the model:

$$Y_{ir} = \alpha_r + \beta X_{ir} + \varepsilon_{ir}$$

where α_r is a race specific intercept, Y_{ir} is the child income percentile, and X_{ir} is parent income percentile.

Simplify things with homogeneous intercept: Chetty Hendren Jones and Porter (2019) find that the intercept is very close to the same across racial groups.

For the NLSY data, we have all the standard OLS moments. For the Census cross-sections, we have:

$$\underbrace{\mathbb{E}[Y_{ir} | r]}_{\text{Child Census}} = \alpha_r + \beta \underbrace{\mathbb{E}[X_{ir} | r]}_{\text{Parent Census}}$$

GMM Estimation

Use two-step GMM. Full set of moments:

$$0 = \mathbb{E} \left[\mathbf{1}_{\text{NLSY}} \cdot (Y_{ir} - \alpha_r - \beta \cdot X_{ir}) \cdot \begin{pmatrix} X_{ir} \\ 1 \\ \mathbf{1}_{\text{Black}} \\ \mathbf{1}_{\text{Other}} \end{pmatrix} \right]$$
$$0 = \mathbb{E} \left[\mathbf{1}_{\text{Parent Census}} \cdot \begin{pmatrix} \mathbf{1}_{\text{White}} \cdot (X_{ir} - \mu^{\text{White}}) \\ \mathbf{1}_{\text{Black}} \cdot (X_{ir} - \mu^{\text{Black}}) \\ \mathbf{1}_{\text{Other}} \cdot (X_{ir} - \mu^{\text{Other}}) \end{pmatrix} \right]$$
$$0 = \mathbb{E} \left[\mathbf{1}_{\text{Child Census}} \cdot \begin{pmatrix} \mathbf{1}_{\text{White}} \cdot (Y_{ir} - \alpha_r - \beta \cdot \mu^{\text{White}}) \\ \mathbf{1}_{\text{Black}} \cdot (Y_{ir} - \alpha_r - \beta \cdot \mu^{\text{Black}}) \\ \mathbf{1}_{\text{Other}} \cdot (Y_{ir} - \alpha_r - \beta \cdot \mu^{\text{Other}}) \end{pmatrix} \right]$$

Data Definitions (NLSY and Census frames)

NLS66/NLSY79 - Data on 1948-1953 and 1961-1964 cohorts. Definitions and data cleaning from Davis and Mazumder (2020a). Use total household income. Take average of all years of data available (up to three years) to reduce measurement error.

Census/ACS - Use decadal censuses (1960 1%, 1970 1% metro fm1, 1980 5% state, 1990 1% metro, 2000 5%) and 2010 ACS. Currently not using 1950 census because detailed household income data only collected for “sample-line persons” (if you know details about this, I’d love to chat).

Measure parent income when child is age 8-17, measure child income at age 35-44.

Construct household income as total income of all family members living in the household. Then, construct percentiles by cohort.

For both data sets, define racial groups as Black, White, and Other

Lifecycle of Racial Income Gap

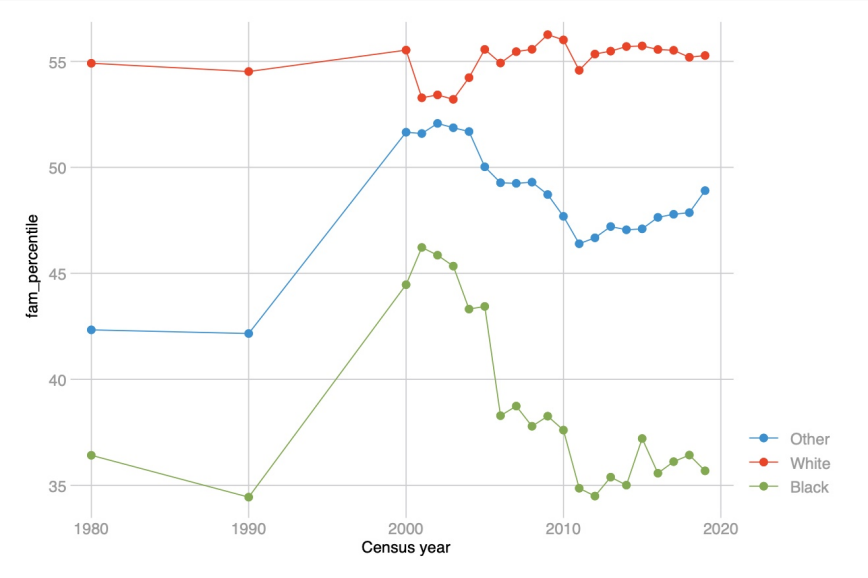
Saying that we are regressing “child household income percentile” on “parent household income percentile” does not fully specify what regression we are running. Need to clarify the age at which each of these is measured.

This is an issue in the Census because we are using cross-sections that are ten years apart: the age of measurement declines as the cohorts get more recent, until there is a switch in census years and there is a discontinuous jump in age at measurement

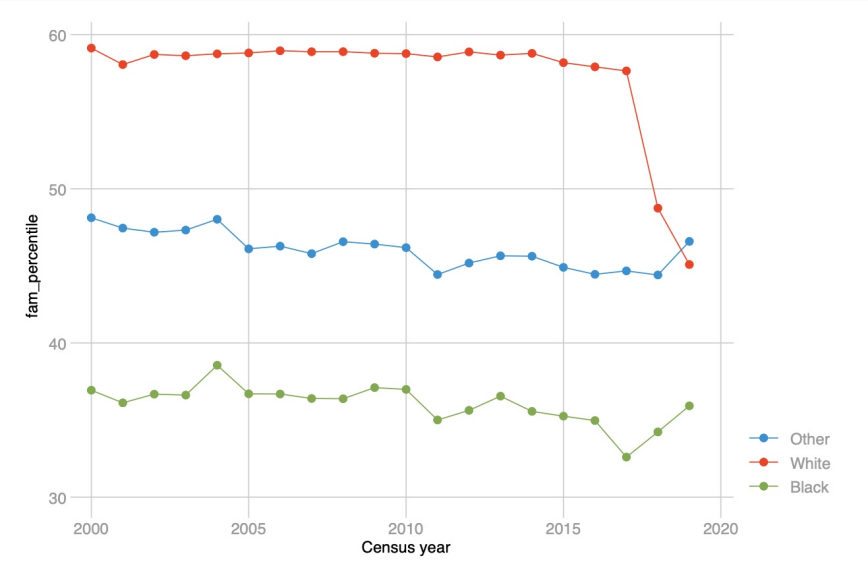
Current solution: Measure child income in 30s, when the age profile of the racial income gap seems to have flattened out

Future solution: Use two years of data, and reweight so that the average age is constant (e.g. for parent income of 1974 cohort, use 70% 1980 data and 30% 1990 data)

Lifecycle of Racial Income Gap: 1980 Cohort (Using Annual ACS After 2000)



Lifecycle of Racial Income Gap: 2000 Cohort (Using Annual ACS)



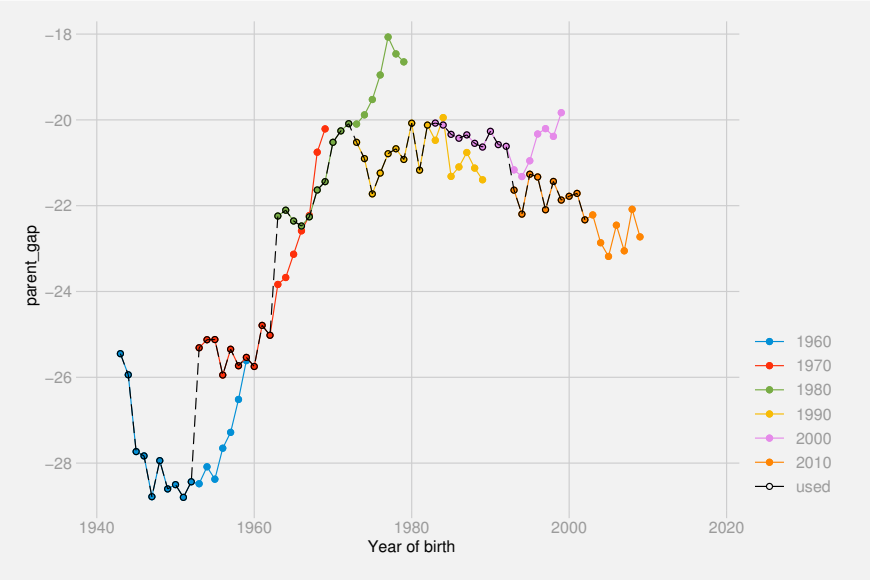
Results

Income gap, in percentile terms, shrank dramatically from 1960s to present; a more modest reversal recently

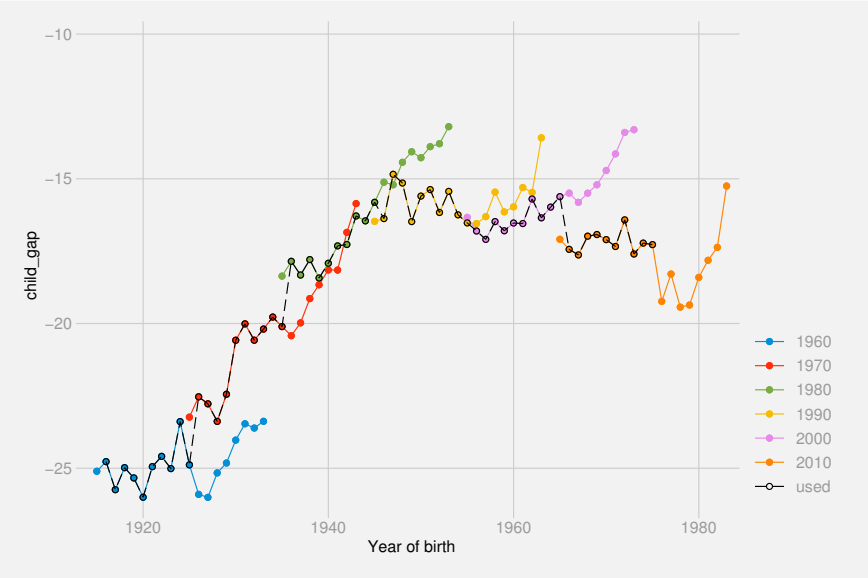
Mechanically, this drives the result: black parents were a lot poorer (in percentiles) than their children, so α_{Black} was higher then;

Question: How to interpret this? Is this saying that “things are getting worse” because mobility is going down somewhat? Or was previous upward mobility only an adjustment period?

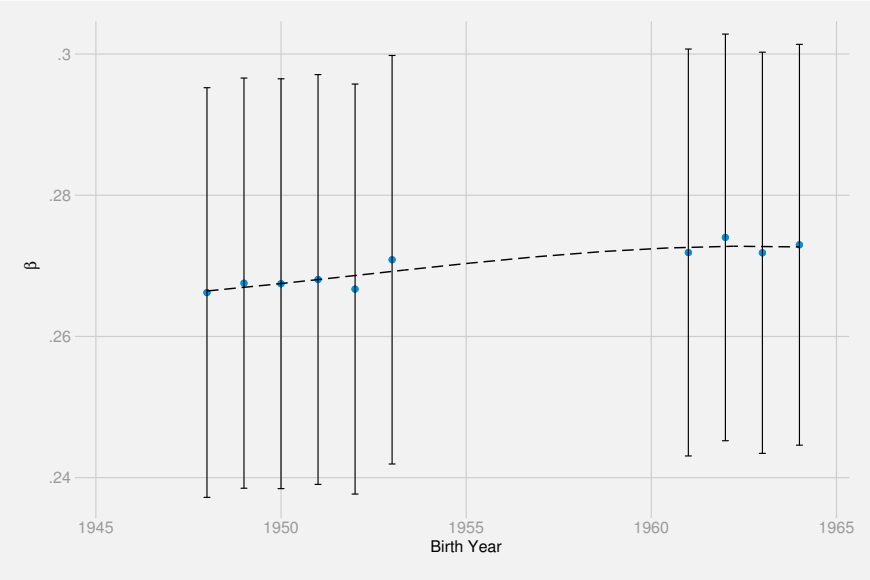
Black-White Gap for Parent Income Percentile



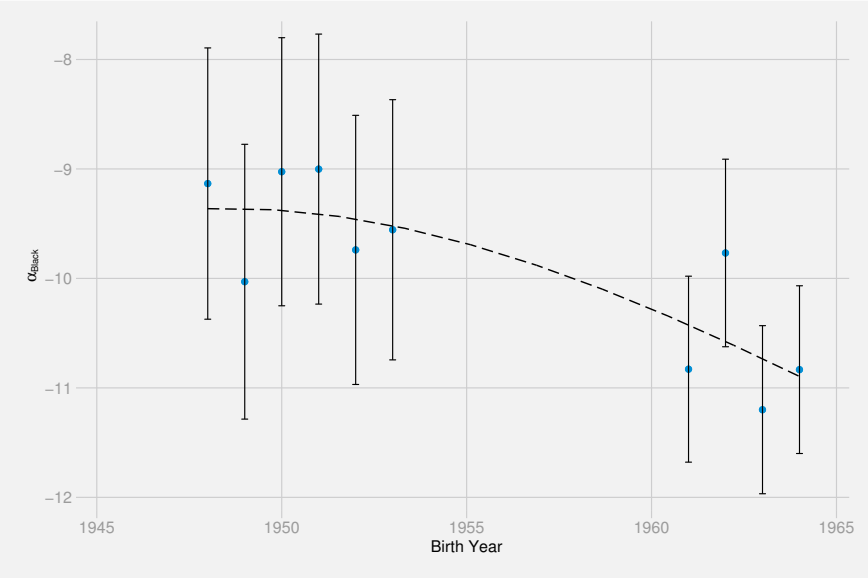
Black-White Gap for Child Income Percentile



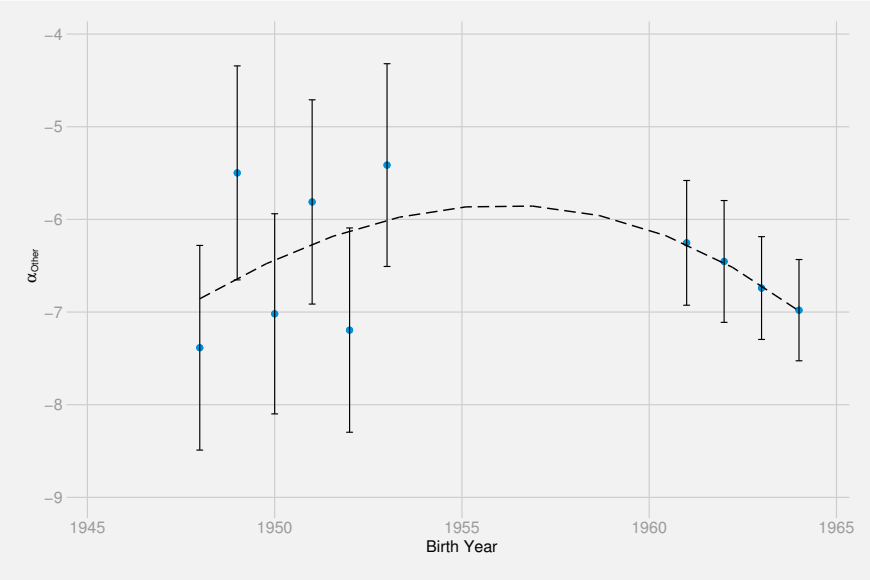
Trends in β



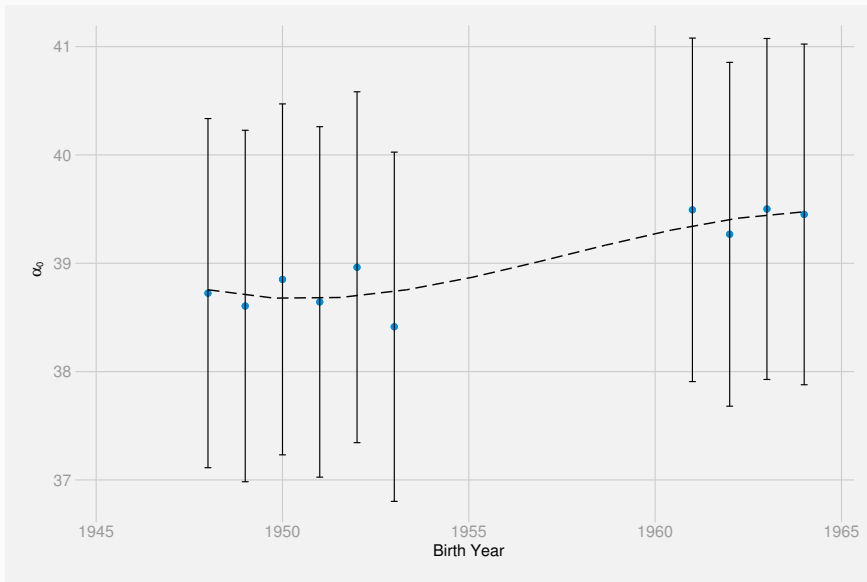
Trends in α_{Black}



Trends in α_{Other}



Trends in α_0



Extending our Estimates: 1945-1975

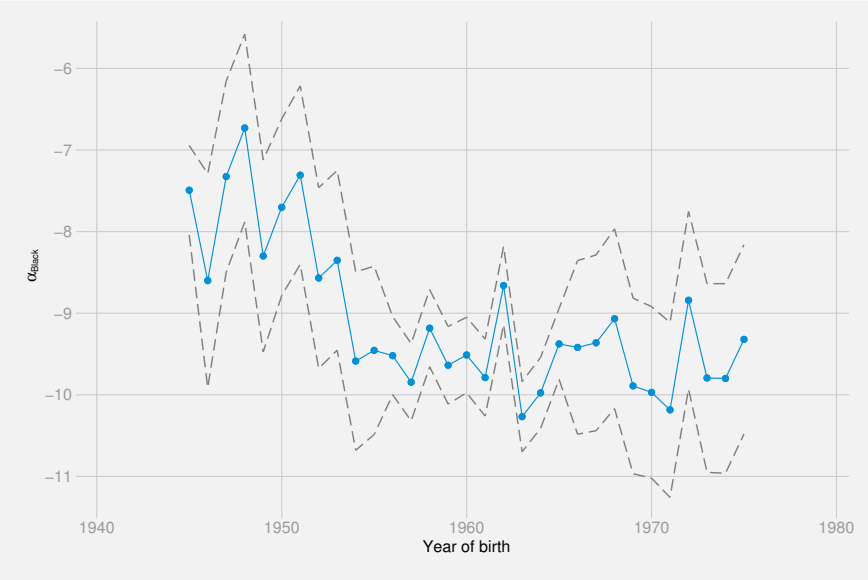
Davis and Mazumder only analyze a few years (10 years, although spans 1948-1964)

We can extend our estimates farther back and forward if we are willing to just impose a β , and drop the NLSY moments

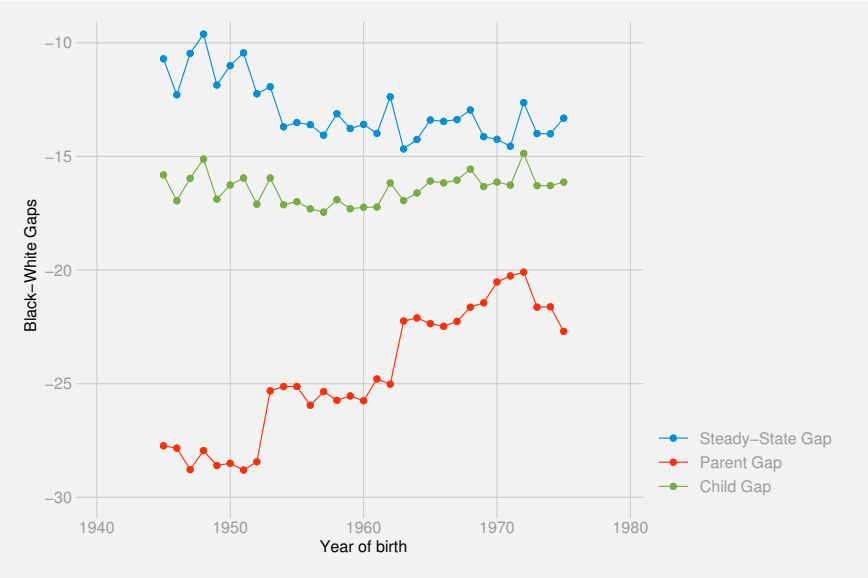
I will impose $\beta = 0.3$

Next Steps: Will be able to go ten years further back if I can get representative household income data from 1950, and can go nine years forward using ACS data through 2019.

Extended Trends in α_{Black}



Black-White Gaps, Parent, Child, and Steady-State



Conclusion

Intergenerational mobility does not appear to be getting better for black children in the US. If anything it is worsening somewhat.

Question of interpretation: Should the high mobility of the 40s and 50s be seen as “steady-state mobility,” or inherently transitory?

Next steps: Breakdown by gender, and extend series forward and backwards if possible. Could also construct regional series (not using NLSY).

Your feedback is very appreciated!