

Derenoncourt, E. (2019). Can you move to opportunity? Evidence from the Great Migration.

The author aims to investigate the causal effects of changes in racial composition during the Great Migration (1940-1970) on social mobility in the United States. In doing so, the paper tries to answer the question whether the Great Migration reduced the ability of Northern cities in the US to facilitate progress between generations for the black population. The main novelty that this paper advertises is the methodology it employs to identify this causal effect. An interaction between predicted southern outflow migration and pre 1940 migration location choices is used as an instrument (shift-share instrument) for increases in the share of Black people in the North of the US. In essence, this instrument can then be used as exogenous shocks in the racial composition in Northern cities, resulting in the identification of the causal effect. Using this identification strategy, the author shows that racial composition shocks in the North reduced the upward social mobility of people born in the 1980s in those regions, where black men were the most affected subgroup.

In order to measure upward mobility and migration patterns, the paper takes data from the well-known Opportunity Insights research group. This Harvard based group is working on the forefront of research on equality of opportunity in the US and has been a driving force in creating social mobility metrics. Using this data, the paper positions itself credibly in the literature. US census data is used to construct the migration patterns. The novelty of this paper partly lies in the combination of these two data sources that have not been used in this context before. In the appendix, the visualizations and graphical representation of the data make a compelling case and strengthen the story-line that the author puts forward. Overall, the paper does a good job at positioning itself in the current literature and using the data to set the stage for the empirical strategy.

While the methodology used by the author is convincing and has a strong foundation in related migration literature, there are two main areas where I feel the author could leave an even stronger impression.

One of the main assumptions underlying the identification strategy in this paper is that the instrument constructed is exogenous and not affected by confounding with respect to the target variable of interest (social upward mobility). The author defends this assumption by making the argument that migrants' origin locations are unrelated to shocks in their destinations, since Black migrants tended to move their livelihood to regions where their families and community members had previously migrated to (i.e. the author assumed there were no significant "pull factors" shaping migrants' decisions to move). In instrumenting for the increase in the share of Black people in the North, the author uses a shift-share approach. This means that the instrument is defined as the interaction between the pre-1940 migration patterns and the predicted outflow of migrants from the south. In order to identify causal effects, the pre-1940 migration patterns should be exogenous. This is a very strong assumption in my opinion and generally untestable. There are some references to literature that address some of these concerns, but none of those truly solve the problem (McKenzie, 2018). The author does perform multiple robustness checks strengthening the confidence for exogeneity, making a strong case for the strategy. However, in recognizing that industrialization (a.o.) could be a potential confounder (which is correctly controlled for), it suggests that there might be a plethora of other confounders not taken into account (see for example the World Bank blog in the references, where these identification assumptions are critically evaluated).

A second important element of the empirical strategy is constructing the predicted outmigration from counties in the South to commuting zones in the North based on characteristics of these locations. These predictions are then used as inputs for the shift-share instrument. Since there is a large group of county and commuting zone characteristics, the author relies on a ML method. The

novelty lies in using LASSO to perform variable selection as a pre-processing step, after which OLS is used to predict outmigration. Since these predictions are used in the construction of the instrument and not in the 2SLS estimation of the causal effects, there is still room for improvement in predictive performance. In terms of predictive performance, a Random Forest or other tree based methods (like gradient boosting) could be considered to create better predictions. Chen et al. (2020) explore this option of using flexible ML methods to construct technical instruments and can be used as guidelines for this paper too. Moreover, the LASSO makes much sense philosophically when the ground truth is sparse and we only expect few variables to be important for prediction. In the case of predicting outmigration, many factors might play a role, whose effects we might not all want to shrink to zero. Tree based methods still provide the flexibility to work in high dimensions (more variables than observations), while not imposing this sparsity restriction.

In order to put the results more in context, the author provided further analysis to investigate two mechanisms behind the decrease in social mobility. This especially puts the results in a broader social context and allows policy makers to interpret the results in a policy relevant way.

Lastly, in terms of delivery of the main conclusions, there could be some more attention to the fact that many findings relate to the 25th and 75th percentile of the upward mobility distribution. A broader analysis cannot be provided at times because of the availability of data, but in terms of presenting the conclusions, I would suggest to make more specific reference to these numbers and relate this to how the results should be interpreted.

Overall, though, the paper gives a very strong and convincing methodology and yields extremely interesting insights, while at the same time introducing some novelties to the field. More granularity can be added to the analysis on location effects and justify further research trajectories.

References:

Chen, J., Chen, D. L., & Lewis, G. (2020). Mostly harmless machine learning: learning optimal instruments in linear IV models. *arXiv preprint arXiv:2011.06158*.

Derenoncourt, E. (2022). Can you move to opportunity? Evidence from the Great Migration. *American Economic Review*, 112(2), 369-408.

Mckenzie, D. (2018, April 30). Rethinking identification under Bartik Shift-Share Instrument. *World Bank Blogs*. <https://blogs.worldbank.org/impactevaluations/rethinking-identification-under-bartik-shift-share-instrument>