Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

_____ _____

Roberson Beauvile Date

The Effects of Black Excellence on Labor-Market Outcomes and Migration

By

Roberson Beauvile

M.A. in Sociology, Emory University, 2023

M.A. in Applied Quantitative Research, New York University, 2020

B.S. in Physics and Sociology, 2019

Department of Sociology

Weihua An
Advisor

Timothy J. Dowd
Committee Member

Heeju Sohn
Committee Member

Kenneth A. Bollen
Committee Member

Accepted:

Kimberly Jacob Arriola, Ph.D, MPH

Dean of the James T. Laney School of Graduate Studies

Date

The Effects of Black Excellence on Labor-Market Outcomes and Migration

By

Roberson Beauvile

M.A. in Sociology, Emory University, 2023

M.A. in Applied Quantitative Research, New York University, 2020

B.S. in Physics and Sociology, 2019

Advisor: Weihua An, PhD

An abstract of

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in the Department of Sociology

2025

Abstract

The Effects of Black Excellence on Labor-Market Outcomes and Migration

By Roberson Beauvile, PhD

This dissertation uses the New Great Migration (1970-Present) to illustrate the inextricable link between black excellence, labor market outcomes, and black migration. Chapter 1 of this dissertation introduces the context behind this link, situating it within the New Great Migration paradigm. The next three chapters make four main contributions to the literature. In chapter 2, I found that black applicants who exuded excellence see their labor gradients doubled when compared to their white counterparts. That is, black applicants who exude excellence see the lowest callback rates among all applicants in cities such as New York but see the highest callback rates in cities such as Atlanta. In chapter 3, I mapped the transformation of community cultural wealth into economic capital using Structural Equation Modeling. I found that for children growing up in a household of black excellence, the six forms of capital (aspirational, navigational, resistance, social, linguistic, and familial capital) are collectively twice as important to economic and social mobility as the pathway through education. In chapter 4, I found that, all things being equal, black excellence almost triples the odds of migration within the US. Additionally, I found that in regard to the effect of black excellence on planning to move in the near future, black excellence is moderated by how strong respondent's immediate interpersonal networks were, nearly doubling the odds of planning to move in the near future. Thus, this dissertation contributes to the literature on internal mass migration by demonstrating that, in modern times, black excellence plays a key role in the mass migration of black people in the US.

The Effects of Black Excellence on Labor-Market Outcomes and Migration

By

Roberson Beauvile

M.A. in Sociology, Emory University, 2023

M.A. in Applied Quantitative Research, New York University, 2020

B.S. in Physics and Sociology, 2019

Advisor: Weihua An, PhD

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Sociology

2025

Acknowledgement

This dissertation is dedicated to my twin brother, Robed and my girlfriend Kiki – the two are the quintessential pillars of my life. I wouldn't be here without'em and couldn't imagine an existence devoid of'em.

I'd be remised if I didn't also thank my advisor Dr. Weihua An for his continuous mentorship and unwavering support. From our first meeting at visitation weekend in 2020 to today, I've never once felt like he didn't have my best interest at heart. Thank you for the great deal of knowledge you've shared with me, introducing me to Mathematical Sociology, and for always having my back. I will always consider him a great friend, now family, and an exceptional human being.

In addition, I'd like to thank Dr. Timothy J. Dowd for his mentorship and guidance. His lectures have been some of the most thought provoking I've ever attended. I credit him for my current understanding of the social world. I also have to thank Dr. Heeju Sohn for her constant uplifting words of encouragement and vital feedback. They have gotten me through some tough situations to say the least.

My many thanks go out to all the members of my dissertation committee: Dr. Weihua An, Dr. Timothy J. Dowd, Dr. Heeju Sohn, and Dr. Kennth A. Bollen. I am grateful for their time and effort, as well as the valuable feedback they've provided.

In closing, I would like to thank the staff and professors at Emory University Department of Sociology for all their help and support. I addition, I'm ever grateful for the support I received from the Laney Graduate School Fellowship and Centennial Scholar Fellowship from 2020-2025.

Table of Contents

List of Tables

List of Figures

Chapter 1

Main Introduction

Background

The Great Migration was one of the largest mass migration events in US history. Between the years of 1900 to 1990 nearly 7 million black Americans moved from the rural South to urban centers in the North, Midwest, and West (Tolnay, 2003). Most scholars break the Great Migration into two distinct phases. The First Great Migration saw 2.5 million southern-born black Americans move to Northern and Midwestern cities such as New York, Chicago, and Detroit from 1910 to 1950. However, black Americans have been leaving the South much earlier than this. The Second Great Migration saw an additional 4 million black Americans residing in non-Southern cites by 1980 (Tolnay, 2003). The economic, political, cultural, and demographic ethos of these urban centers and cities had been fundamentally reshaped by black migrants. But why did they leave the South in the first place?

During the late 19th century and early 20th century, black Americans in the South were exposed to unfathomable levels of lethal violence on two omnipresent fronts; first coming from the white mobs, composed primarily of white civilians, and second the white criminal justice system in the forms of lynchings and public legal executions of many forms (Ayers 1984; Shapiro 1988; Williamson 1984). It is truly astonishing the levels of legal unbridled violence black Americans had to endure during these times. In addition to violent persecution, black Americans during this time saw an incredible and deliberate suppression of their economic mobility, educational attainment, and voting rights by all manner of white people. Taken together, the Great Migration was an attempt to be relieved from the extraordinary obstacles to a decent life. But the out-flow of black Americans from the South did not last forever.

The migratory patterns of black Americans came to a halt in the closing decades of the 20th century and reversed. By the early 21st century the South had become a regional magnet for black Americans more than white Americans, particularly those with *higher education levels* (Frey 2005). The level of violence experienced by black Americans at the hands of whites had significantly subsided when compared to the early 20th century, but did not vanish. Moreover, the economic landscape of Southern cities had been burgeoning compared to non-Southern cities during the late 20th century as well. However, the increasing economic growth and reduced racism doesn't fully explain why blacks are more attracted to the South than whites. The main focus of my dissertation is to shed some light on the attractiveness of the South by the black diaspora. The South used to be a place of extreme violence and oppression for blacks of all walks of life. *Why is it then, that in modern times black folk are more attracted to the South than any other race or place in the US?*

The theoretical approaches to explaining migration are generally dominated by the use of "push" and "pull" factors to account for migratory behavior. That is, there exist attractive and repulsive forces that push and pull potential migrants towards and away from particular regions. Researchers tend to favor 3 main explanations for the mass black exodus from the US South and subsequent reversal: (1) economic forces such as regional wage differentials and expansion of employment opportunities in the North and other regions outside of the South for the Great Migration and the opposite for the New Great Migration; (2) social forces such as educational opportunities, racial violence and voter disenfranchisement; and (3) environmental changes such as floods or the boll weevil infestations (Tolany and Beck, 1990; Frey 2005). However, for simplicity, this paper will mainly focus on the two major forces stemming from *economic* and *social* factors as environmental forces do not seem to play a major role in the New Great Migration.

Moreover, this dissertation introduces a new pull factor that has not previously been talked about in the literature, that is the effects of black excellence.

Black excellence played a crucial role during the Great Migration and is inextricably linked to the New Great Migration as well. During the Great Migration, black folk were primarily driven by the prospects of escaping the extremely hostile environment in the South, as well as the potential economic and educational opportunities in the North and West. In fact, by the 1920s and 1930s, black Americans were experiencing an unprecedented level of socioeconomic growth, and cultural revolution in Harlem, New York known as the Negro Renaissance, today referred to as the Harlem Renaissance (Mitchell II 2010). By the 1960s and 1970s the concept of black excellence was created. The purpose of black excellence was to celebrate the achievement of black people demonstrating excellence through education, creativity, career, and politics as a non-violent weapon against racism and discrimination. Exuding black excellence was a source of pride and evidence that black people could achieve all the things white scholars, entrepreneurs, scientist, and political leaders claimed they were not physically and intellectually capable of.

Thus, black excellence is inextricably linked to the New Great Migration for several reasons. Black professionals, entrepreneurs, and highly educated black people are moving back to the South where the economic opportunities are booming. Black people moving back to the South want to build and sustain black-owned businesses, institutions, and communities. Cities like Atlanta, "The Black Mecca of the South" foster environments where Black professionals can network, support each other, and contribute to generational wealth. The migration of highly educated black people contributes to the strengthening of HBCUs (Historically Black Colleges and Universities) and the overall intellectual capital of the region. Schools like Spelman, Morehouse, and North Carolina A&T benefit from increased enrollment and alumni reinvestment,

reinforcing cycles of black academic excellence. Black excellence translates into greater political engagement and leadership in growing Southern cities, influencing policies on housing, education, and economic development. This shift has led to higher black representation in government, reshaping local and state political dynamics. Finally, many black professionals prioritize homeownership, and the South offers more affordable housing markets compared to Northern and Western cities.

This results in stronger black middle-class neighborhoods, fostering generational wealth and financial stability, as well a higher quality of life. Exploring all these topics in relation to migration is beyond the scope of this dissertation so I will focus on economic factors related to the expansion of employment opportunities and educational opportunities. Thus, this dissertation will primarily focus on the effects of black excellence on the black labor market, black education, and black migration.

Significance

Why is this research significant? Many black Americans moving back to the South are seeking better economic opportunities, lower costs of living, and job growth in cities like Atlanta, Houston, and Charlotte. This shift challenges the traditional narrative that the South is a place of economic stagnation for black folk and highlights how Southern cities have become hubs for black economic growth.

Many Black Americans are returning to places where their ancestors lived, fostering a cultural and historical reconnection. The movement is also leading to the revitalization of historically black communities, with new businesses, cultural institutions, and social networks forming. Black migration to Southern metros is reshaping urban and suburban development, with rising black

populations in suburban areas. Understanding these patterns can help policymakers address housing, education, and infrastructure needs.

The original Great Migration (1910–1970) saw millions of black Americans leave the South due to extreme levels of racial discrimination and economic hardship. The New Great Migration (1970s–present) signals a reversal, indicating that the South now offers better economic and social conditions for many black Americans.

As the black population grows in Southern cities, there is a need for housing policies that prevent displacement and gentrification. Understanding migration trends can guide investment in black-owned businesses and workforce development. Growing populations require better schools, public transit, and healthcare access. Although this research does not delve deeply into policy implications explicitly, understanding the mechanisms underlying black migration in relation to black excellence is important, given how their migratory patterns have and will continue to fundamentally reshape the ethos of the South.

Structure of Dissertation

This dissertation is broken up into three parts. The structure takes the form of three research papers exploring the negative effects of black excellence on the labor market outcomes (Chapter 2), a causal structural analysis of the effects of coming from a household of black excellence on educational aspirations, educational attainment, childhood and adult health, and adult socioeconomic status (Chapter 3), and the relationship between black excellence, tie strength, and migration (Chapter 4). Chapter 5 will be a discussion of my findings and their implications and concluding thoughts about the direction of future research in regard to the direction of research on black excellence in the coming years.

References

Ayers, E. L. (1984) Vengeance and Justice: Crime and Punish Nineteenth-Century American South. New York: Ox.

Frey, William H. 2005. "The New Great Migration: Black Americans' Return to the South, 1965–2000." Pp. 87–110 in *Redefining Urban and Suburban America: Evidence from Census 2000*. Washington, DC: Brookings Institution Press.

Mitchell II, Ernest J. 2010. "'Black Renaissance': A Brief History of the Concept." *Amerikastudien / American Studies* 55(4):641–65.

Shapiro, H. (1988) White Violence and Black Response: From Reconstruct to Montgomery. Amherst: University of Massachusetts.

Tolnay, Stewart E. 2003. "The African American 'Great Migration' and Beyond." *Annual Review of Sociology* 29(1):209–32.

Williamson, J. (I984) The Crucible of Race: Black-White Relation American South since Emancipation. New York: Oxford U.

Chapter 2

The Effects of Black Excellence on Labor-Market Outcomes: A Resume Audit Study on Race, Status, and Labor Gradients in Atlanta and New York City

Abstract

Black excellence is a term used to describe and celebrate black individuals, or black people in general, with high levels of achievement, success, or ability. This study demonstrates that the transfer of cultural capital to economic capital presumed by the assertation of black excellence may act as push factor for migration outside of predominantly black cities such as Atlanta. I used a resume audit to show that black applicants with high perceived socioeconomic status see less economic opportunities in New York City's labor market and are rewarded in Atlanta's labor Market. Moreover, I found that the labor market differentials between the two metropolitan areas were larger for black applicants living in relatively affluent neighborhoods than their white counterparts, as well as black and white applicants living in non-affluent neighborhoods. This implies that black excellence may be a strong push factor for high achieving black people to migrate back to Southern metropolitan areas. The results show that black applicants living in affluent neighborhoods in Atlanta had the highest call-back rates among all applicants at 12%, and black applicants living in affluent neighborhoods in NYC had the lowest call back rates among all applicants at 4%. Moreover, the differential in call-back rates between cities was highest for black applicants living in affluent neighborhoods at 8% and lowest for white applicants living in affluent neighborhoods at 4%.

Introduction

Following slavery, discrimination, oppression, and a Civil War, blacks had a plethora of reasons to leave the South in the early 20th century: destitute schools, economic dislocation, a dearth of political power, and whites finding a myriad of ways to demean, harass, and violently kill them (Alexander et al. 2017; Beck 2015). By 1910 blacks began migrating out of the South at an unprecedented rate (Black et al. 2015; Leibbrand et al. 2019; Stack 1997). From 1910-1970, the number of blacks that lived in the South was nearly cut in half, with a large percentage of black migrants residing in just four cities, New York, Chicago, Philadelphia, and Detroit (Tolnay 2003; White et al. 2005). By 1970 the "Great Migration," as it came to be called, came to a halt and scholars began noticing a reversal of the that trend (Adelman et al. 2000; Alexander et al 2017; Frey 2004; McHugh 1987; Sharkey 2015).

From 1950s to 1960s, black people embarked on the largest non-violent campaign against racial segregation, discrimination, and disenfranchisement in the US, namely, the Civil Rights Movement (Sundstrom 1994). Coming out of the Civil Rights Movement (and by extension, the South) was a cultural object created by black people specifically tailored to be a weapon against racism and centuries of subjugation, namely, black excellence. With some scholars calling it the "highest form of resistance" (Etienne 2004). Black excellence describes and celebrates black individuals, or black people in general, with high levels of achievement, success, or ability. This includes blacks with above average incomes, education, housing and neighborhood quality, health status, etc. By the 1970s the term "black excellence" was on the rise in the scientific literature (see next section). This rise coincided with the reversal of the Great Migration.

Not only were blacks leaving non-southern metropolitan areas at record numbers, but a large portion of them were also returning to the South, particularly Atlanta (Frey 2004). Arguably,

some of the factors that influenced black's migration patterns during the Great Migration may be the ones directing blacks back to the South today: there is increasing economic opportunities for high achieving blacks and an improving "racial climate" in Southern metropolitan areas (Tolnay 2003). It is important to note that, just as it was the case during the Great Migration, today, blacks and whites are similar in terms of out-migration patterns but differ substantially in regard to destination. Blacks and whites are both migrating out of northeastern metropolitans, with whites tending to settle in the South and West and the vast majority of blacks settling primarily in the Atlantic South (Frey 2004). Moreover, both groups have been following this pattern since the 1980s (Smelser, Willson, and Mitchel 2001; Toppo and Overberg 2015).

Atlanta's post-industrial economy has been burgeoning since the 1960s, "with substantial population growth, growth of jobs in the service sector, rising income, and rising employment" (Sharkey 2013:49). Meanwhile, according to some scholars, hiring discrimination, at the first point of hiring, has not changed since 1989 and seems to be increasing in the northeastern region (Quillian et al. 2017). Furthermore, "in 40 years not one [Chicago] community changed from black to white" (Sampson 2013:106). Suggesting that the stigma of disorder has irremovably stained black neighborhoods; neighborhoods in which middle-class blacks, poor blacks, and working-class blacks tend to live in or around (Patterson 1997; Pattillo-McCoy 1999; Wilson 1978).

Just as they did at the beginning of the Great Migration, middle- and upper-class blacks are leading the way today in the 'New Great Migration,' that is, the majority of black migrants leaving non-southern metropolitans are highly educated and have the economic means to migrate across the country, subsequently residing primarily in Atlanta (Frey 2004). Gaddis (2014) showed that blacks from elite colleges have only as much economic opportunity as whites from less selective universities. This implies that the economic opportunities for high achieving blacks are

diminished due to their racial status, in spite of their educational attainment. Therefore, based on the neoclassical macroeconomic theory of migration (Lewis 1954), the unprecedented migration of educated blacks, compared to their white counterpart, out of non-Southern metropolitan areas would suggest that there is more economic opportunity in the South.

However, economic opportunity and an improving racial climate may not be the only factors motivating blacks' exodus from non-Southern metropolitan areas. Migrants tend to reside in regions where their racial or ethnic group are highly concentrated in a historically significant way (Smelser, Wilson, Mitchell 2001). Atlanta, for example, has always held a special place in the hearts and minds of blacks. It has long been referred to as the "black mecca of the South," where "black folks have more, live better, accomplish more and deal with whites more effectively than they do anywhere else in the South – or north" (Garland 1971:152).

Furthermore, when one takes into consideration cost of living, particularly housing cost, Georgia's residents have a higher median income than New York's residents (Mulhere 2018). Taken together, there seems to be four main reasons that are influencing black's current migration pattern: labor market differentials between non-southern and southern regions, an improving racial climate in the South, the higher concentration of blacks in the South, and lower cost of living in the South. This paper focuses on the former three reasons: labor market differentials, improving racial climate, and a higher concentration of blacks in the South.

This paper seeks to compare the effects of perceived socioeconomic status (SES) on labor market differentials for blacks and whites in the two metropolitan areas where the vast majority of black migrants are migrating to and from, Atlanta and New York City, respectively (Frey 2004). No study to date has looked at how one's perceived SES affects labor market differentials. Labor market differentials are an important factor when explaining the migration decisions of black

people. In particular, how being perceived as a high achieving black person affects one's labor market outcomes in metropolitan areas where black excellence is celebrated heavily but occupy different historically significant spaces.

Thus, this paper seeks to contribute to the literature by measuring how black excellence affects labor market differentials and labor market outcomes in Atlanta and New York City using a resume audit. These differentials serve as a metric for the potential push and pull forces induced by black excellence. The first section situates black excellence as a form of *resistance capital* and defines the concept of "black excellence" as a cultural object, created and mainly recognized by black people as a form of resistance against racism. The next section will discuss labor market differentials as a measure of labor market opportunities that can be a factor in predicting migration patterns. Following this section, I will elaborate on the factors that affect migration decisions. Particularly, how the concentration of black people heavily affects black migrant's destination.

Black Excellence as Capital

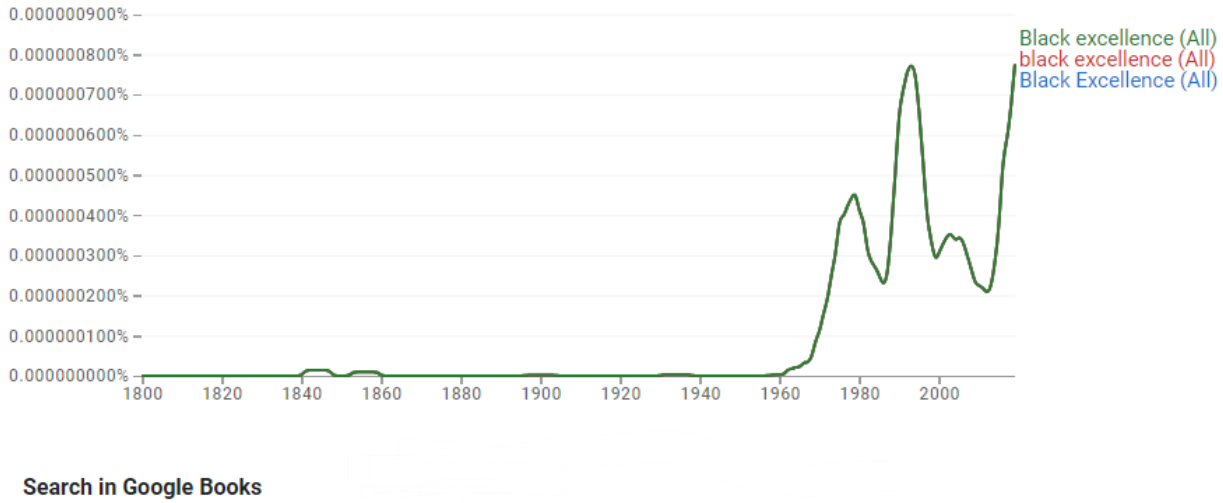
"Black Excellence" as Culture (Cultural Capital???)

The Civil Rights Act of 1968 was undoubtedly a response to a string of decades long protests by, and the sometimes brutal and highly public murders of, black people. It was the U.S. government's attempt to combat discrimination on the bases of race, religion, and national origin. However, coming out of the Civil Rights Movement was a cultural object created by black people specifically tailored to be a weapon against racism, namely, black excellence. With some scholars calling it the "highest form of resistance" (Etienne 2004).

Given the history of black people in America, the concept of black excellence is relatively new. In fact, the term itself did not start seeing significant usage in scientific literature until 1992.

Figure 2.1 shows the yearly percentage of the term “black excellence” in Google’s American Text Corpora, containing roughly 155 billion words from published books and articles spanning the years 1800 to 2019. There were three peaks of interest in 1978, 1992, and 2019.

Figure 2.1. Yearly Percentage of “black excellence” (case-sensitive) in Google’s Text Corpora



From figure 2.1, the increase of the term “black excellence” began to rise at the tail end of the Civil Rights movement, peaking in 1978. This is expected as the concept of black excellence saw its inception during this period. During this time, black excellence came in the form of leadership roles in the Civil Rights Movement. Some examples include Jo Ann Robinson (the head of the Montgomery Women’s Political Council and a key organizer in the Montgomery Bus Boycott), Fannie Lou Hamer (lead Freedom Summer of 1964), John Lewis (leader of the Student Nonviolent Coordinating Committee), and Edward Brooke (the first African American to serve in the U.S. senate in 1967) were all pivotal to the movement and exemplified resistance through excellence (Gaddis 2023).

Fourteen years after the initial peak in 1978, there was an additional peak in 1992. This peak began a little after 1980, which coincides with the proliferation of black film and television

in what has been called “the golden age of black film and television” in the early 90s and 2000s (Rivers 2021). During this period, television shows with all black cast such as In Living Color, Martin, and Living Single became powerhouses for network television, competing with shows such as Friends, Seinfeld, Mad About You, and ER, with black viewership in particular.

The prosperity of black entertainment during this time undoubtedly came from the unprecedented success of The Cosby Show. Where a show about an upper middle class black family, with a doctor and lawyer as parents, demonstrated that an all-black cast could indeed be a successful sitcom. Moreover, black audiences worldwide uniformly reported feelings of pride when watching the show (Havens 2000). This is due to the fact that the portrayal of blacks on the show broke from the traditionally derisive portrayal of blacks in white popular culture.

After the golden age of black entertainment, there was a precipitous drop in the use of the term black excellence until a sharp rise in 2011. One year before the second election of the 44th president of the United States, Barack Obama (America’s first black president). Micheal and Barack Obama represented the pinnacle of black excellence for many reasons. To many black Americans the Obama’s were the manifestation of black excellence, changed their thoughts of themselves and their nation, and exemplified black romantic and familial love. Following 2011, the use of the term black excellence has exploded and has continued to rise ever since.

In this paper, I define black excellence to be synonymous with any and all achievements that subverts the deleterious effects of racism. That is, achievements by black people that challenge the stereotypes associated with their income, education, housing, health, etc.

Black Excellence as Community Cultural Wealth

The mechanism by which black excellence subverts the effects of racism and discrimination is simple. The collective and vicarious celebration of the embodiment of excellence, that is a high achieving black person, leads to adoration and emulation; what psychologist call vicarious reinforcement (Horsburgh & Ippolito 2018). Yosso's (2005) *community cultural wealth* framework, which critiques Pierre Bourdieu's traditional concept of cultural capital that centers white, middle-class norms as the standard for social mobility, highlights the unique forms of capital developed by marginalized communities to navigate and resist oppression. Yosso (2005) introduced 6 forms of cultural wealth: (1) Aspirational Capital – maintaining hope and resilience despite structural barriers; (2) Linguistic Capital – bilingualism, storytelling, and communication skills shaped by cultural heritage; (3) Familial Capital – knowledge and values passed down through family and community networks; (4) Social Capital – community networks that provide support, resources, and advocacy; (5) Navigational Capital – skills to successfully move through institutions not designed for marginalized groups; (6) Resistance Capital – knowledge and strategies gained through collective struggles against oppression.

That is, through the struggles of being black in a highly racialized world, high achieving black people who exude excellence, in one way or another, provide a beacon of hope, a blueprint, networks, language, skills, and resistance for those black folks still on the fringes of success. This is due to the fact that high achieving black people see a multitude of social and economic rewards from their success that can then be passed down to the overall community.

In the majority of cases, high achieving black people tend to have better life outcomes than non-high achieving blacks and whites. For example, the Journal of Blacks in Higher Education (JBHE) Foundation has long shown that higher education is a major mechanism for closing the black-white income gap. In 2007 they published findings that illustrated the median income for

blacks at each educational level. They found that in 2004, the median income for black adults with a high school diploma was roughly $10,000 and the median income for black adults with a Ph.D. was roughly $74,000. This substantial difference in income did not go unnoticed. According to US Census Bureau (2023), by 2021, the percent of black people with a bachelor's degree nearly doubled from the early 2000s.

Despite the income boost that highly educated blacks receive, they still fall short in the labor market compared to equally educated whites. Gaddis (2015) looked at how black applicants with degrees from an elite university faired in the labor market compared to their white counterparts. He found that white applicants with a degree from an elite university had the highest response rate from potential employers at 17.5%, followed by black applicants with a degree from an elite university at 12.9%, white applicants with a degree from a less selective university at 11.4%, and, lastly, black applicants with a degree from a less selective university at 6.5%. These results make it clear that race and college selectivity affect labor-market outcomes among bachelor's degree holders.

Black candidates with an elite education came second only to white candidates with an elite education. This is expected as candidates with a degree from a well-known elite university would necessarily attenuate the negative effects of race when compared to candidates of any race from less selective schools. Thus, it is unclear how black excellence itself affects labor-market outcomes given that employers value candidates from highly selective schools. This leads to the research question: (1) *How does black excellence affect labor-market outcomes?*

Labor Gradients

The oldest and widely used theory of international and internal migration was developed to explain labor migration in the process of economic development (Lewis 1954; Ranis and Fei 1961; Massey et al. 1993). According to this theory, migration is a result of geographic differences in the supply of and demand for labor. Regions with a large supply of labor relative to capital have a low equilibrium market wage. Regions with a limited supply of labor relative to capital have a high market wage. Thus, the differential in wages relative to capital induces the flow of labor from high-labor regions relative to capital to low-labor regions relative to capital.

However, this paper focuses on the differential in labor-market opportunities, as measured by the likelihood of receiving a call back from a potential employer. I call this differential in labor-market opportunities "*labor gradients*." This is similar to temperature gradients in physics where heat flows from warmer regions of space to cooler regions of space over some fixed distance. Similarly, when labor-market opportunities begin to rise in the capital-poor region, they fall in the capital-rich region until equilibrium is reached and migration due to labor gradients reflect just the pecuniary and psychological cost of movement.

Mirroring this flow of labor is a flow of investment capital from the capital-rich to capital-poor regions. Juxtaposing this model onto the trends seen in the New Great Migration, it is clear that blacks' migration patterns maybe the result of the *labor gradients* described above. Moreover, capital investments are also flowing into the South from non-southern regions and can be seen in the growth of industries such as agribusiness, defense, technology, real estate and construction, and tourism. This leads to my second research question: (2) given the different migration patterns for blacks and whites, do black migrants have larger labor gradients than whites? (3) And are these labor gradients affected by class? If current black migration patterns reflect *labor gradients*, then

these differentials should be clearly measured via a resume audit. A resume audit would also be able to highlight the effects of perceived SES on labor market outcomes and labor gradients.

Migration

The 'New Great Migration' and Labor Gradients

By 1975 the economic advantages afforded to black migrants of the Great Migration had diminished, thus, the New Great Migration represented labor reallocation in response to market needs (Heckman 1990; Ritchey 1976). As the low-skilled jobs in the northern region began to dwindle, black migrants of the New Great Migration started taking advantage of the burgeoning service economy of the South. This reasoning is underpinned by the classical competitive model of factor mobility. That is, labor demands and supply are in equilibrium, with labor supply adjusting in response to the relative real-wage rates between areas (Heckman 1990). The underlying assumption is that the differential between the northeastern regions and southeastern regions are diverging causing a *labor gradient*. Today, this differential is referred to as "regional price parity" (Mulhere 2018).

Regional price parity standardizes income, effectively gives an indication of how far their money will go for a given region. The higher a state's price-parity score, the more residents will pay for goods and services such as housing, food, and transportation. New York's price parity score is among the highest in the country at 115.3. This means that New York residents can expect to pay over 15% more for goods and services than the average American. However, Georgia's price-parity score is 92.6. Indicating that Georgia's residents can expect to pay almost 7.5% less for goods and services than the national average. Thus, New Great Migration migrants may also

be responding to the break in equilibrium or diverging real-wage rates that lead to the increasing *labor gradient* between Atlanta and New York City.

The Changing Southern Racial Climate

Frey (2004) and Tolnay (2003) assert that the racial climate in Atlanta is improving. This would cause a dampening effect on the social closure mechanism generating the discriminatory acts of employers. Thus, creating an avenue of economic opportunity in Atlanta as employers see less of a difference between themselves and blacks at the first point of hiring. The improving racial climate in Atlanta can also serve to dampen the effect of neighborhoods perceived to be disordered or the effect of perceiving blacks from affluent neighborhoods in a more negative light than whites from similar neighborhoods. Effects that I hypothesize influence hiring practices of employers. According to Sampson (2013:13), "neighborhoods differ dramatically in their quality, feel, sights, sounds, and smells… [but have] a diversity of behaviors and social actions that cluster together in space and that define the social organization of the city."

This means that one's neighborhood could intrinsically connotate a whole host of socioeconomic characteristics that employers might recognize in an address on a resume which then in turn may affect labor market outcomes and labor gradients. This leads to my fourth and fifth research questions: (4) do neighborhoods affect labor market outcomes and labor gradients? And (5) do call back rates reflect an improving racial climate in Atlanta?

Discrimination

Scholars have noticed that blacks have been migrating back to the South for years (McHugh 1987). DeWaard, Curtis, and Fuguitt (2016) illuminate a "key" component to migration patterns, that is, new migrants' integration is dependent on duration of residency. According to the

authors, scholars of the 'New Great Migration' have omitted this detail from their discourse. The authors argue that black migrants' duration of residence differs by birth cohort and birthplace, that is, black migrants', born in the South, duration of residency is higher than black migrants not born in the South. This would suggest that black migrants residing in the South, and staying long enough to be well integrated, were born there, thus, are the larger proportion of black migrants residing in southern regions and taking part in the New Great Migration. But the authors also found that there is a net positive and non-trivial migration of non-southern blacks to the South, that is, non-southern born black migrants are experiencing an affinity towards the South despite never living there. This means that black migrants are not merely residing in the South because it is where they grew up but because there exist more salient "push" factors that lead black migrants towards southern metropolitan areas like Atlanta, and away from non-southern metropolitan areas like New York City.

One factor scholars attribute to the in-migration of blacks to the South from metropolitans across the nation is that economic opportunities are burgeoning in the South while diminishing in metropolitans outside the southern region (Alexander et al. 2017; Frey 2004; McHugh 1987; Stack 1996; White et al. 2005). Some scholars have simultaneously concluded that racial discrimination was waning, and race was no longer a significant factor in determining blacks' life chances (Patterson 1997; Wilson 1978). Surveys indicated that whites favored equal treatment for everyone, regardless of race (Schuman 1998). However, Quillian et al. (2017) concluded that hiring discrimination, at the first point of hiring, has not changed since the 1980s.

Moreover, the authors found that hiring discrimination, at the first point of hiring, is significantly increasing in the northeastern region. Coupled with the blossoming economy of the South, this steady increase in hiring discrimination in northeastern regions may serve as a "push"

factor influencing blacks' levels of economic opportunities in northeastern regions, leading to their migration to more economically prosperous regions. This study seeks to test these levels of economic opportunity and hiring discrimination for blacks using a field experiment method, namely, a resume audit. This leads to my sixth research question: (6) to what extent, if any, is there hiring discrimination in Atlanta or New York City? No study to date compares the degree of discrimination at the first point of hiring towards blacks in historically significant cities such as New York City and Atlanta. This paper seeks to fill this gap in the literature by comparing call-back rates of blacks, in relation to whites, in these two metropolitan areas.

On the other hand, Darolia et al. (2015) sent fictitious resumes to online job postings in seven cities across six occupational categories, racializing surnames for black, Hispanic, and white applicants. The authors asserted that no hiring discrimination at the first point of hiring was detected. Regarding surnames selected, the authors selected Washington and Jefferson to indicate African American applicants and Anderson to indicate the white applicant. Although the authors note that 90 to 75 percent of individuals with the surnames Washington are African American and 70 percent of individuals with the surname Anderson are white, a fundamental limitation of this study is that the authors combined the surnames with the matched first names, Chloe and Ryan. The names Ryan Jefferson and Chloe Washington may not be strong enough indicators of black applicants which would explain why no racial differences were found in this study. If so, their purpose as indicators of race is necessarily moot. To avoid this limitation, I used a distinctively black sounding first name, Tyrone, for the black applicant and a distinctively white sounding first name, Connor, for the white applicant. According to Fryer and Levitt (2004) the first name Tyrone is popular among blacks but is virtually unheard of for whites and the first name Connor is

distinctively white. I will be matching them with the surnames Jefferson and Anderson, respectively.

Method

Resume Audit

A resume audit is a field experiment that measures the total number of interview requests made by real employers to fictitious applicants; this number divided by the total number of fictitious applications sent to said employers' results in a "*call-back rate*" for each applicant. In general, the experiment went as follows: first, I created 8 fictitious resumes that corresponded to 8 separate email addresses that I created through the free email service provided by Google.com – I also created one master email account for information recovery purposes; second, using these resumes, I created 8 accounts on 3 job matching websites (Careerbuilder.com, Glassdoor.com, and Indeed.com) and sent out 2,150 applications responding to job postings on these websites; and finally, I read over 4,800 emails (an average of over 600 emails per fictitious applicant) in order to construct a database that included variables that indicated the date I applied to a job posting, the date I received a *call-back*, job posting position title, *call-back* status (coded as "1" for a call-back and "0" otherwise), race (black or white), neighborhood type (ordered or disordered), city or neighborhood of the job posting, metropolitan area of the job posting, and website used to send out resumes.

Indicators of Race and Status

Resumes were virtually identical except for indicators that signaled race and residence. To signal race, I used a first and a last name that are distinctly white sounding, Connor Anderson, and a first and last name that are distinctly black sounding, Tyrone Jefferson. To ascertain

neighborhoods that are perceived to be disordered (relatively poor), I watched months' worth of news clips aired on CBS Atlanta and CBS New York, compared rent maps for each metropolitan area, and utilized "broken windows" theory in conjunction with google maps to verify signs of a disordered neighborhoods. Based on the news clips, neighborhoods in The Bronx can be perceived as disordered, as almost all news reporting's of The Bronx were related to crime or disorderly behavior. The entire borough also saw some of the lowest average rent in New York City (see Figure 1A in appendix). However, in Atlanta news reporting on crimes were placed by counties which made this method of ascertaining disordered neighborhoods ineffective.

Utilizing rent maps (see Figure 2A in the appendix), I observed that clusters of neighborhoods in the northwestern part of Atlanta were indicative of a disordered neighborhood by virtue of their extremely low average rent relative to most areas in Atlanta. The logic behind this is simple, as the renting market in areas with low average rent might be in a state of low demand because the way they appear is perceived as an area of high crime and disorder. I verified this assumption using google maps and "broken windows" theory. The theory indicates that small forms of public disorder led to serious crime and urban decay (Kelling and Coles 1996), and "visual cues such as graffiti, public intoxication, garbage, and abandoned cars are thought to attract criminal offenders, who assume from these cues that residents are indifferent to what goes on in the neighborhood" (Sampson and Raudenbush 2004). I noted a plethora of signs of disorder ranging from destitute sidewalks and houses to loitering and graffiti. Thus, I chose Bankhead, Atlanta as a neighborhood of perceived disorder. Using the same techniques and logic as in the previous paragraph, I concluded that the neighborhoods indicative of little or no perceived disorder are the Upper East Side, New York and Midtown, Atlanta given their relatively high rent. Addresses displayed on applicants' resumes from one of these four areas will indicate whether an

applicant resides in a perceived disordered or perceived ordered space. This is why I created a total of 8 separate resumes; 2 racial indicators and 2 neighborhood indicators in two metropolitan areas correspond to 8 resumes, 4 applicants, 8 emails accounts, 24 job matching website accounts, etc.

Addresses in Manhattan can be placed readily by employers because of New York City's grid system, with East 59th street to East 96th street being indicative of the Upper East Side. Street addresses in The Bronx all have indicators that place residents in this borough. Atlanta's neighborhoods, however, have almost all named roads with the expectation of Midtown and its surrounding areas. Almost all numbered roads with the Northeast (NE) indicator are in Midtown. Given the large variety of road names in Atlanta, large roads like Boulevards might be more distinct because they would most likely be frequented by more drivers and people taking public transportation. Bankhead, Atlanta is circumvented by two large Boulevards that bisect the entire low-rent region (Joseph E. Boone Blvd NW and Joseph E. Lowery Blvd NW) and a major highway (Donald Lee Hollowell Pkwy NW), enclosing this neighborhood within three large roads. Thus, I used the following addresses to increase the likelihood an employer could place the address as disordered and ordered: E 149th & E 152nd St, Bronx, NY 10455 and Joseph E. Lowery Blvd NW, Atlanta, GA 30314 (disordered neighborhood); E 66th & E 77th St, New York, NY 10075 and 8th & 10th St NE, Atlanta, GA 30309 (ordered/more affluent neighborhood).

Randomization

All resumes indicated that applicants were awarded a B.S. degree in finance with a minor in mathematics in May 2019. Applicants held positions such as Administrative Assistants in Brooklyn and central Atlanta, and applicants were all part of the same honors society. Tyrone, however graduated from the University of Florida and Connor graduated from Florida State University, two similarly ranked large state schools (U.S. News & World Report). Both applicants

had identical work experience, language and programming skills, and awards. Before applying to job postings, I created 8 groups of applicants. Group 1 was composed of Tyrone living in the perceived ordered neighborhood in New York City and Connor living in the perceived disordered neighborhood in New York City. I named this group CT-OD-NY, where C stands for Connor, T stands for Tyrone, O stands for ordered neighborhood and denotes the residence of Connor, D stands for disordered neighborhood and denotes the residence of Tyrone, and NY represents New York City, the metropolitan area for this pair. Using this naming convention, the 8 groups were as follows: 1) CT-OD-NY, 2) CT-DO-NY, 3) CT-DD-NY, 4), CT-OO-NY, 1A) CT-OD-ATL, 2A) CT-DO-ATL, CT-DO-NY 3A) CT-DD-ATL, 4A) CT-OO-ATL, where ATL represents Atlanta, the metropolitan area for the pair.

Before applying to any job in Atlanta or New York City, I used python to randomly select a number from the set {1,2,3,4}. When 1 was selected, I sent out the group 1 or 1A pair of resumes to the job posting in their respective areas; when 2 was selected, I sent the group 2 or 2A pair of resumes to the job posting; etc. This method effectively randomly assigned neighborhood indicators while controlling the black-white pair comparison for resumes sent out. I sent out a total 2,150 resumes to 1,075 employers across the Atlanta and New York City metropolitan areas over a 2-month period. The high internal and construct validity of this method allowed me to directly measure labor market differentials by race and neighborhood residence (comparing levels of economic opportunity) and hiring discrimination based on racial or neighborhood status at the first point of hiring.

Over 95% of the advertised positions I applied to were part of the office and administrative support, sales, management, material moving, and healthcare support occupational groups. Table 2.1 shows the top 6 job titles by number of applications sent in descending order of frequency.

Other titles included Office Clerk, Inventory Clerk, A/R Clerk, Health Aide, Sales Assistant, Warehouse Associate, etc. Office and administrative support occupations had the largest concentration of black workers than any other occupational group (14% compared to 12% of workers as a whole),

Table 2.1. Percentage of Jobs Titles

Job Title	Frequency	Percent	Cumulative Percentage
Administrative Assistant	396	18	18
Sales Representative	386	18	36
Manager	273	13	49
Accountant	229	11	60
Custom Service Representative	171	8	75
Sales Associate	158	7	77
Other†	537	25	100

†Subsumed in the other category are over 80 more job titles. Each position makes up between 1 and .05 percent of the entire sample. However, the vast majority of these positions fall under the office and administrative support, sales, material moving, service, management, and healthcare support occupational groups.

with the sales (10% black concentration compared to 11%), material moving (10% black concentration compared to 6%), management (7% black concentration compared to 12%), and food preparation and serving (7% black concentration compared to 6%) occupational groups rounding out the list of the top five occupational groups with the largest concentration of black or African American workers, ages 16 and older, in 2016 (Rolen and Toosi 2018). Moreover, for all workers, the office and administrative support group was projected to have the largest number of annual occupational openings, on average, from 2016 to 2026 (Rolen and Toosi 2018). 95% of the jobs I applied to fell into one of these occupational groups.

Table 2.2 shows the distribution of locations for each job opening I applied to from descending order according to frequency. More than half of the open positions were in the cities of Atlanta and New York, with almost a quarter of the job openings I applied to being on the island of Manhattan, NY, or Atlanta. In fact, these two areas combine make up almost half of the job

Table 2.2 Percentage of Jobs' Location

Job Location	Frequency	Percent	Cumulative Percentage
Manhattan	514	24	24
Atlanta	479	23	47
Brooklyn	96	4	51
Queens	46	2	53
The Bronx	22	1	54
Staten Island	16	1	55
Other†	977	45	100

†Subsumed in the other categories are over 100 cities in and around Atlanta and New York City, including Jersey City, Decatur, Alpharetta, etc. All cities are within a commutable distance from the metropolitan areas of Atlanta and New York City. Each city in this category makes up 1 to .05 percent of the entire sample.

locations I applied to. Other areas include cities like Jersey City, NJ, Decatur, GA, and Alpharetta, GA. Every job opening was within a reasonable commute distance from applicants' residences. Both tables' statistics bolster the external validity of this study. By focusing on jobs that are within a commutable distance from applicants' residence, as well as positions that fall into occupational groups that most workers are a part of, I experimentally controlled for the average characteristics of workers in the labor market, as well as the positions the average worker would hold when applying to these jobs (Rolen and Toosi 2018). Moreover, applicants were given relatively average educations at large state schools, along with work histories that mirror the composition of the average worker in the workforce. Thus, this study represents the experience of the average person in the American labor market.

Sample Size Estimation

To determine an estimate of the sample size I would need to detect each effect given its existence, I conducted a two-group independent sample power analysis using the statistical software R. For any power analysis, if three of four parameters are known the fourth one can be determined using the other three. The four parameters are as follows: each group's sample size, the effect size d, the power level p (the probability of detecting an effect given that it exists), and the significance level $sig.level$ (the probability of finding an effect that is not there, or the p-value). I set $sig.level$ to be .05, as is the convention in sociology and p to be .9, a relatively large power level. The effect size, or Cohen's d, is given by the equation below:

$$d = \frac{|\bar{x}_1 - \bar{x}_2|}{\sigma_{pd}}$$

Where σ_{pd} = Pooled Standard Deviation and is given by:

$$\sigma_{pd} = \sqrt{\frac{SD_1^2 + SD_2^2}{2}}.$$

SD_i^2 are the squared of the standard deviations for each group, where i = 1,2. In an experiment like a resume audit, call-backs are coded as either a "1" or "0." Thus, the standard deviation for resume audits SD_{Ri} for each group (1 and 2) can be expressed in terms of only their means, or the call-back rates for each group – expressed by the last equation below:

$$SD_{Ri} = \sqrt{\frac{n_i}{n_i - 1}(\bar{x}_i - \bar{x}_i^2)},$$

$$for\ n_i \geq 51\ and\ 1 < \frac{n_i}{n_i - 1} \leq 1.01,$$

$$SD_{Ri} = \sqrt{1.01(\bar{x}_i - \bar{x}_i^2)}.$$

$\bar{x}_i$ are the call-back rates for each group. Inserting the final equation above into the formula for the effect size I derived Cohen's d for resume audits d_R:

$$d_R = \sqrt{\frac{2|\bar{x}_1 - \bar{x}_2|^2}{1.01(\bar{x}_1 + \bar{x}_2 - \bar{x}_1^2 - \bar{x}_2^2)}}.$$

It is clear from the formula above that the effect size is only dependent on the magnitude of each group's call-back rate. This means that in order to determine a sample size that will detect an effect 90% percent of the time with a 95% confidence interval I had to consider sample sizes for call-back rates ranging from .03 to .15 in my analysis, reasonable estimates for call-back rates according to previous studies (Gaddis's 2014; Pager's 2003; Nunley's et al. 2015; Quillian's et al. 2017). I stratified the range of call-back rates in order to account for the small to large effect sizes. The results of the power analysis indicate that to detect a large effect size requires a sample size of more than 79 employers per resume pair, which amounts to over 316 employers in total. However, to detect a small effect size would require a sample size of more than 234 employers per resume pair, which amounts to 936 employers. Thus, my sample size of 1,075 employers should be more than enough to detect large and small effects 90% of the time, given that they exist.

Results

Analysis

Table 3 shows the call-back rates and differences in call-back rates for the black and white applicants between metropolitan areas. Applicants in New York City received a lower percentage of call-backs irrespective of race or neighborhood residence. The black applicant living in a relatively affluent neighborhood in New York City received the lowest percentage of call-backs, with the highest percentage of call-backs going to the black and white applicant living in a relatively affluent and relatively poor neighborhood in Atlanta, respectively. In general, applicants received a request for an interview at a rate of 11 requests per 100 applications sent in Atlanta and

5 requests per 100 applications sent in New York City. The difference in these call-back rates was highly significant at a p-value less than .0005 but the effect size $d_R = .22$, a relatively small effect size. Nevertheless, this result suggests that applicants' probability of receiving a call-back is doubled merely by living in Atlanta compared to New York City.

Table 2.3 Call-back Rate Differential by Metropolitan Area (Labor Gradient), Race, and Class.

	Atlanta	New York City	Labor Gradient
Total Resumes Sent	968	1,182	
Call-back Rates	.11	.05	.06****
Black Applicant			
Total Resumes Sent	490	602	
Call-back Rates	.11	.05	.06****
Ordered Neighborhood Resumes Sent	243	301	
Ordered Neighborhood Call-back Rates	.12	.04	.08****
Disordered Neighborhood Resumes Sent	247	301	
Disordered Neighborhood Call-back Rates	.10	.05	.05*
White Applicant			
Total Resumes Sent	478	580	
Call-back Rates	.11	.06	.05**
Ordered Neighborhood Resumes Sent	237	279	
Ordered Neighborhood Call-back Rates	.10	.06	.04
Disordered Neighborhood Resumes Sent	241	301	
Disordered Neighborhood Call-back Rates	.12	.05	.07**

Note: p-values are two-tailed.
**p<.05. **p<.01. ***p<.001. ****p<.0005*

The difference in call-back rates in terms of an absolute difference in call-back rates between metropolitan areas, significance level, and effect size for the black applicant was equivalent to the whole sample. Both the significance level and effect size decreased for the white applicant; the p-value being less than .01 and $d_R = .18$, respectively. After, controlling for race

and neighborhood residence, the significant difference in call-back rates between metropolitan areas for the white applicant living in a relatively affluent neighborhood completely vanished.

Table 2.4 Differences in Call-back Rates by Race and Class

	Black Applicant	White Applicant	Difference
Total Resumes Sent	1,092	1,058	
Call-back Rates	0.08	0.08	0
Atlanta			
Total Resumes Sent	490	478	
Call-back Rates	0.11	0.11	0
Ordered Neighborhood Resumes Sent	243	237	
Ordered Neighborhood Call-back Rates	0.12	0.10	0.02
Disordered Neighborhood Resumes Sent	247	241	
Disordered Neighborhood Call-back Rates	0.10	0.12	0.02
New York City			
Total Resumes Sent	602	580	
Call-back Rates	0.05	0.06	0.01
Ordered Neighborhood Resumes Sent	301	279	
Ordered Neighborhood Call-back Rates	0.04	0.06	0.02
Disordered Neighborhood Resumes Sent	241	301	
Disordered Neighborhood Call-back Rates	0.05	0.05	0

Note: p-values are two-tailed.
**p<.05 **p<.01 ***p<.001 ****p<.0005*

However, the call-back rate for the black applicant living in a relatively affluent neighborhood was 3 times greater in Atlanta than in New York City and was significant at p<.0005. Moreover, the effect size was at a moderate $d_R = .30$. When only looking at the difference in call-back rates with respect to neighborhood type and irrespective of race, the results mirrored that of the overall sample in terms of significance level and effect size: p<.0005 and $d_R = .22$. This result couple with the vanishing of significant difference in call-back rates for the white applicant

living in an affluent neighborhood suggest neighborhoods have an effect on labor market differentials and is only observable when race and place are controlled for simultaneously.

Table 4 shows the call-back rates and differences in call-back rates for the white and black applicants. In general, the white and black applicants received an equivalent percentage of call-backs, each at 8 percent. This is also the case for the black and white applicant in Atlanta, with both applicants receiving 11 call-backs for every 100 applications sent out. However, the black applicant saw the lowest percentage of call-backs in New York City, particularly when they resided in a relatively affluent neighborhood. The differences in call-back rates, however, were not significant, with some differences being zero. For example, black and white applicants living in Atlanta saw identical call-back rates. Moreover, the differences in call-back rates that were non-zero were not significant with effect sizes that were practically zero. This result did not change after controlling for neighborhood residence.

Logistic Regression Results

Table 5 shows the logistic regression results. The coefficients were originally expressed in the form of log odds, that is, the log of an odds ratio. An odds ratio is simply the quotient of two odds. I chose to display the coefficients in the form of odds ratios because it is difficult to interpret the effect size based on coefficients that are in the form of log odds. Thus, converting the log odds coefficients into odds ratios coefficients effectively makes these odd ratios standardized effect size statistics if the predictor variables are dichotomous (Ialongo 2016), which is the case for this study. The independent variables related to the metropolitan area seemed to be the strongest predictors out of the four effect types of a call-back after controlling for all other variables. The odds of receiving a call-back is 2.32 times greater if applicants apply in Atlanta than in New York City,

after controlling for race, neighborhood type, and job matching website. This result is highly significant at p<.0005 with a moderate effect size of 2.32.

Table 2.5 Odds Ratio Logistic Regression Models Predicting Call-backs

Effect Type	Model 1	Model 2	Model 3
Black Applicant	0.927	0.929	1.07
	(0.15)	(0.15)	(0.397)
Metropolitan Area			
Atlanta		2.32***	2.69**
		(0.389)	(0.891)
Neighborhood Type			
Ordered Neighborhood		0.99	1.33
		(0.161)	(0.479)
Job Matching Website			
Glassdoor.com	.484*	.454**	0.452**
	(0.139)	(0.131)	0.131
Indeed.com	0.838	0.797	0.795
	(0.214)	(0.204)	(0.204)
Interactions			
BlackxAtlanta			0.764
			(0.359)
BlackxOD			0.554
			(0.295)
AtlantaxOD			0.592
			(0.276)
BlackxAtlantaxOD			2.78†
			(1.87)
N	2150	2150	2150

Note: Standard errors are in parentheses
*†p<.10 *p<.05 **p<.01 ***p<.001*

These results, along with the results found in Table 3, indicate a clear but moderate labor market differential between New York City and Atlanta. Particularly, in the office and administrative support, sales, management, material moving, and healthcare support occupational

groups, 95% call-backs came from employers advertising positions in one of these occupational groups.

Discussion

According to the neoclassical macroeconomic theory of migration, labor market differentials induce a flow of labor from high-labor supply and relatively low capital regions to low-labor supply and relatively high capital regions over some fixed distance, a process I call *labor gradients.* Similar to temperature differentials in physics that induce the flow of heat from warmer regions of space to cooler regions of space (measured as temperature gradients), labor gradients can be observed by measuring the difference in labor market opportunities (or difference in call-back rates) between two geographical regions.

I found that the labor gradient for the black applicant living in a relatively affluent neighborhood was the largest of all applicants while the white applicant living in a relatively affluent neighborhood had the lowest labor gradient, however this result was not significant. This mirrors the findings of Frey (2004), who observed that blacks who exuded excellence (by virtue of their education) are leading the way out of all migrants heading to the South, while white migrants residing in the South make up a much smaller portion of Southern migrants. These results support the idea that being black and excellent may be a driving force for many black migrants choosing to reside in the South. Moreover, class seems to affect labor gradients differently for blacks and whites depending on their class. With affluent blacks seeing the most labor market opportunities in the South and affluent whites seeing the least. The mechanism explaining these results is illuminated by the call-back rates for each city.

Quillian et al. (2017) observed no change in the level of hiring discrimination in the U.S. against African Americans since 1989, as well as an increase in the levels of hiring discrimination in northeastern regions. However, I cannot confirm nor deny these claims with the results from this study, as seen in Table 4, no difference was statistically significant. This can mean one of two things. Either there was no difference in callback rates for black and white applicants or the difference was not large enough to detect statistically. More research needs to be done in this area to answer the sixth research question conclusively.

However, the findings of zero difference in total call-back rates in Atlanta and 20% less total call-backs for blacks than whites in New York City would suggest that hiring discrimination at the first point of hiring is still prevalent in the northeastern metropolitan areas. Moreover, when class is included in the analysis, affluent lacks received 20% more call-backs than affluent whites in Atlanta. In New York City, however, affluent blacks received 50% less call-backs than affluent whites. Additionally, affluent blacks received 20% less call-backs than poor blacks in New York City. The results seem to suggest that black excellence negatively affects labor market outcomes in New York City but positively affects labor market outcomes in Atlanta.

Given that these results were not statistically significant, I cannot conclusively say that black excellence is penalized outside of Southern metropolitan areas. However, this is the first study pointing to the possibility that black excellence can have negative effects on black people. The results may indicate that there is an underlying mechanism, such that black excellence as a cultural object is mostly celebrated in the South but penalized outside of the South. This would logically follow as the concept of black excellence was born out of the Civil Rights Movement, and, thus, born out of the South. Moreover, it is possible that whites want to see blacks do better socioeconomically but not as good or better than themselves. More research should be focused on

the negative effects of black excellence as there is a growing conversation, not yet prevalent in the literature, on the negative effects of black excellence on black people in general.

The challenge and limitation of this study was ensuring employer's recognition of neighborhoods of perceived disorder in Atlanta through applicants' addresses. Street addresses in Atlanta do not have clear indications of one's neighborhood of residence, with the exception of Midtown. Moreover, roads are distinguished by names and do not follow any obvious pattern. Thus, I am reliant on the employer's knowledgeability of Atlanta's streets to place applicants of the perceived disordered neighborhoods based on their addresses.

Another limitation of this study is that it cannot distinguish why blacks might have more economic opportunities in Atlanta than New York City. I cannot know why employers might perceive blacks as more employable in Atlanta than in New York City. Thus, while the audit study establishes a difference in economic opportunity between metropolitan areas, it is impossible to measure the underlying mechanism for these effects using this methodology. What is known however is that the South has seen increasing levels of economic development since the 1970s. This influx of capital coupled with an improving racial climate for middle-class blacks could account for the significantly higher call-back rates for the blacks in Atlanta than New York City.

Another limitation of this study is that it cannot explain why the trend of the Great Migration has reverse. Migration is a complex subject matter involving many free parameters. To better understand why migrants of the New Great Migration are residing in Atlanta, perhaps an ethnographical or survey methodology can illuminate the factors influencing their behavior. Moreover, this study, also, cannot measure the effect their migration has on Atlanta's and the nation's economy, labor market, housing market, etc., or on their quality of life compared to their previous residence outside of the South.

Despite the intrinsic complexity of the process of migration, it is a physical process and, thus, can be broken down to a series of components that can be measured and modeled using thermodynamic principles in physics. This would make the physics and mathematical concept of gradients an ideal model because migration has long been thought of as the movement of labor induced by labor market differentials between regions of high-labor supply relative to capital and regions of low-labor supply relative to capital. Thinking of migration in terms of gradients alleviates some of the inherent complexities associated with the migration process by firstly, associating migration patterns with underlying differentials over set distances that are both measurable.

Secondly, migration gradients have no way of predicting individual or average migration patterns. This means that in order to determine the underlying gradient, an associated migration pattern must be present. Lastly, migration patterns would imply and underlying gradient, but gradients may not imply migration patterns. I hypothesize that there are three types of gradients: labor gradients, network gradients, and security gradients. Labor gradients would be characterized by differentials in labor supply relative to capital between regions. Network gradients would be characterized by differentials in weak ties relative to strong ties between regions. And security gradients would be characterized by differentials in deleterious entities (crime, natural disasters, etc.) relative to non-deleterious entities between regions. I surmise that thinking of migration in terms of gradients illuminates the procedural and chaotic (in the thermodynamic sense) nature of migration, as well as providing a bases for studying migration in a systematic way. This way of thinking could help shape a single, self-contained theory of migration.

References

Adelman, Robert M., Chris Morett, and Stewart E. Tolnay. (2000). "Homeward Bound: The Return Migration of Southern-Born Black Women, 1940 to 1990." *Sociological Spectrum,* 20,433–63.

Alexander, J. Trent, Christine Leibbrand, Catherine G. Massey, and Stewart E. Tolnay. (2017). "Second-Generation Outcomes of the Great Migration." *Demography* 54,2249–71.

Anon. (2019). CBS46 Crime Alert. *Wild Shootout Caught on Camera*. Retrieved October 19, 2019 (https://www.youtube.com/watch?v=MHe8aULOxzI).

Anon. 2007. "Higher Education Is the Major Force in Closing the Black-White Income Gap." *The Journal of Blacks in Higher Education* (39):26.

Beck, E. M. (2015). "Judge Lynch Denied: Combating Mob Violence in the American South, 1877–1950." *Southern Cultures,* 21(2),117–39.

Black, Dan A., Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor. (2015). "The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South." *American Economic Review,* 105(2),477–503.

Bureau, US Census. 2023. "Black Education on the Rise." *Census.Gov*. Retrieved October 16, 2023 (https://www.census.gov/library/visualizations/2023/comm/black-education.html).

Darolia, Rajeev, Cory Koedel, Paco Martorell, Katie Wilson, and Francisco Perez-Arce. (2015). "Race and Gender Effects on Employer Interest in Job Applicants: New Evidence from a Resume Field Experiment." *Applied Economics Letters,* 23(12),853–56.

Etienne, Germaine. (2004). "Excellence is the highest form of resistance: African American reformers in the pre -Civil War *North." *Doctoral Dissertations Available from Proquest*.

AAI3152692. https://scholarworks.umass.edu/dissertations/AAI3152692

Frey, William H. (2004). *The New Great Migration: Black Americans' Return to the South, 1965–2000*. New York, NY: Living Cities: The National Community Development Initiative. Retrieved October 11, 2019 (https://www.brookings.edu/wp-content/uploads/2016/06/20040524_Frey.pdf)

Fryer, Roland J. and Steven D. Levitt. (2004). "The Causes and Consequences of Distinctively Black Names." *The Quarterly Journal of Economics,* 119(3),767–805.

Gaddis, S. Michael. (2014). "Discrimination in the Credential Society: An Audit Study of Race and College Selectivity in the Labor Market." *Social Forces,* 93(4),1451–79.

Ghaddis, Vincent. 2023. "2023: Resistance through Black Excellence: Its History, Meaning, and Leaders." *Elgin Community College*. Retrieved October 9, 2023 (https://elgin.edu/news/ecc-stories/2023/resistance-through-black-excellence-its-history-meaning-and-leaders.php#:~:text=Jo%20Ann%20Robinson%2C%20the%20head,of%20Freedom%20Summer%20of%201964.)

Havens, Timothy. 2000. "`the Biggest Show in the World': Race and the Global Popularity of The Cosby Show." *Media, Culture & Society* 22(4):371–91.

Henry, Michael. (2004). *Race, Poverty, and Domestic Policy*. New Haven, CT: Yale University Press.

Horsburgh, Jo and Kate Ippolito. 2018. "A Skill to Be Worked at: Using Social Learning Theory to Explore the Process of Learning from Role Models in Clinical Settings." *BMC Medical Education* 18(1).

Ialongo, Cristiano. (2016). “Understanding the Effect Size and Its Measures.” *Biochemia Medica,* 150–63.

Kelling, George and Catherine Coles. (1996). *Fixing Broken Windows: Restoring Order and Reducing Crime in Our Communities*. New York, NY: Free Press

Leibbrand, Christine, Catherine Massey, J. Trent Alexander, and Stewart Tolnay. (2019). “Neighborhood Attainment Outcomes for Children of the Great Migration.” *American Journal of Sociology,* 125(1),141–83.

Lewis, W. Arthur. (1954). “Economic Development with Unlimited Supplies of Labour.” *The Manchester School*, 22(2),139–91.

Massey, Douglas S., Joaquin Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino, and J. Edward Taylor. (1993). “Theories of International Migration: A Review and Appraisal.” *Population and Development Review,* 19(3),431.

McHugh, Kevin E. (1987). “Black Migration Reversal in the United States.” *Geographical Review,* 77(2),171–82.

Mulhere, Kaitlin. (2018). “The Average Income in Every State—and What It's Really Worth.” *Money,* 47(4),67–70.

Nunley, John M., Adam Pugh, Nicholas Romero, and R. Alan Seals. (2015). “Racial Discrimination in the Labor Market for Recent College Graduates: Evidence from a Field Experiment.” *The B.E. Journal of Economic Analysis & Policy,* 15(3),1093–1125.

Pager, Devah. (2003). “The Mark of a Criminal Record.” *American Journal of Sociology* 108(5),937–75.

Pattillo-McCoy, Mary. (1999). *Black Picket Fences: Privilege and Peril Among the Black Middle Class*. Chicago, IL: University of Chicago Press.

Quillian, Lincoln, Devah Pager, Ole Hexel, and Arnfinn H. Midtbøen. (2017). "Meta-Analysis of Field Experiments Shows No Change in Racial Discrimination in Hiring over Time." *Proceedings of the National Academy of Sciences,* 114(41),10870–75.

Ranis, Gustav, and J.C.H. Fei. (1961). "A theory of economic development*." American Economic Review* 51, 533-565.

Ritchey, P. Neal. (1976). "Explanations of Migration." *Annual Review of Sociology,* 2,363–404.

Rivers, Kendall. 2021. "Black Excellence: Why the 90's and Early 2000's Were the True Golden Ages for Black Film And..." *Medium*. Retrieved October 9, 2023 (https://kendallrivers.medium.com/black-excellence-why-the-90s-and-early-2000s-were-the-true-golden-ages-for-black-film-and-c50044711039).

Rolen, Emily and Mitra Toosi. (2018). "Blacks in The Labor Force." *U.S. Bureau of Labor Statistics*. Retrieved (https://www.bls.gov/careeroutlook/2018/article/mobile/blacks-in-the-labor-force.htm).

Sampson, Robert J. (2013). *Great American City: Chicago and the Enduring Neighborhood Effect*. Chicago, IL: University of Chicago Press.

Sampson, Robert J. and Stephen W. Raudenbush. (2004). "Seeing Disorder: Neighborhood Stigma and the Social Construction of 'Broken Windows.'" *Social Psychology Quarterly,* 67(4),319–42.

Schuman, Howard. (1998). *Racial Attitudes in America: Trends and Interpretations*. Cambridge, MA: Harvard University Press.

Sharkey, Patrick. (2013). *Stuck in Place: Urban Neighborhoods and the End of Progress Toward Racial Equality*. Chicago: University of Chicago Press.

Sharkey, Patrick. (2015). “Geographic Migration of Black and White Families Over Four Generations.” *Demography,* 52(1),209–31.

Smelser, Neil, William Julius Wilson, and Faith Mitchell. (2001). *America Becoming: Racial Trends and Their Consequences*. Vol. I. Washington, D.C.: National Academy Press.

Stack, Carol B. (1996). *Call to Home: African Americans Reclaim the Rural South*. New York: Basic Books.

Sundstrom, William A. (1994). “The Color Line: Racial Norms and Discrimination in Urban Labor Markets, 1910–1950.” *The Journal of Economic History*, 54(2),382–96.

The Best National Universities in America. *U.S. News & World Report*, U.S. News & World Report, www.usnews.com/best-colleges/rankings/national-universities.

Tolnay, Stewart E. (2003). “The African American ‘Great Migration’ and Beyond.” *Annual Review of Sociology,* 29,209–32.

Tomaskovic-Devey, Donald and Dustin Avent-Holt. (2019). *Relational Inequalities: An Organizational Approach*. New York, NY: Oxford University Press.

Toppo, Greg and Paul Overberg. (2015). “After Nearly 100 Years, Great Migration Begins Reversal.” *USA Today*. Retrieved May 14, 2020 (https://www.usatoday.com/story/news/nation/2015/02/02/census-great-migration-reversal/21818127/).

White, Katherine J. Curtis, Kyle Crowder, Stewart E. Tolnay, and Robert M. Adelman. (2005). “Race, Gender, and Marriage: Destination Selection During the Great Migration.” *Demography,* 42(2),215–41.

Wilson, William J. (1978). *The Declining Significance of Race: Blacks and Changing American Institutions*. Chicago, IL: University of Chicago Press.

Wilson, William J. (2012). *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. Chicago, IL: University of Chicago Press.

Yosso, Tara J. 2020. “Whose Culture Has Capital? A Critical Race Theory Discussion of Community Cultural Wealth.” *Critical Race Theory in Education* 114–36.

Chapter 3

A Causal Structural Analysis of the Effects of Black Excellence on Household Income

Abstract

This study examines the causal structure underpinning the effect of black excellence has on household income. Specifically, this study looks at the total effect of having black parents or guardians who are highly educated on one's future household income. I found that respondents with black parents holding bachelor's degrees saw the most socioeconomic stability and upward mobility on average. Moreover, *educational aspirations* played a larger role in the total effect of childhood socioeconomic status (SES) on future household income for respondents with black parents holding graduate degrees than any other group in this study. This is the first study to explicitly map the transformation of aspirational capital into economic capital within the community cultural wealth framework as a function of black excellence. Finally, the results suggest that childhood health plays a nearly negligible role in the total effect of childhood SES on future household income.

Introduction

A plethora of studies have looked at how one's childhood socioeconomic status (SES) predicts one's life outcomes, specifically, outcomes related to one's adult SES and overall health (see Grossman 2006 for a review). However, few studies have directly analyzed the causal pathway that connects this relationship and consider racial difference by class. That is, does coming from a black household have differing effects on individuals' life outcomes depending on their class? Most studies tend to focus on the direct relationship between childhood SES and adult SES, childhood SES and adult health (Phelan, Link, and Tehranifar 2010), or the indirect relationships that make up this relationship. For example, Currie (2009) looked at the causal relationship between parental SES and childhood health, as well as the relationship between childhood health and adult SES. Chau et al. (2016) and Hjorth et al. (2016) looked at the relationship between childhood health and childhood (high school completion) to adult educational attainment. Grioux (2005) looked at the relationship between educational attainment and adult SES.

The study that does analyze the relationship between childhood SES and adult SES directly (see Diagram 2 in Haller and Portes 2015), as well as theorize the causal pathways between this relationship (Lareau 1987), tend to neglect the profound effect that childhood health has on educational attainment as well as the role of racially stratified class. That is, the lower, middle, and upper class within racial groups. Currie's (2009) survey finds multiple examples in the literature that demonstrate a causal relationship between child health and future educational attainment; suggesting that childhood health plays an important role in the intergenerational transmission of SES.

Haller and Portes (2015), used the Wisconsin Model of Educational and Early Occupational Attainment to construct a path diagram that illustrated the causal pathway between

childhood SES and occupational attainment. Two things were missing from this analysis. First, this model does not account for the effect of childhood health on educational attainment. This paper intends to analyze the causal structure between childhood SES and adult SES/health, while accounting for the effect of childhood health. Second, Haller and Portes (2015) did not analyze the path diagram directly. That is, they did not derive the structural equation for the total effect (sometimes referred to as gross effect) associated with the path diagram. Instead, they relied on calculations made by Blau and Duncan (1967) to quantify the model (See Table 4.1, Haller and Portes 2015). Thus, this paper also intends to derive the structural equation for the total effect directly using calculus.

Like many social scientists, Haller and Portes (2015) adopted the calculations made by Blau and Duncan (1967) to quantitatively demonstrate the contributions of the direct and indirect effect of the total effect of father's occupation and education on children's early in life occupational attainment. This means that the path diagrams used by Haller and Portes (2015) and the works of many social scientists tend to be underutilized. The power of path diagrams can be unlocked with an analysis of the total effect associated with any given pathway in the path diagram. Thus, fully understanding the strengths and weaknesses of the causal chain requires an explicit calculation of the total effect associated with the causal pathway.

This paper intends to address these problems by first introducing childhood health as a variable in the causal pathway between childhood SES and adult SES. This will be done by constructing the casual structure of this relationship using a path diagram and relevant literature. Second, after constructing the path diagram for all households, this paper will specify models by race (black households and white households) and a racially stratified group by class (black households with parents educated at the college and graduate levels). In this study, households

with black parents who are college educated are defined as “households of black excellence.” Finally, this paper will derive the total effect analytically for pathways of interest in a path diagram using calculus to determine the contribution of the direct, indirect, spurious, and unanalyzed effects.

Constructing the Path Diagram

To build the path diagram illustrating the casual pathways of interest, this study begins by considering the causal relationships between one’s childhood SES and adult SES and overall health. The relationship between childhood SES and one’s life outcomes have been widely studied in the social science and medical literature (see Grossman 2006, Nobel et al. 2015 for reviews). Moreover, one’s early life experiences are greatly shaped by their physical and socioeconomic wellbeing, which are primarily related to the SES of one’s parent(s) and/or guardian(s). This implies that parents SES dictates childhood SES. Which, in turn, implies the consideration of directionality and parental SES in the causal pathway of interest.

Class and SES frequently exhibit significant "stickiness" across generations. That is, individuals frequently remain in the same economic class as their parents. This phenomenon is particularly evident in the black community, who face systemic barriers that hinder upward mobility. Research routinely shows that black children are more likely to experience downward economic mobility compared to their white counterparts, even when born into middle- or upper-class families. Factors such as residential segregation, disparities in educational opportunities, and labor market discrimination contribute to this persistent intergenerational immobility (Chetty et al. 2018).

However, black excellence highlights the remarkable achievements of black individuals who overcome systemic obstacles. These individual successes do not always translate into sustained intergenerational wealth. Historical discriminatory policies, such as redlining and unequal access to credit, have limited black families' ability to accumulate and transfer wealth across generations. Even when black people attain professional success, the absence of inherited wealth and ongoing systemic inequities can impede the transmission of economic advantages to their descendants (Massey 2016).

Educational attainment is often viewed as a pathway to upward mobility, and many black people have leveraged academic success to improve their SES. However, studies show that even with higher education credentials, black people often face a persistent racial wealth gap. For example, black college graduates tend to have higher student loan debt and lower family wealth compared to white graduates, limiting their capacity for wealth accumulation and intergenerational transfer (Addo, Houle, and Simon 2016). This means that while education can facilitate individual advancement, structural barriers continue to obstruct the broader economic mobility of black families.

The literature specifies two causal mechanisms underpinning this relationship: the hypothesis of *social causation* ("positive" direction) and the hypothesis of *social selection/"drift"* ("negative" direction) (Mossakowski 2014). The mechanism of social causation posits that those experiencing economic troubles have an increased risk for subsequent poor health; this relationship is defined as the positive causal direction. Antithetically, following the negative causal direction, the mechanism of social selection/drift asserts that one's health can reduce socioeconomic mobility, leading individuals to "drift" into the lower social class and reduce their personal income. Many scholars tend to focus on the positive causal direction, emphasizing the social causation

hypothesis. In fact, some scholars (Phelan, Link, and Tehranifar 2010) even argue that SES is a "fundamental" cause of health outcomes. This study, however, emphasizes both directions simultaneously by also considering each mechanism at different points in the causal pathways.

This study, then, considers the factors between childhood SES and adult SES and overall health. Recently, scholars have noted that the casual pathways between these relationships are difficult to pin down (Kim and Radoias 2019). Phelan et al (2010) demonstrated that the relationship between SES and health has a multitude of mediating factors because SES subsumes money, knowledge, prestige, power, and beneficial social connections that have a preventative effect on poor health outcomes. This implies that the pathway between early life SES and adult overall health includes many other important factors. For example, Currie (2009) demonstrated that there is an abundant amount of evidence suggesting a strong causal relationship between parental SES and child health and child health and future SES (as measured by educational attainment and labor market outcomes). Moreover, many scholars have asserted that education can boost one's SES (Giroux 2005). This means that educational attainment preludes adult SES and health (contributing to both simultaneously) and follows childhood SES and health (both contributing simultaneously) in the causal pathway. Hence, this study considers educational attainment in the causal pathway of interest.

A corollary of the causal pathway between parental SES, childhood health, and educational attainment is that adults who had poorer health in their childhood would be less educated than adults who were healthier during their childhood. There are two reasons for this. First, poorer health leads to less frequent school attendance and lower performance. Chau et al. (2016) found that school difficulties and absenteeism were strongly associated with health-related problems such as poor physical health, psychological health, social relationships, and living environment.

Additionally, poor health leads to less encouragement to continue schooling given absenteeism and subsequent poor academic performance. The Wisconsin Model of Educational and Early Occupational Attainment illustrate the known causal pathways between parental SES and occupational attainment (see Diagram 2 in Haller and Portes 2015). Lareau's (1987) study qualitatively demonstrated the theoretical mechanisms underpinning these pathways. Lareau (1987) demonstrated how class backgrounds can affects parents and students think about schooling. Middle-class parents see themselves as partners with teachers, while working-class parents see teachers owning sole responsibility for educating their child. Thus, students from middle-class backgrounds see higher educational advantages in the form of higher academic performance. This, presumably, yields differing mental representations of themselves as students. Which leads to differing educational aspirations.

Given that parents from middle-class backgrounds actively participate in their child's education, it follows that these students perform better academically compared to their working-class background counterparts. However, students (and parents) from middle-class backgrounds tend to attribute their academic performance to their own ability rather than being generated by parental participation. These same students also attribute the performance of working-class peers to their "lack of effort" as opposed to circumstantial reasons. This phenomenon is known by social psychologists as a self-serving attribution bias (Fiske and Taylor 2021). Which implies that students from middle-class backgrounds see themselves in relation to school in a more positive light than students from a working-class background regarding schooling.

Their bias stems from self-regulation. Self-regulation describes how individuals control and direct their actions, emotions, and thoughts, as well as how they formulate and pursue goals. In this scenario, one's self-efficacy influences self-regulation such that middle-class students

performing better academically would undoubtedly hold greater belief in their own academic ability given their aforementioned self-serving attribution biases. Taken together, this implies that students from a middle-class background, as well as healthier students would be more likely to be encouraged to go to college and subsequently complete more schooling than students from a working-class background. Hence, this study considers educational aspirations in the casual pathway of interest.

The causal pathway containing educational aspirations is embedded within the *community cultural wealth* framework introduced by Yosso (2005). That is, educational aspirations act as aspirational capital, maintaining hope and resilience against structural barriers to education in the US. This form of cultural capital is transformed into educational attainment and then economic capital. Yosso's (2005) *community cultural wealth* framework, which critiques Pierre Bourdieu's traditional concept of cultural capital that centers white, middle-class norms as the standard for social mobility, highlights the unique forms of capital developed by marginalized communities to navigate and resist oppression. Yosso (2005) introduced 6 forms of cultural wealth, however this study focuses on aspirational capital in the form of education aspirations. Note that aspirational capital in the form of educational aspirations is passed down between generations of black folk, particularly in the most educated households (households of black excellence), and helps navigate and resist oppression by maintaining hope and resilience despite perpetual systemic barriers to social mobility for black people in the US.

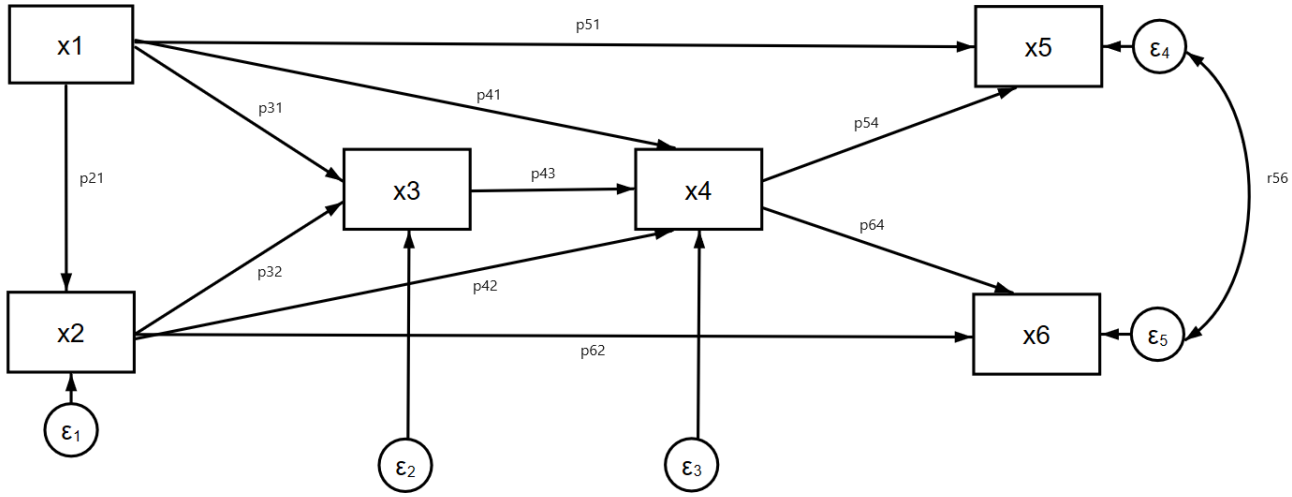


Figure 3.1 General path diagram for the causal pathways of interest.

This study also extends on the Wisconsin Model by considering childhood health in the causal pathway between childhood SES and adult SES. Figure 3.1 is the path diagram (generated through STATA's SEM builder) that illustrates all casual pathways of interest. To my knowledge, this is the first study to explicitly look at these pathways together and analytically examine the path dependencies between early life conditions (childhood SES and health) and later life outcomes (adult SES and overall health) using differential calculus. The next section introduces a method for analytically examining any causal pathway of interest using calculus to calculate the structural equations.

Analytical Casual Modeling

Path analysis can be considered a simpler version of structural equation modeling (SEM) that can and has historically used latent variables (Bollen et al. 2022). Both of which are under the commonly known moniker causal modeling. Casual modeling is a confirmatory method aimed at specifying the process of interrelated effects and testing whether a set of observed data is consistent with hypothesized processes (Bentler 1988). Casual modeling allows researchers to discern and

assess the effects of a set of factors on outcomes through multiple causal pathways. The power of casual modeling comes in its ability to analytically distinguish between the direct, indirect, spurious, and unanalyzed effects of any pathway in a path diagram. It is important to note that this means that the terms "direct," "indirect," "spurious," and "unanalyzed" are used to distinguish between the terms in the equation associated with the total effect.

Like all SEMs, classic path analysis relies on a system of equations (common phrasing in mathematics), or a "series of structured regression equations" (common phrasing in the social sciences). These equations makeup up the "structural equations" in structural equation modeling. Path analysis was developed by Wright (1922) and introduced to sociology by Blalock (1964) and Duncan (1966). Duncan's (1966) article pioneered path analysis in sociology, providing sociologists with a useful analytic tool that helped advance the discipline in an immeasurable way. Later, Duncan and Colleagues (Duncan, Haller, and Portes 1971) offered an addendum in Blalock [1971] due to mistakes related to confusing the population parameter with sample estimations and the confounding of noncausal effects with indirect causal effects.

Nevertheless, Duncan (1966) illustrated a method for deriving the expression for the total effect associated with any pathway in a given path diagram, using a system of linear equations as a starting point. Then, deriving the total effect by utilizing statistical objects (specifically the correlation coefficient formula). Thus, Duncan's (1966) formulation requires the successive expansions of the correlation coefficient in each term in the basic theorem of path analysis (see equation 5 in Duncan 1966). This means that deriving the total effect requires explicitly multiple structural equations and multiple multivariate algebraic expansions over all variables in the pathway. For path diagrams including more than 5 variables, it is clear that this formulation may be analytically cumbersome.

However, this problem is attenuated with statistical software such as STATA. The algorithm that underpins this calculation relies on the formulation proposed by Sewall Wright, Judea Pearl, Kenneth Bollen, Karl Jöreskog, and Bengt Muthén (Bollen et al. 2022), which extended on the formulation proposed by Duncan (1966). Both methods have similar starting points, that is, a system of linear equations. And neither approach includes the spurious and unanalyzed effects in the equation associated with the total effect. It is important to note that "spurious effect" refers the sum of terms associated with the non-casual pathways and "unanalyzed effects" refers to the sum terms associated with the pathways that include correlation coefficients, that is pathways that include double arrows.

This paper, instead, proposes the use of an analytical approach following the calculations of path coefficients calculated by statistical software. The mathematics involved will be described in detail later in this section. The procedure goes as follows: Starting from a single generalized mathematical object (specifically the functional form of casual pathways in any path diagrams), the derivation of the total effect requires the chain rule for multivariate functions and differentials to extrapolate the equation associated with the total effect for any given pathway of interest in a path diagram. This formulation, thus, analytically accounts for the spurious and unanalyzed effects in addition to the direct and indirect effects.

For example, in Figure 3.1, x_1 = parent's education, x_2 = childhood health, x_3 = educational aspirations, x_4 = educational attainment, x_5 = household income, x_6 = current health status. In the path diagrams in this study, straight arrows may only point in one direction. This assumes that variables cannot be both causes and effect of each other. While curved, double-headed arrows indicate a correlation between exogenous variables or error terms. Given that each causal pathway can be denoted by a regression equation, the pathway between any two variables can be expressed

as a function of the explanatory variables. For instance, the casual pathway of interest from x_1 to x_3 can be expressed by:

$$x_3 = f(x_1, x_2) \quad (1)$$

Moreover, given the relationship between x_1 and x_2, x_2 can be expressed as:

$$x_2 = g(x_1). \quad (2)$$

Thus, x_3 can be re-written as:

$$x_3 = f\big(x_1, g(x_1)\big). \quad (3)$$

This equation is important because it allows for the calculation of a small variation in x_3 (denoted by dx_3) when there is small variation in x_1 (denoted by dx_1). These small variations are commonly referred to as *differentials* in mathematics. Given equation 3, dx_3 can be expressed using the differential formula for two dependent variables:

$$dx_3 = \frac{\partial x_3}{\partial x_1} dx_1 + \frac{\partial x_3}{\partial x_2} dx_2 \quad (4)$$

To calculate the small variation in x_3 when there is a small variation in x_1, one simply has to divide both sides of the equation by dx_1:

$$\frac{dx_3}{dx_1} = \frac{\partial x_3}{\partial x_1} + \frac{\partial x_3}{\partial x_2}\frac{dx_2}{dx_1} \quad (5)$$

Equation 5 tells a number of stories. Generally, however, this equation states that total effect of x_1 on x_3 depends on the variation in x_3 when x_2 is constant plus the product of the variation in x_3 when x_1 is constant and the correlation between x_2 and x_1. Thus, the structural equation that describes the relationship between x_1 and x_3 can be expressed as:

$$r_{31} = p_{31} + p_{32}r_{12} \quad (6)$$

Where $r_{31} = \frac{dx_3}{dx_1},\ p_{31} = \frac{\partial x_3}{\partial x_1},\ p_{32} = \frac{\partial x_3}{\partial x_2},\ r_{12}\frac{dx_2}{dx_1}$. In this example r_{ij} represents the correlation coefficient from the standard correlation matrix and p_{kl} represent path coefficients (standardized beta coefficients). This derivation is of the simplest possible path between two variables with a connecting 3 variable and serves as an example of how to construct a structural equation from a path diagram. In the following example I will use the path diagram in Figure 3.1 to calculate the total effect of the pathway between childhood health and adult health.

Using the same logic for the construction of equation 1, the causal pathway of interest between x_2(childhood health) and x_6 (adult health) can be expressed as:

$$x_5 = f(x_{1,}x_4, x_6) \tag{7}$$

Expressing equation 7 in terms of x_1 gives:

$$x_5 = f\{x_1,\ g[x_1, h(x_1, j(x_1)),\ j(x_1)],\ m[g(\ x_1, h(x_1, j(x_1)),\ j(x_1)), j(x_1)]\} \tag{8}$$

Where $x_4 = g(x_1, x_3, x_2)$, $x_6 = m(x_4, x_2)$, $x_3 = h(x_1, x_2)$, and $x_2 = j(x_1)$. Using the differential formula on equation 8 and dividing dx_6 by dx_2 the total effect of x_2 on x_6 can be written as:

$$\frac{dx_5}{dx_1} = \frac{\partial x_5}{\partial x_1} + \frac{\partial x_5}{\partial x_4}\frac{dx_4}{dx_1} + \frac{\partial x_5}{\partial x_6}\frac{dx_6}{dx_2} \tag{9}$$

Using the chain rule for multivariate functions on the second derivatives of the second and third terms on the right side of the equation, and simplifying the resulting answer, equation 9 becomes:

$$\frac{dx_5}{dx_1} = \frac{\partial x_5}{\partial x_1} + \frac{\partial x_5}{\partial x_4}\frac{\partial x_4}{\partial x_1} + \frac{\partial x_5}{\partial x_4}\frac{\partial x_4}{\partial x_3}\frac{\partial x_3}{\partial x_1} + \frac{\partial x_5}{\partial x_4}\frac{\partial x_4}{\partial x_3}\frac{\partial x_3}{\partial x_2}\frac{\partial x_2}{\partial x_1} + \frac{\partial x_5}{\partial x_4}\frac{\partial x_4}{\partial x_2}\frac{\partial x_2}{\partial x_1} + \frac{\partial x_5}{\partial x_6}\frac{\partial x_6}{\partial x_4}\frac{\partial x_4}{\partial x_1} + \frac{\partial x_5}{\partial x_6}\frac{\partial x_6}{\partial x_4}\frac{\partial x_4}{\partial x_3}\frac{\partial x_3}{\partial x_1} + \frac{\partial x_5}{\partial x_6}\frac{\partial x_6}{\partial x_4}\frac{\partial x_4}{\partial x_3}\frac{\partial x_3}{\partial x_2}\frac{\partial x_2}{\partial x_1} + \frac{\partial x_5}{\partial x_6}\frac{\partial x_6}{\partial x_4}\frac{\partial x_4}{\partial x_2}\frac{\partial x_2}{\partial x_1} + \frac{\partial x_5}{\partial x_6}\frac{\partial x_6}{\partial x_2}\frac{\partial x_2}{\partial x_1} \tag{10}$$

Utilizing the notation convention from equation 6, equation 10 can be written as:

$$r_{51} = p_{51} + p_{54}p_{41} + p_{54}p_{43}p_{31} + p_{54}p_{43}p_{32}p_{21} + p_{54}p_{42}p_{21} + r_{56}p_{64}p_{41} + r_{56}p_{64}p_{43}p_{31} + r_{56}p_{64}p_{43}p_{32}p_{21} + r_{56}p_{64}p_{42}p_{21} + r_{56}p_{62}p_{21}$$

(11)

Equation 11 represents the structural equation estimating the total variance for the effects of variable 1 on variable 5. This equation indicates the sum of the flow along every possible pathway that connects variables 1 and 5. The derivation can be done for any pathway in the path diagram in Figure 3.1 Thus, when path and correlation coefficients have been calculated, the utility of an analytical approach to casual modeling is explicit. The total variance (or total effect) between any two factors can be partitioned into specific types: direct (p_{51}), indirect ($p_{54}p_{41}$, $p_{54}p_{43}p_{31}$, $p_{54}p_{43}p_{32}p_{21}, p_{54}p_{42}p_{21}$), spurious and unanalyzed ($\mathrm{r}_{56}p_{64}p_{41} + \mathrm{r}_{56}p_{64}p_{43}p_{31} + \mathrm{r}_{56}p_{64}p_{43}p_{32}p_{21} + \mathrm{r}_{56}p_{64}p_{42}p_{21} + r_{56}p_{62}p_{21}$) (unknown directionality, that is, the curved arrows) effects. Equation 11 will be used for the analysis of the total effect later in this paper. Using the chain rule to compute total effects is a reasonable and precise approach, particularly for specific paths and small models, as it the case in this study. I do not use the traditional matrix algebra calculations because this study prioritizes analytical clarity for 4 specific paths. For this study, the chain rule is a valid and insightful approach.

I utilized the SEM builder in STATA to calculate the path and correlation coefficients and the derivation above to calculate the contribution of the direct, indirect, spurious, and unanalyzed effects to the total effect. To my knowledge, this is the first study that uses an analytical approach (casual modeling) in the social sciences to analyze the effects of growing up in a household of black excellence on future income. Moreover, this paper contributes to the literature by providing a brief guide for conducting analytical causal modeling in the social sciences. Although researchers have provided a broad overview of how to conduct SEM (see Kline 2016), this paper extends the

SEM methodology by explicitly calculating the total effect of any pathway of interest after path coefficients are estimated.

Data

The Panel Study of Income Dynamics is a nationally representative sample of 18,000 individuals living in 5,000 families in the United States. It is the longest-running longitudinal household survey in the world having started in 1968(?). The data covers information on income, health, child development, education, and many other topics. The complete dataset contains hundreds of variables with information on demographics and geographic variables along with information about early life conditions and life course events. The data were collected from 1968 to 2021. This study used data from the years 2011, 2013, 2020, and 2021.

Methods

Variables of Interest

The key variables include respondent's *current household income, health status, childhood health status, educational aspirations, father's education, mother's education.* and *respondent's education.*

Current household income ranged from 0 to 1,165,000 US dollars, with a mean of 87,231 US dollars and a median of 60,900 US dollars (n = 1,239). Data were collected in the year 2021. The distribution of household income was highly skewed to the right. I normalized these data by taking the logarithm of income, creating the variable *logincome*. I used this variable in my analysis.

Respondent's health status is defined as respondent's self-reported overall health at the time they took the survey in 2021. Respondents were asked: "Now I have a few questions about

(your/Reference Person's) health. Would you say (your/Reference Person's) health in general is excellent, very good, good, fair, or poor?" Respondents' answers fell on a five-point Likert scale ranging "excellent" (coded as "1") to "poor" (coded as "5") (n=1,264). Roughly 17 percent of respondents indicated that they were in "excellent" health. Thirty-six percent indicated they were in "very good" health. Thirty-five percent indicated they were in "good" health. Roughly 10 percent indicated they were in "fair" health. And Roughly 2 percent indicated they were in "poor" health.

Childhood health status is defined as the respondent's self-reported overall health as a child. In 2013, respondents were asked: "These next questions ask about your health during childhood, that is, before age 17. Would you say that your health during that time was excellent, very good, good, fair or poor?" Respondents' answers fell on a five-point Likert scale ranging "excellent" (coded as "5") to "poor" (coded as "1") (n=495). Roughly 46 percent of respondents indicated they were in "excellent" health as a child. Roughly 36 percent indicated they were in "very good" health as a child. Roughly 14 percent indicated they were in "good" health as a child. Roughly 3 percent indicated they were in "fair" health as a child. And roughly 1 percent indicated they were in "poor" health as a child.

Educational aspiration is defined as how far respondents would like to go in school. In 2011, respondents were given the statement "How far would you like to go in school? Would you like to graduate from high school, graduate from a two year community college, earn a specialized certificate from a vocational or trade school, attend a four-year college, graduate from a four-year college, get more than four years of college, or do something else?" (n=1,571). Respondent's responses fell on a 6-point scale ranging from "Graduate from high school" (coded as "1") to "Get more than 4 years of college" (coded as "6"). Roughly 2 percent of respondents indicated they

would like to “Graduate from high school.” Roughly 16 percent indicated they would like to “Graduate from a two year community college.” Roughly 10 percent indicated they would like to “Earn a specialized certificate from a vocational or trade school.” Roughly 3 percent indicated they would like to “Attend a 4-year college.” Roughly 41 percent indicated they would like to “Graduate from a 4-year college.” Roughly 32 percent indicated they would like to “Get more than 4 years of college.”

Father’s education and *mother’s education* are defined as how much education respondent’s father and mother completed in 2020. Respondents were given the statement “How much education did (your/his/her) father complete (in the United States)?” (n=1,264). Respondent’s answers fell on 8-point scale ranging from “Completed 0-5 grades” (coded as “1”) to “Completed 17 or more years; College, advanced or professional degree, some graduate work; close to receiving degree” (coded at “8”). For father’s education, roughly 1 percent of respondents “Completed 0-5 grades.” Roughly 4 percent “Completed 6-8 grades; "grade school"; DK but mentions could read and write.” Roughly 9 percent “Completed 9-11 grades (some high school); junior high.” Roughly 39 percent “Completed 12 grades (completed high school); "high school".” Roughly 4 percent “Completed 12 grades plus nonacademic training; R.N. (no further elaboration).” Roughly 11 percent “Completed 13-14 years; Some college, no degree; Associate's degree.” Roughly 15 percent “Completed 15-16 years; College BA and no advanced degree mentioned; normal school; R.N. with 3 years college; "college".” Roughly 9 percent “Completed 17 or more years; College, advanced or professional degree, some graduate work; close to receiving degree.” And roughly 8 percent were either not educated in the US or had no education.

For mother's education, roughly 1 percent of respondents "Completed 0-5 grades." Roughly 4 percent "Completed 6-8 grades; "grade school"; DK but mentions could read and write." Roughly 6 percent "Completed 9-11 grades (some high school); junior high." Roughly 33 percent "Completed 12 grades (completed high school); "high school"." Roughly 3 percent "Completed 12 grades plus nonacademic training; R.N. (no further elaboration)." Roughly 18 percent "Completed 13-14 years; Some college, no degree; Associate's degree." Roughly 19 percent "Completed 15-16 years; College BA and no advanced degree mentioned; normal school; R.N. with 3 years college; "college"." Roughly 9 percent "Completed 17 or more years; College, advanced or professional degree, some graduate work; close to receiving degree." And roughly 8 percent were either not educated in the US or had no education.

I created a new variable called *parents' education* that is equal to the sum of *father's education* and *mother's education.* I used this variable as a proxy for respondent's childhood socioeconomic status for my analysis. Additionally, I analyzed respondents by whether respondent's parents were white, black, black with a college education at the bachelor's level, and black with a college education at the graduate level.

Respondent's education is defined as the number of years of schooling the respondents have completed in 2020. *Respondent's education* ranged from "Completed no grades of school" (coded as "0") to "Actual number" of years of education ranging from 1-17 (n=1,264).

Analysis Methods

To examine the causal pathways for the effect of parent's education on future household income, I will first compared the path coefficients for all households, black households, white households, and highly educated black households. This comparison will illuminate how parent's

race and education may moderate one's future income, education, childhood health, and adult health. In particular, I will highlight how having highly educated black parents may moderate the effect size on the aforementioned outcomes when compared to the average black household.

I will subsequently conduct a casual structural analysis of the path diagrams using calculus, specifically the chain rule for multivariate functions. This method is appropriate given the linear and multivariate nature of the relationships illustrated by the path diagrams. This method allows for the assessment of the effects of a set of factors on outcomes through multiple causal pathways. The power of this method comes in its ability to analytically distinguish between the direct from indirect effects, as well as the spurious and unanalyzed effects using formalized mathematics. This allows for more analytical clarity when compared to more traditional methods. I utilized the SEM builder in Stata to calculate the path and correlation coefficients and the derivation above to calculate the different effects between key variables.

Findings

Path and Correlation Coefficients

The next seven figures are directed graphs path diagrams. Each path diagram shows estimation for the path and correlation coefficients between the variables of interest. Straight arrows indicate the direction of the causal relationship with path coefficients floating on the line. Curved arrows indicate correlation between variables where causal relationship is unclear with correlation coefficients floating on the curved line. There are arrows outside the model that indicate variance contributed by error and any unmeasured variable.

The correlation coefficient comes from a standard correlation matrix containing all of the variables in the model. The path coefficients are standardized and unstandardized beta coefficients

from a linear regression model in which the variables at the end of the pathway of interest are dependent and the other variables in the pathway are independent.

Figures 3.2 through 3.6 illustrate the causal pathways between respondent's parent's education and current household income. Figure 3.2 estimates the path coefficients, correlation coefficients, and error coefficients between all variables. All the coefficients in Figures 3.2 through 3.6 are statistically significant ($p<0.001$). Figure 3.2 will serve as the reference point in my analysis.

Figure 3.3 specifies the path diagram in Figure 3.2 to respondents who grew up in households with white parents or guardians. Figure 3.4 specifies the path diagram in Figure 3.2 to respondents who grew up in households with black parents or guardians. Figure 3.5 specifies the path diagram in Figure 3.2 to respondents who grew up in households with at least one black parent holding a bachelor's degree. Figure 3.6 specifies the path diagram in Figure 3.2 to respondents who grew up in households with at least on black parent holding a graduate level degree. The latter two categories represent households of black excellence.

Figure 3.2.1 Standardized Path Diagram for All Households

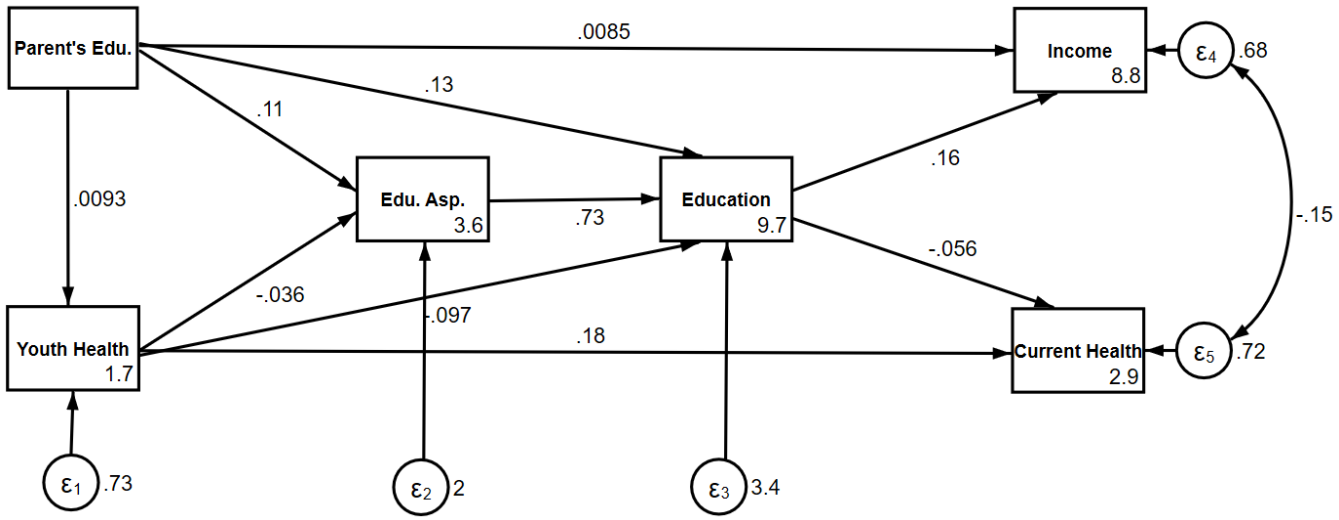


Figure 3.2.2 Unstandardized Path Diagram for All Households

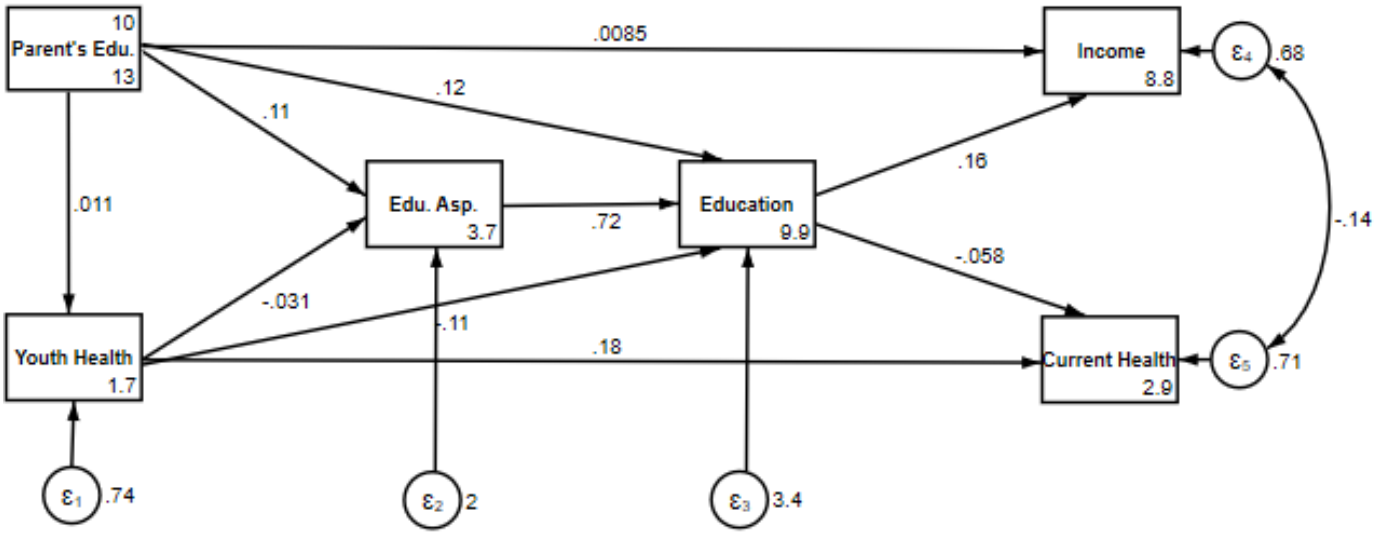


Figure 3.3.1 Standardized Path Diagram for White Households

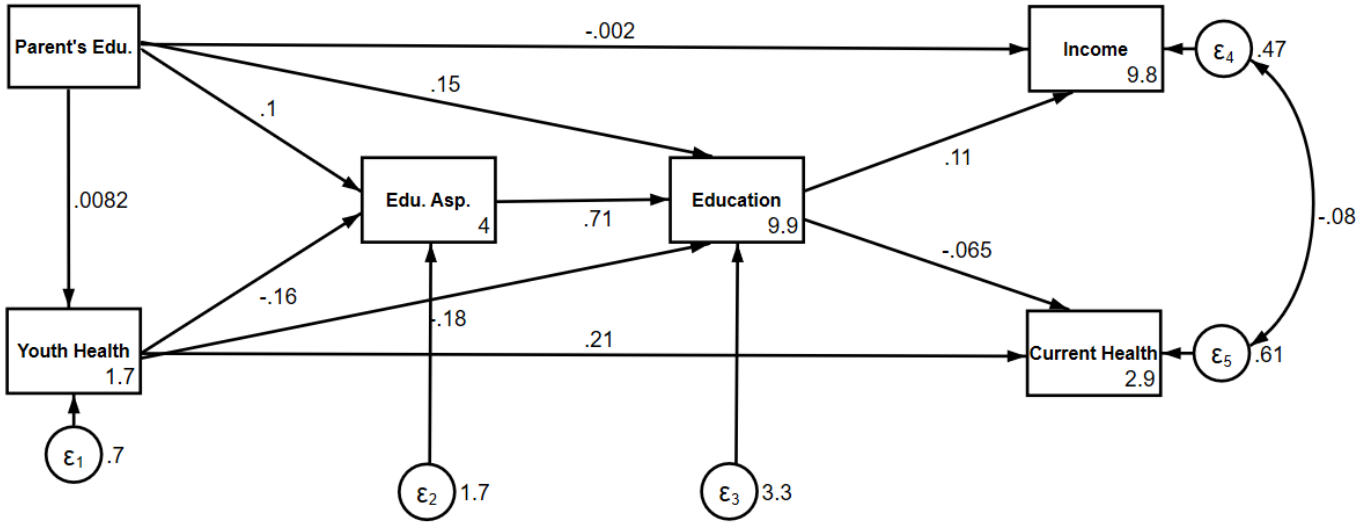


Figure 3.3.2 Unstandardized Path Diagram for White Households

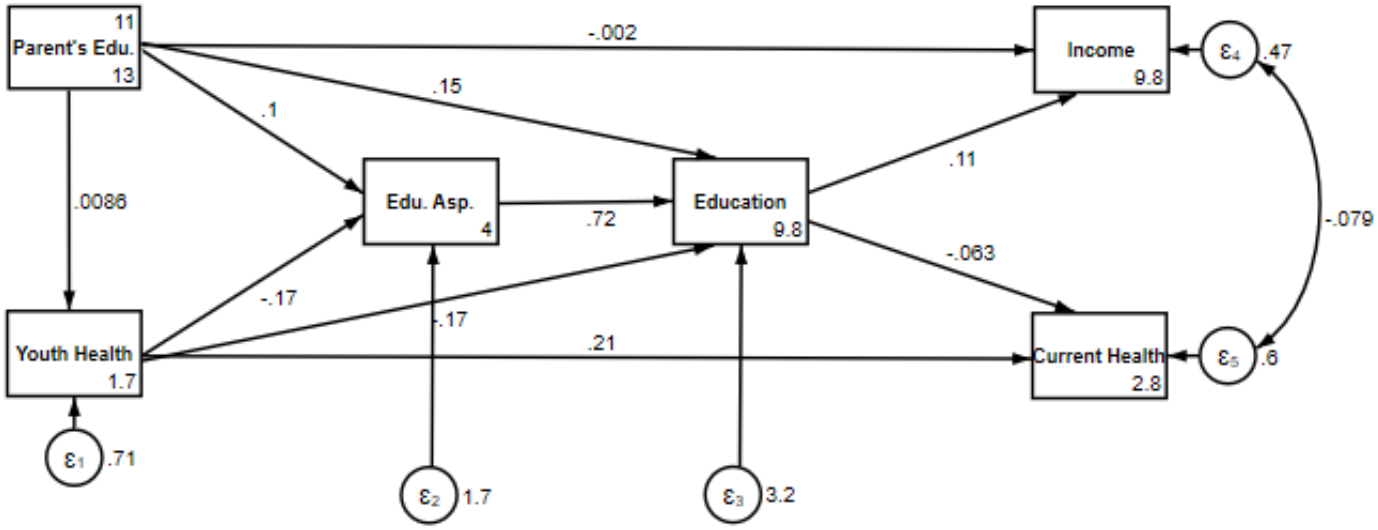


Figure 3.4.1 Standardized Path Diagram for Black Households

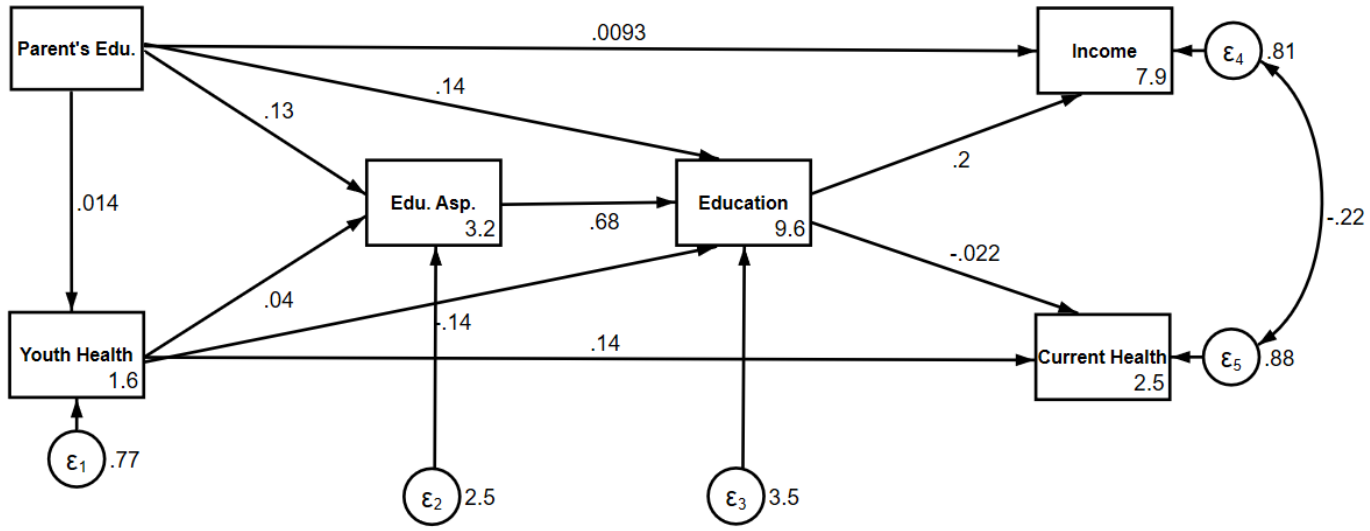


Figure 3.4.2 Unstandardized Path Diagram for Black Households

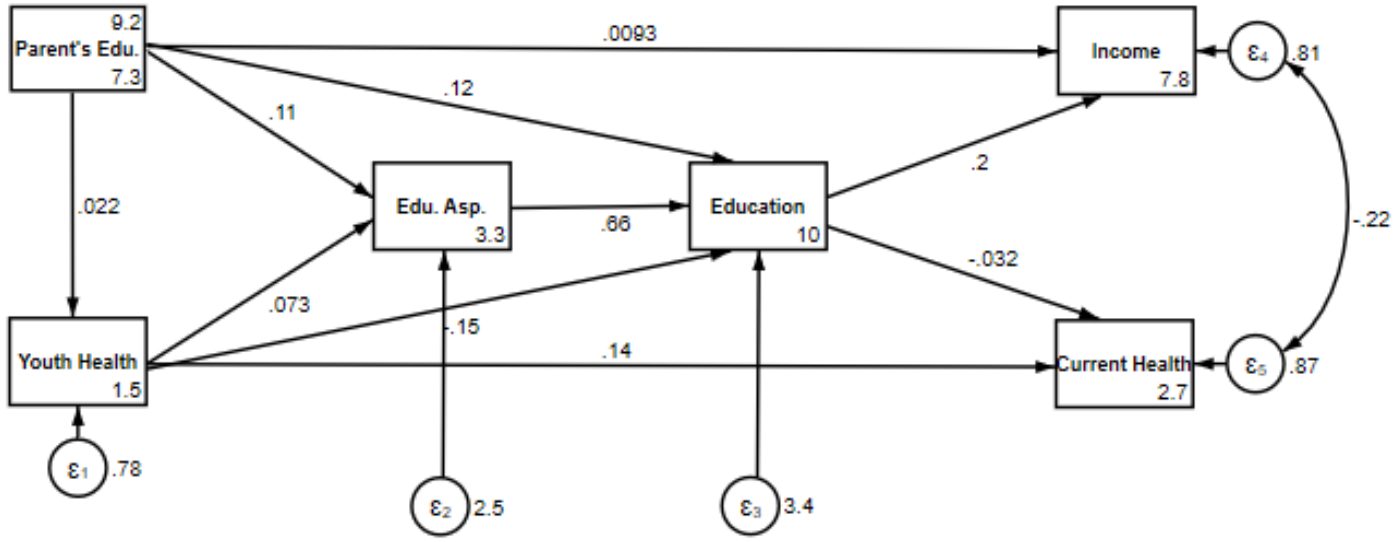


Figure 3.5.1 Standardized Path Diagram for Black Households with College Educated Parents at Least Bachelor's Level

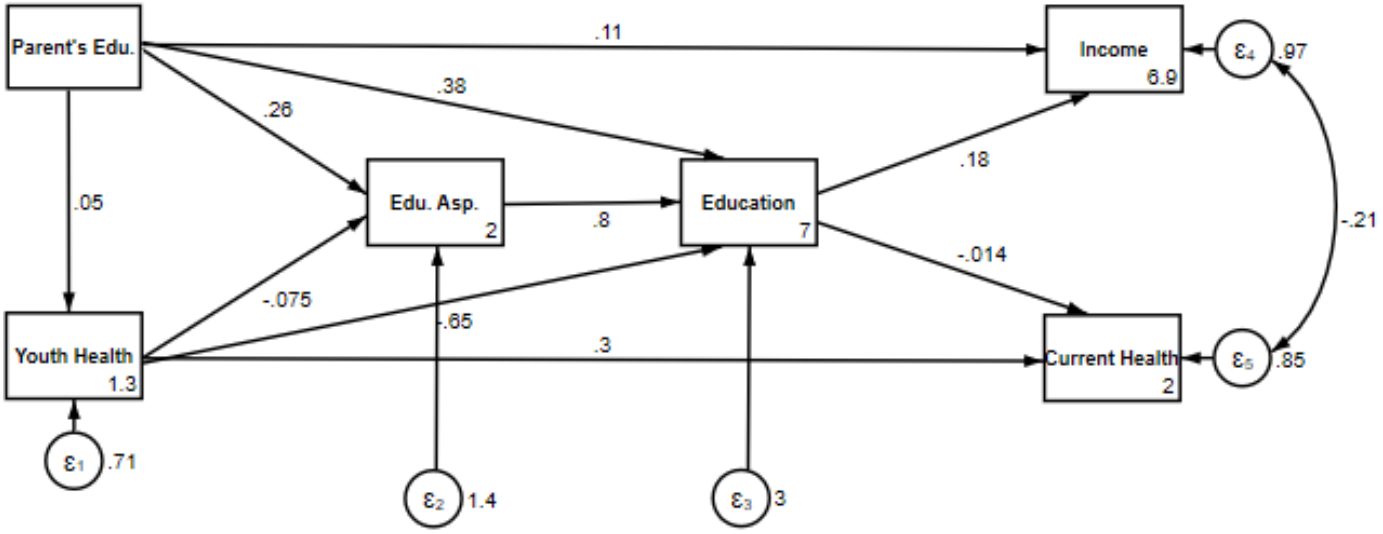


Figure 3.5.2 Unstandardized Path Diagram for Black Households with College Educated Parents at Least Bachelor's Level

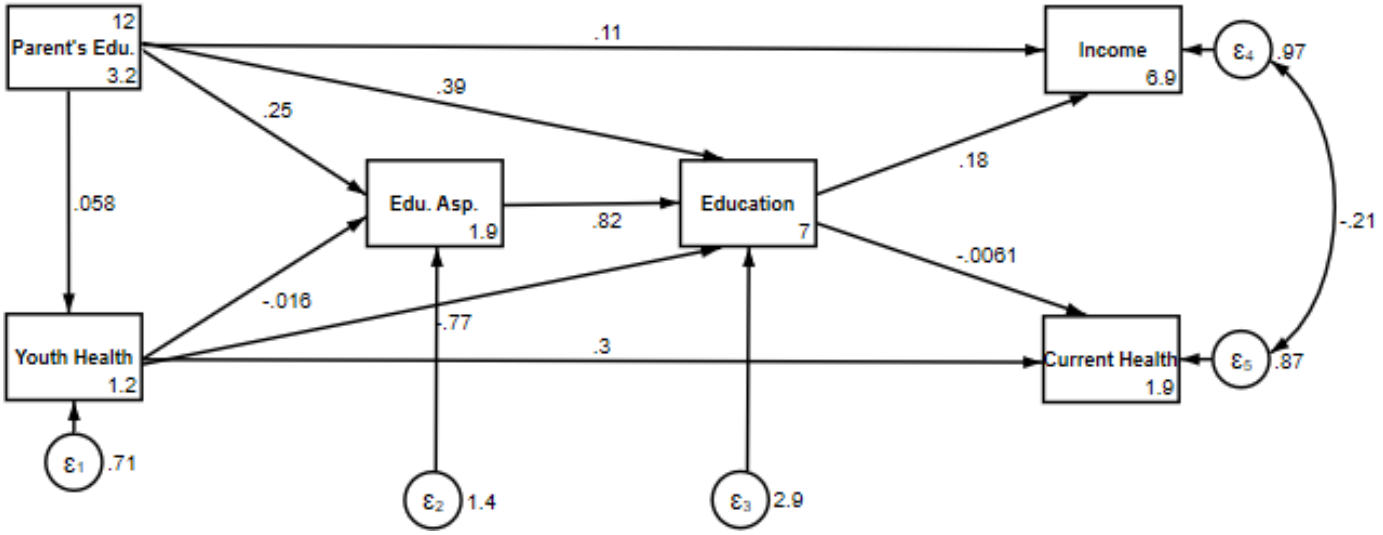
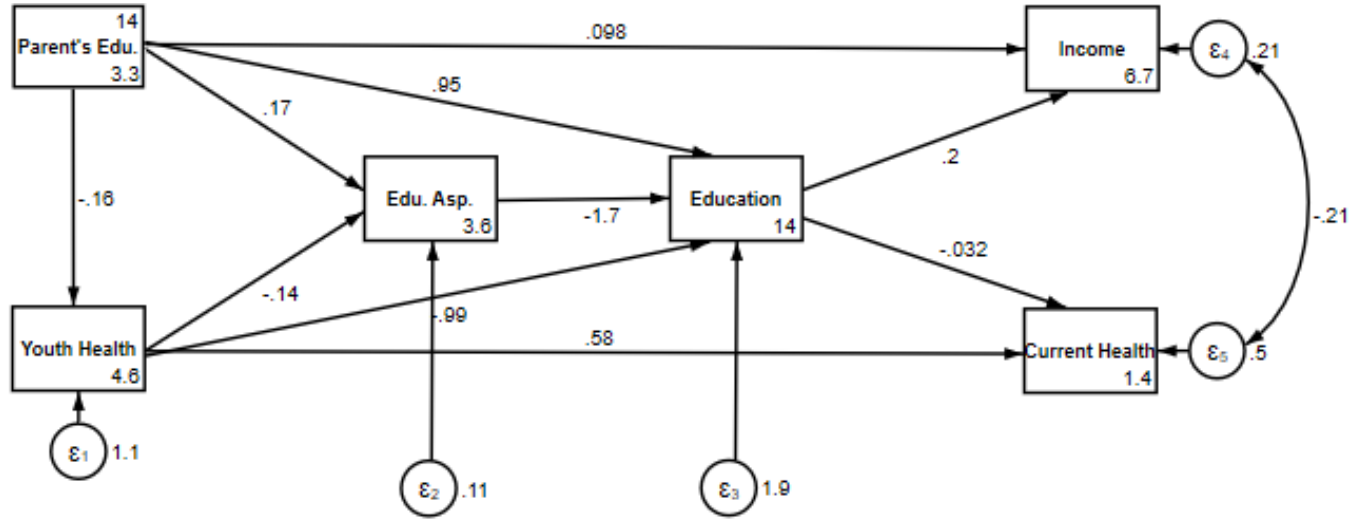


Figure 3.6 Unstandardized Path Diagram for Black Households with College Educated Parents at Least Graduate Level



When comparing Figure 3.3 to Figure 3.2, that is white households in the data to the sample dataset, the coefficients are nearly identical with the exception of the coefficients for the effect of

childhood health (youth health) on educational aspirations, the effect of childhood health on educational attainment, and the correlation between current health and current household income.

Focusing on the effect of childhood health on educational aspirations, it appears that growing up in a white household quadruples the negative effect that childhood health has on educational aspirations in the sample dataset. That is, the better the childhood health of the respondent the less educational aspirations they had. However, this effect is reversed in black households and more than doubled in households of black excellence. This implies that respondents with better health during childhood growing up in black households and households of black excellence sought out more education than those with poorer health. However, the effect of childhood health on actual educational attainment for households of black excellence is negative and double the average but between 11 and 41 percent less than white households. Thus, for all respondents, the better the health, the less education they attained.

Assessing the correlation between current health and current household income for all models shows that respondents who grew up in a household of black excellence, specifically, had the highest correlation, in terms of magnitude of the correlation coefficient, between their current health and current income. It is important to note that this is not a multiple-group analysis here and is strictly descriptive. Respondents who grew up with black parents or guardians with a graduate level education being double the correlation of the average household and over triple the white household. It is important to note that current health is measured on a 5-point scale where excellent health is 1 and poor health is 5. Thus, the negative sign implies that the better the health of the respondent the higher their income and the better the income of the respondent the better their current health. The magnitude of this correlation being the greatest for respondents who grew up in households of black excellence, specifically with parents with graduate level degrees.

The main effect this paper is focused on is the effect of respondent's parent's education on current income. What is interesting is that for all models except Figures 3.5 and 3.6, the path coefficient for the effect of respondent's parent's education on current household income is nearly zero. This supports the "stickiness" theory of SES. That is, many people are not upwardly or downwardly mobile but instead spend most of their lives in the class they were born in. However, respondents who grew up in households of black excellence, specifically with parents or guardians holding bachelor's and graduate level degrees, see a boost in household income compared to all other respondents. Additionally, the effects of educational aspirations on educational attainment is strongest for respondents with a parent that at least has a bachelor's degree and weakest for respondents with black parents with at least a graduate level degree.

Total Effect

In this section I examine the total effect of respondent's parents' education on their current household income by using equation 11 in conjunction with the estimated coefficients in Figures 3.2 through 3.6 to calculate the structural equation associated with total effects mentioned above. Resulting in the calculation of the contributions of the direct, indirect, spurious, and unanalyzed effects to the total effect. I constructed the path diagram illustrated in Figure 3.1 (with all path coefficients being standardized and unstandardized) using the SEM builder in STATA as a blueprint for Figures 3.2 through 3.6. Because all the variables are observed, this type of SEM is commonly referred to as a path analysis.

However, unlike path analysis, I employed an analytical approach to the calculations of the total effect, rooted in the chain rule in calculus. This approach is appropriate given that path analysis involves the solution of multiple linear regression equations, with dependent variables

being normally distributed and the relationship between variables being linear, causal, and additive. Thus, calculus, the mathematical study of change is sufficient to determine the structural equation associated with the casual pathway of interest. That is, the total effect of childhood SES on current household income.

Additionally, by logical temporal ordering, respondent's parent's education and respondent's childhood health preceded or was simultaneously occurring with respect to educational aspirations, and subsequently their educational attainment and household income at the time of the survey, as well as their current health status at the time of the survey. Moreover, respondent's educational attainment is highly correlated with their household income ($p<0.001$).

Thus, in order to determine the total effect of the path dependencies between childhood SES and current household income, I used equation 11 to determine the following structural equation:

$$r_{51} = p_{51} + p_{54}p_{41} + p_{54}p_{43}p_{31} + p_{54}p_{43}p_{32}p_{21} + p_{54}p_{42}p_{21} + r_{56}p_{64}p_{41} + r_{56}p_{64}p_{43}p_{31} + r_{56}p_{64}p_{43}p_{32}p_{21} + r_{56}p_{64}p_{42}p_{21} + r_{56}p_{64}p_{42}p_{21} + r_{56}p_{62}p_{21}$$

Where the direct (p_{51}), indirect ($[1]p_{54}p_{41}$, $[2]p_{54}p_{43}p_{31}$, $[3]p_{54}p_{43}p_{32}p_{21}$, $[4]\, p_{54}p_{42}p_{21}$), spurious and unanalyzed ($\mathrm{r}_{56}p_{64}p_{41} + \mathrm{r}_{56}p_{64}p_{43}p_{31} + \mathrm{r}_{56}p_{64}p_{43}p_{32}p_{21} + \mathrm{r}_{56}p_{64}p_{42}p_{21} + + r_{56}p_{62}p_{21}$) (unknown directionality, that is, the curved arrows) effects are calculated using the path and correlation coefficients in Figures 3.2 through 3.6.

Table 3.1 delineates the contribution of the direct, indirect, spurious, and unanalyzed effects to the total effect of each causal pathway based on unstandardized coefficients. Examining the causal pathway r_{51} (the effect of respondent's parent's education on current household

income), table 3.1 indicates that the racial and educational level of parents or guardians matters in determining which causal pathway contributes the most or least to the total variance.

Table 3.1 Contribution of the Total Effect of Respondent's Parent's Education on Current Household Income

	Effect Type (% of Total Effect)					
	Direct	Indirect 1	Indirect 2	Indirect 3	Indirect 4	Spurious/ Unanalyzed
All Households	20.0	45.2	29.8	0.1	0.5	4.5
White Households	7.2	59.0	28.3	0.4	0.6	4.5
Black Households	18.3	47.3	28.6	0.4	1.3	4.1
Black Households of Excellence (College Degrees)	47.9	30.6	16.1	0.1	3.5	1.9
Black Households of Excellence (Graduate Degrees)	23.7	45.9	14.0	1.8	7.6	7.0

I use the category All Households to mean the average results, and this serves as a point of reference in the subsequent analysis. Focusing on the direct effect, it is clear that respondent's race and the education levels of respondent's parents or guardians matter. Given the "stickiness" of SES, on average the direct effect contributes about 20 percent to the total effect. When comparing white and black households, the direct effect is more than double that of white households than black households. This implies that SES is stickier for white households than black households. That is, respondents who grew up with white parents or guardians saw less of a change in SES throughout their lifetime with increasing parental education when compared to respondents who grew up with black parents and the average respondent in the dataset.

The contribution of the direct effect for respondents who grew up in a household of black excellence, however, differed by the educational level of their parents. The contributions of the direct effect of childhood SES on current household income for respondents with black parents holding bachelor's degrees was nearly 50 percent of the total effect and about 24 percent for respondents with black parents holding graduate level degrees, which is close to the average. This

implies that respondents in black households of excellence where the parents hold bachelor's degrees saw massive gains in SES a direct result of their parents holding a college degree. However this result returned to baseline with graduate level education.

Table 3.1 indicates that the first indirect causal pathway ($p_{54}p_{41}$) contributes the most to the total effect of childhood SES on current household income and when specifying race and educational attainment for all respondents except those with black parents with at least one parent holding a bachelor's degree. This causal pathway follows the path from respondent's parent's education to respondent's educational attainment to respondent's current household income. For all households, the contribution of the first indirect effect ranges between 31 and 59 percent. The contribution of this effect type is largest for white households at 59 percent and smallest for black households of excellence where respondent's parents holding bachelor's degrees at 30 percent. Thus, according to this study, in the causal pathway of the total effect of childhood SES on current household income, the largest determining factor of upward or downward mobility of respondents throughout their lifetime depends on the relationship between their parent's education and their education. Moreover, the amount of this contribution also depends on race, to a small but nontrivial degree.

According to table 3.1, the second indirect effect ($p_{54}p_{43}p_{31}$) is the second largest contributor to the total effect. This causal pathway follows the path from respondent's parent's education to respondent's educational aspirations to respondent's educational attainment to respondent's current household income**.** Unlike the first indirect effect, the second indirect effect sees the largest contribution to the total effect for all households on average and the smallest for black households of excellence at 14 percent. Thus, adding educational aspirations to the first indirect effect seems to be less impactful to the total effect of childhood SES on current household

income for all households in general and childhood households of black excellence in particular. This may imply a lower level of importance of educational aspirations for black respondents than white respondents. It is important to note that the path coefficient between educational aspirations and educational attainment is the largest for respondents with parents with bachelor's degrees in the positive direction. Thus, educational aspirations are higher for these respondents but its role in the total effect seems to be smaller.

The third and fourth indirect effects ($[3]p_{54}p_{43}p_{32}p_{21}, [4]\, p_{54}p_{42}p_{21}$) have the least contribution for all households in the data, with the vast majority being no more than 8 percent of the total effect. The third indirect effect follows the path from respondent's parent's education to respondent's childhood health to respondent's educational aspirations to respondent's educational attainment and finally to respondent's current household income**.** Moreover, the fourth indirect path follows the path from respondent's parent's education to respondent's childhood health to respondent's educational attainment and finally to respondent's current household income. Thus, adding childhood health to the first and second indirect effects has little effect on the total effect of childhood SES on current household income. This is counterintuitive to the assumption that poor health as a child would limit the education potential of said child, thereby limiting their earning potential. However, clearly, the results indicate otherwise.

Discussion

This study directly analyzes the casual structure of the effect of childhood SES on one's current household income. That is, this study directly calculates the contribution of the direct, indirect, spurious, and unanalyzed effects to the total effect of respondent's parent's educational attainment on their current household income. Specifically, this study looks at respondents who

grew up in black households of excellence, where their parents either held a college degree or graduate degree. As a comparison, I also examine the effect of childhood SES on current household income for respondents who grew up in black and white households, as well as all respondents in general. Moreover, an analysis of the path diagrams for each group was done to look at how path and correlation coefficients differed by respondents' race and the educational status of respondent's parents or guardians.

There were several key findings on both fronts. An analysis of the path diagrams showed that the path coefficients for the effect of childhood health (youth health) on educational aspirations, the effect of childhood health on educational attainment, and the correlation between current health and current household income varied by race and the educational attainment of black parents.

The better the childhood health of the respondents the less educational aspirations they had. However, this effect is reversed in black households and more than doubled in households of black excellence. Thus, respondents with better health during childhood growing up in black households sought out more education than the average household in the dataset. However, the effect of childhood health on actual educational attainment for households of black excellence is negative and double the average, which is expected as poorer health as a child would necessarily limit one's potential for higher education.

However, for respondents who grew up with black parents or guardians with graduate degrees the correlation between their current health and current household income is double the correlation of the average household and over triple that of white household in terms of magnitude. This implies that the better the health of the respondents the higher their income and the better the income of the respondent the better their current health in respondents who grew up in black

households, with the magnitude of this correlation being the greatest for respondents who grew up in households of black excellence, specifically with black parents with graduate degrees.

The path coefficient for the effect of respondent's parent's education on current household income was similarly close to zero for all models except for respondent's with at least one black parent with at least a bachelor's degree. This result supports the "stickiness" theory of SES. That is, many people are not upwardly or downwardly mobile but instead spend most of their lives in the class they were born in. However, the path coefficient for respondents who grew up in households of black excellence, specifically with parents or guardians holding at least a bachelor's degrees saw the largest gain in household income compared to all other respondents including black respondents in general. These findings suggest that respondents with black parents holding at least a bachelor's degrees were most likely out of all respondents to see upward mobility.

After calculating the structural equation for the total effect of childhood SES on household income, I broke the result into 4 effect types that illuminate the largest and smallest contributors to the total effect by household type. Several key findings supported the conclusions made when comparing the path diagrams directly.

When comparing white and black households, the direct effect is more than double that of white households than black households. This suggests that SES is stickier for white households than black households and play a larger role a life time SES. Respondents who grew up with white parents or guardians saw less of a change in SES throughout their lifetime when compared to respondents who grew up with black parents and the average respondent in the dataset.

However, for respondents who grew up in a household of black excellence the contribution of the direct effect of childhood SES on current household income for respondents with black

parents holding college degrees was nearly 50 percent and about 24 percent for respondents with black parents holding graduate level degrees, which is close to the average but more than double respondents in white households. This result suggests that respondents in black households of excellence where the parents held college degrees gained the most in terms of SES compared to their parents throughout their lifetime, even being dramatically less than white households on average. I conclude this because not only were the percentages higher but the path coefficients were higher as well.

An analysis of the structural equation for the total effect of childhood SES on current household income also shows that the path through education is the most important by being the largest contributor to the total effect. However, according to the results, this path is more important for respondents in white households than black households, with this pathway contributing 59 percent to the total effect for white households and 47 percent for black households. The contribution is even less for respondents with black parents with graduate degrees at 30 percent.

Additionally, when educational aspirations are included into the pathway mentioned above, the contribution to the total effect becomes greatest for black households than white households but not by much. This may suggest that the educational aspirations of respondents who grew up in black households of excellence were less important to their financial future when compared to other households. It is beyond the scope of this study to reach any conclusion as to why this may be the case, but future researcher should focus on this line of inquiry.

A limitation of this study is that data may be biased in two ways. First, respondents may be limited in regard to their memory. However, the bias here is likely to be negligible given that questions focused on significant and generalized collections of moments in their childhood and adulthood. That is, their health as a child, mother's and father's education, educational aspirations,

educational attainment, household income, and current health status. Future work might implement sensitivity analysis to address these potential biases.

Path diagrams are used by scientists across all disciplines to visualize causal structures. However, many scientists do not go beyond illustrating the pathways and calculating correlation and path coefficients between factors. The true power of path diagrams stems from the calculation and interpretation of the structural equations associated with the causal pathways of interest. That is, the analysis of all the dependencies associated with a given causal pathway in order to precisely ascertain the constituents of the total effect associated with any given causal pathway, along with their relative strength. Thus, researchers can more accurately determine which effects contribute the most or the least to the total effect of any given pathway.

The utility of path diagrams as a powerful visualization tool, however, cannot be understated. Many models rely heavily on this tool, given it is the simplest way to clearly communicate the casual pathways between multiple factors. Thus, its ubiquity is not surprising. However, along with the inclusion of any path diagram should be causal structural analysis of the pathways most pertinent to the proposed model, as well as factors that have profound effects on key variables.

References

Addo, Fenaba R., Jason N. Houle, and Daniel Simon. 2016. "Young, Black, and (Still) in the Red: Parental Wealth, Race, and Student Loan Debt." *Race and Social Problems* 8(1):64–76.

Bentler, P. M. 1988. "Causal Modeling via Structural Equation Systems." *Handbook of Multivariate Experimental Psychology* 317–35.

Blalock, Hubert M., O. D. Duncan, A. O. Haller, and A. Portes. 1971. "Peer Influences on Aspirations: A Reinterpretation." Pp. 219–44 in *Causal models in the social sciences*. Chicago, IL: Aldine Pub. Co.

Blau, Peter M. and O.D. Duncan. 1978. *The American Occupational Structure*. The Free Press.

Bollen, Kenneth A. 1987. "Total, Direct, and Indirect Effects in Structural Equation Models." *Sociological Methodology* 17:37.

Chau, Kénora et al. 2016. "Associations between School Difficulties and Health-Related Problems and Risky Behaviours in Early Adolescence: A Cross-Sectional Study in Middle-School Adolescents in France." *Psychiatry Research* 244:1–9.

Chetty, Raj, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2018. "Race and Economic Opportunity in the United States: An Intergenerational Perspective." *The Quarterly Journal of Economics* 135(2):711–83.

Currie, Janet. 2009. "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development." *Journal of Economic Literature* 47(1):87–122.

Duncan, O. D., A. O. Haller, and A. Portes. 1970. “Peer Influences on Aspirations: A Reinterpretation.” Pp. 219–44 in *Causal models in the social sciences,* edited by H. M. Blalock. Chicago, IL: Aldine-Atherton, Inc.

Duncan, Otis Dudley. 1966. “Path Analysis: Sociological Examples.” *American Journal of Sociology* 72(1):1–16.

Fiske, Susan T. and Shelley E. Taylor. 2021. *Social Cognition: From Brains to Culture (4rd edition)*. Thousand Oaks, CA: Sage

Giroux, Henry. 2005. *Schooling and the Struggle for Public Life: Democracy’s Promise and Education’s Challenge*. 2nd ed. Boulder, CO: Paradigm Publishers.

Grossman, M. 2006. “Education and non-market outcomes.” In: Hanushek, E, Welch, F (eds) *Handbook of the Economics of Education*. Amsterdam: North-Holland, pp. 577–633.

Haller, Archibald O. and Alejandro Portes. 2015. “Status Attainment Processes,” Pp. 33-41 in Richard Arum, Irenee Beattie, and Karly Ford (Eds.) *The Structure of Schooling: Readings in the Sociology of Education.* Thousand Oaks, CA: Sage Publications, Inc.

Hjorth, Cathrine F. et al. 2016. “Mental Health and School Dropout across Educational Levels and Genders: A 4.8-Year Follow-up Study.” *BMC Public Health* 16(1).

Kim, Younoh and Vlad Radoias. 2019. “Subjective Socioeconomic Status, Health, and Early-Life Conditions.” *Journal of Health Psychology* 26(4):595–604.

Lareau, Annette. 1987. “Social Class Differences in Family-School Relationships: The Importance of Cultural Capital.” *Sociology of Education* 60(2):73.

Massey, Douglas S. 2016. "Segregation and the Perpetuation of Disadvantage." *The Journal of African American History* 98(1):1–12.

Mossakowski, Krysia N. 2014. “Social Causation and Social Selection.” *The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society* 2154–60.

Phelan, Jo C., Bruce G. Link, and Parisa Tehranifar. 2010. “Social Conditions as Fundamental Causes of Health Inequalities: Theory, Evidence, and Policy Implications.” *Journal of Health and Social Behavior* 51(1).

Wright, Sewall. 1922. “The Theory of Path Coefficients a Reply to Niles's Criticism.” *Genetics* 8(3):239–55.

Yosso, Tara J. 2020. “Whose Culture Has Capital? A Critical Race Theory Discussion of Community Cultural Wealth.” *Critical Race Theory in Education* 114–36.

Noble, Kimberly G. et al. 2015. “Family Income, Parental Education and Brain Structure in Children and Adolescents.” *Nature Neuroscience* 18(5):773–78.

Blalock, Hubert M., Jr. 1964. *Causal Inferences in Nonexperimental Research.* Chapel Hill, NC: University of North Carolina Press.

Chapter 4

Tie Strength, Black Excellence, and Migration Prediction: Results from Logistic Regressions and Machine Learning

Abstract

Weakly tied group members allow for a broader scope of information and resources to diffuse throughout the group than strongly tied groups. Then, in terms of information and resource flow between ties, does a strongly tied personal network inhibit or encourage one's odds of internal migration in the US? And does this effect differ by race and class? This paper found that internal US migrants with strongly tied interpersonal networks were more likely to migrate and potential migrant's expectation to move in the near future was not influenced by interpersonal tie strength but instead were better predicted by whether they lived in a metropolitan area, owned a home, or were employed. Additionally, black respondents in the top 5% in terms of household income are the most likely to migrate within the US, all things being equal, but black respondents in the bottom 5% in terms of household income are the most likely to migrate in general. This study also found that black respondents (in the top and bottom 5%) expectation of moving in the near future were moderated by their interpersonal tie strength, with stronger interpersonal network ties increasing the odds of planning to move in the near future.

Introduction

Within groups, people's opinions and behaviors are more homogeneous than between groups (Burt 2004). Thus, networks with little to no structural holes (individuals connected to non-group members) tend to repeat the same information while rejecting new information from non-group members. In fact, Burt (2004) demonstrates that individuals that serve as bridges across groups are more familiar with alternative ways of thinking and behaving and are less likely than group members not at the peripheral of these structural holes to have their "new" ideas rejected. This implies that one's personal network, to some extent, dictates the quality and quantity of new information they are privy to, as well as their behavioral patterns. Additionally, Burt (2004) emphasizes the critical role "weak ties" play in information access and diffusion. This is because all weak ties are bridges and strong ties are bridges only if the two individuals forming said strong tie have no other strong ties, which is unlikely in all but small groups (Granovetter 1977).

The more two individuals' networks overlap the stronger the ties between them tend to be (Granovetter 1977). Then, could a strongly tied personal network inhibit one's migration inclination? Current migration scholars concerned with the obstruction of migration tend to focus on political, economic, and social barriers such as provincial restrictions in Canada (Zaman 2020), regional income disparities in the US (Kaplan and Schulhofer-Wohl 2017), and the language barrier preventing potential migrants seeking to enter East Germany for work (Dolejš, Glorius, and Hruška 2016). This paper seeks to contribute to this literature by examining how one's personal network may inhibit migration for people living in developed countries such as the US.

The size, frequency of interaction, and proximity of one's social network are affected by race and age (Ajrouch, Antonucci, and Janevic 2001). Blacks and whites tend to be similar in terms of proximity of their social networks but differ in regard to frequency of interaction and size.

Blacks tend to have smaller social networks than their white counterparts and interact with relatives more in general, with greater fact-to-face contact. Age, however, is associated with smaller, less frequent interaction with, and less proximity with social networks, with a higher proportion of one's network being kin at older ages.

The strength of ties within social networks are affected by class, specifically cultural taste impacts the network density of strong and weak ties (Lizardo 2006). High-brow cultural consumption increases the density of strong ties in a network but has no effect on the density of weak ties in said network. While popular cultural consumption has a positive effect on the density of weak ties in a social network with no effect on the density of strong ties in said network. This implies that penetrating the personal network of upper-class individuals requires the knowledge and experience of specific cultural objects not easily accessible by the general public. This is due to the closed nature of strongly tied networks. Thus, it would not be unreasonable to assume that those in the upper class tend to have tightly bound personal networks.

The first hypothesis of this paper is, then, that individuals with stronger ties more than weaker ties are less likely to move, as a strongly tied personal network would serve as stronger anchor than a weakly tied personal network to one's current location. The second hypothesis, by extension, is that individuals with stronger ties more than weaker ties are less likely to anticipate movement in the near future for the reason mentioned in hypothesis 1. Following from hypothesis 1, hypothesis 3 posits that upper class blacks (blacks exhibiting black excellence), given their higher tie strength by virtue of their race and class, are less likely to migrate than lower class blacks. And by extension, hypothesis 4 would posit that they are less likely to anticipate moving in the near future.

This paper structures migration into three possible categories. Two of which (initiation and perpetuation) have been heavily explored by researchers (Erdelen and Richardson 2020; Katz and Stark 1986; Kritz, Lim, and Zlotnik 1992; Lewis 1954; Massey et al. 1993; Piché and Dutreuilh 2013) and the latter (migration inhibition) has received little attention from researchers (Belot and 2011; Li et al. 2016). Thus, in order to fully understand the nature of migration, the structures that might inhibit people's movement must also be explored. Additionally, to my knowledge, no paper to date solely looks at the strength of one's personal network as a direct inhibitor of migration. Instead, scholars focus on how networks serve as a mechanism by which social capital is acquired and distributed in order to sustain initiated migration flows (Kritz, Lim, and Zlotnik 1992; Paul 2019). This paper also seeks to fill this gap in the literature.

The Migration Process: Initiation, Perpetuation, and Termination

The Axiom of Migratory Behavior

Generally, the cycle of mass migration, excluding return migration (which is the reverse of this cycle), follows three fundamental Markovian steps: (1) the initiation or inhibition of migratory behavior, (2) the subsequent perpetuation of this behavior, and (3) its eventual termination. This can be considered the axiom of migratory behavior and is true for internal and international migration alike because all migratory behavior necessarily must be initiated to be perpetrated and subsequently terminated. However, the mechanisms responsible during each step can be and are often independent. That is, the mechanism for migration initiation may be different from the mechanisms for migration perpetuation. However, migration termination tend to be similar to the mechanisms for migration initiation. Massey et al. (1993) reviewed several competing theories that illustrate this process.

According to Massey et al. (1993), a disparate variety of theories have been proposed to explain the initiation of international migration. Each theory (neoclassical economics, the new economics of migration, dual labor market theory, and world systems theory) reveals that the initiation of international migration can be thought of as a function of labor differentials across regions. Kritz, Lim, and Zlotnik (1992) demonstrate structures such as social networks and organizations that can serve as mechanisms that maintain or perpetuate migration patterns. However, neoclassical economic theories on labor migration, upon labor differentials converging towards equilibrium (differentials reaching zero), migration flow also slows to gradual termination. This is an example of how migratory behavior can be initiated by one mechanism yet perpetuated by another mechanism and terminated by the initiating mechanism.

Piché and Dutreuilh (2013) extends Massey et al.'s appraisal by including an additional initiator of migration, political differentials: asymmetries in power and wealth across regions. Unlike the previous examples, political differentials tend to serve as an initiating, perpetuating, and terminating mechanism to migration. Erdelen and Richardson (2020) go a step further than Massey et al. (1993) and Piché and Dutreuilh (2013) by also including environmental initiators such as climate change, extreme weather differentials, and natural resource abundance.

Migration Initiation and Inhibition

Neoclassical economic theory attributes the initiation of migration to geographic labor differentials stemming from burgeoning economies in receiver regions (Lewis 1954; Sjaastad 1962). This can be seen in the wealthy, burgeoning economies of Canada, Norway, and the United Arab Emirates that have become some of the largest receivers of labor migrants (Valenta et al. 2017). New economics of migration theory challenges the assumption of individual actors seeking to maximize returns for their labor made by neoclassical economic theory by proposing that

households, not individuals, seek to minimize the risk of labor market failure (Katz and Stark 1986). That is, actors migrate for greater returns on their labor to cushion the blow a potential economic downfall would have on their family. Thai (2014) found that immigrants, who mostly earn low wages, send a substantial portion of their earnings back home.

Dual labor market theory, however, attributes the initiation of international migration to intrinsic labor demands of modern industrial societies (Piore 1979). That is, the initiation of international migration is due to permanent demands for immigrant labor in developed countries, a bifurcation of labor markets. Campbell (2019) found that false claims of labor shortages by employer associations and campaigns to alter immigration rules in Australia are linked to aspiration to find vulnerable workers willing to accept poor wages and working conditions. This theory, then, strays away from the rational choice models proposed by neoclassical and new economic theories.

World systems theory extends upon dual labor market theory by arguing that economic globalization and market penetration across national boundaries create labor differentials that naturally lead to migration (Sassen 1988). That is, when capitalist firms enter peripheral and semi-peripheral markets in search of land, raw materials, labor, and new consumer markets, they create a mobile population prone to migrate abroad. Poland, for example, by virtue of its semi-periphery position as an immigration center resembling western Europe and a migration periphery such as eastern Europe is viewed peripherally by its own citizens who decided to emigrate and became an immigration sub-center for migrants from less-developed countries (Żołędowski 2020). Finally, extreme political differentials increase the likelihood of conflict within a region which leads migrants to seek freedom from persecution (Zolberg, Suhrke, and Aguauo 1986).

There is a symmetry to the process of migration, notwithstanding its clear complexity. That is, the initiators of migration can also serve as inhibitors (Belot and Ederveen 2011; Li et al. 2016). That is, if the lack of economic, social, political, and environmental differentials or the presence of their corresponding barriers preclude the initiation of migration, networks and institutions sustain such barriers or the lack of differentials. For instance, Belot and Ederveen (2011) found that there is a negative effect of cultural differences on international migration flow between OECD countries. The authors assert that social barriers, such as cultural differences between regions, better explain migration patterns between developed countries than traditional economic variables.

Migration Perpetuation

If the initiation of international migration can be attested to labor and political differentials across regions, then the perpetuation of international migration is sustained through structures and institutions such as social networks and organizations (Massey et al. 1993). Institutional theory asserts that an imbalance between the number of people seeking entry into capital-rich countries and access to limited offered visas gives rise to private institutions and voluntary organizations seeking to satisfy this demand. Thus, organizations gradually develop to sustain, promote, and support international migration, institutionalizing the flow of migrants regardless of the original cause. This is similar to dual labor market theory mentioned earlier and exemplifies one mechanism initiating and perpetuating migration.

Network theory asserts that migrant networks increase the likelihood of international movement by lowering the cost and risks associated with migration, while increasing the expected net returns of movement (Kritz, Lim, and Zlotnik 1992). In fact, Paul (2019) found that Filipino and Indonesian migrant workers in Singapore and Hong Kong have divergent migration

experiences and aspirations due to differences in pre-migration overseas networks; with Filipino migrants being more likely to have known existing migrants before their first migration, having significantly larger networks, being more geographically dispersed, and having networks being comprised more of white-collar contacts. Additionally, the vector of movement grows larger in magnitude with each act of migration, a process called cumulative causation (Mydral 1957). This is because the social context in which migration decisions are made (seeking to increase one's income or land, reduction of agrarian labor in sender region, culture of migration, regional distribution of human capital, and social labeling) all serve to reinforce migration patterns.

Garip and Asad (2016) identified three types of self-feeding social mechanisms (cumulative causation of migration) that underlie network effects in Mexico-US migration: social facilitation (network peers provide information or help that reduces the cost or increases the benefits of migration), normative influence (network peers encourage of discourage migration through social rewards or sanctions), and network externalities (pooled resources common to prior migrants that increase the value or reduces the cost to potential migrants). Garip (2008) also finds that resources from weak ties have a higher effect than resources from strong ties on Thai migrants. These migrants rely more on resources from village members than household members in their migration decision. This builds on the work of Granovetter and Burt, Lin (2000) by showing how one's location in a social network (weak ties) effects the quantity and/or quality of resources (information or help) one has access to.

This paper seeks to illustrate how the nature of personal networks, particularly one's tie strength between themselves and their close ties, may inhibit migration. I utilized data from the Panel Study of Income Dynamics (PSID) to estimate the likelihood of migration given the strength of ties between the respondent and their close friends and relatives. The PSID sample was

originally a nationally representative sample of 18,000 people in 5,000 households and oversampled for low-income families. The sample grows naturally over time due as children and grandchildren are invited to join the PSID. Families are followed regardless of where they live which makes these data excellent for migration studies.

Methods

Measuring Migration

In 2017, 28940 respondents were asked if they have lived anywhere else since January 2015. Specifically, variable *ER66156* asked: "(Have you/Has [he/she]) lived anywhere else since January 2015?" Respondents could have answered: Yes (coded as 1), No (coded as 5), Do not know (DK) (coded as 8), or refuse to answer (coded as 9). Moreover, in variable *ER66158* (following the previous question in variable *ER66156*), respondents were asked which was the most recent year they had moved: 2015, 2016, or 2017. I used the second variables to construct a variable, *Moved in 2 Years,* that indicates whether the respondent moved in 2017 or 2016. I coded movement in 2017 or 2016 as 1 and all other years coded as 0 for no movement. Thus, the recoded respondents that moved in 2015 were subsumed in the category of respondents that did not move at all since January 2015. The respondents that did not know whether they had moved or refused to answer were not included in the analysis. In 2017, 16608 (57%) respondents had not moved since 2015, 6801 (24%) respondents moved in 2017 or 2016, 3028 (10%) respondents moved in 2015, 8 respondents did not or refused to answer (<0%), and 2495 (9%) responses were missing due to one of three reasons: from Latino sample (ER30001=7001-9308); main family nonresponse by 2017 or mover-out nonresponse by 2015 (ER34502=0). Given that the PSID does not follow respondents outside of the US, respondents' migration is internal, that is, respondents are migrating

within the US. Thus, *Moved in 2 years* is a measure of respondent's internal migration in the US. Given that this measure is binary, I will use logistic regression to estimate respondent's likelihood to migrate within the US given the tie strength between themselves and friends and family.

Measuring Future Migration

In 2017, respondents were asked if they thought they might move in the next couple of years. Specifically, variable *ER66162* asked respondents: "Do you think (you/HEAD) might move in the next couple of years?" Respondent's responses were coded as followed: 1 for "Yes, might or maybe," 5 for No, 8 for DK (Do not Know), and 9 for NA (missing) or refused to answer. 9360 (32%) respondents answered "Yes, might, or maybe," 16283 (56%) respondents answered "No," 795 (27%) respondents did not know whether they would move in the next couple of years, 7 (<0%) respondents refused to answer, or responses were missing, and 2495 responses were missing due to one of three reasons: from Latino sample (ER30001=7001-9308); main family nonresponse by 2017 or mover-out nonresponse by 2015 (ER34502=0). Using variable *ER66162,* I created the variable *Plans to Move* by recoding the "No" responses as 0, not changing the coded value for the "Yes" responses and designating all other responses as missing. Similar to the *Moved in 2017* variable, the small amount of missing data (<0%) of the *Plans to Move* variable makes casewise deletion an appropriate method for dealing with missingness. A limitation of this measure of future migration, as well as the variable M*oved in 2017* or *2016,* is that they do not specify a destination. Thus, it is unclear whether or not respondents think they might move or have moved to a different neighborhood, city, or state in the US. Thus, *Plans to Move* is a measure of respondent's inclination to anticipate external and internal migration in the near future. Given that this measure is binary, I will use logistic regression to estimate respondent's likelihood to be inclined to migrate given the tie strength between themselves and friends and family.

Why People Move?

Table 4.1 Reasons for Internal Migration in 2017

Reasons:	Percentage
1. Purposive consumptive (house related): get own home/place; got married; physical conditions of the previous housing unit	31%
2. Purposive consumptive reasons (expansion of housing): more space; more rent; better place	20%
3. Response to outside events (involuntary reasons): Housing unit coming down; being evicted; armed services, etc.; health reasons; divorce; retiring because of health	15%
4. Purposive consumptive (neighborhood related): better neighborhood; go to school; to be closer to friends and/or relatives	10%
5. Purposive consumptive reasons (contraction of housing): less space; less rent	7%
6. Ambiguous, mixed, or other reasons, including reasons such as to save money, all my old neighbors moved away, retiring	6%
7. Purposive productive reasons: to take another job; transfer; stopped going to school	6%
8. To get nearer to work	4%
9. Homeless	1%

Notwithstanding the unclear destination of respondents' migration, the data do allow for something to be said about why respondents moved in 2017. In 2017, respondents were asked why did they move? Specifically, variable *ER66159* asked respondents "Why did (you/he/she) move? Table 4.1 shows a breakdown of the percentages behind migrant's internal migration in US in 2017. Table 4.1 indicates that of the respondents that migrated in 2017, over 65% of them were for consumptive reasons. That is, respondents were expanding or purchasing a new/better home, moving to a better

neighborhood, move closer to work or school, or moving closer to friends and family. Essentially, most respondents migrated for a better life.

Tie Strength

The PSID does not have variables that directly ask respondents to estimate tie strength between themselves and their close ties. Nor does the PSID have a way to construct a network egocentrically. However, the PSID includes multiple measures that can approximate tie strength between respondents and friends/relatives based on whether or not respondents communicated or spent time with their friends and families the previous day. This paper assumes the existence of respondent's friend and family network. Using these measures, I created a 5-point index that approximated the strength of ties between respondents and their close ties (friends and family) as a linear combination of 5 yes or no questions.

In 2016 respondents were asked 5 yes or no questions that corresponded to these PSID variables: *WB16C8A, WB16C10A, WB16C10B, WB16C10C,* and *WB16C13E*. WB16C8A asked respondents "Which of the following CARE activities did you do yesterday? [MARK ALL THAT APPLY]--Helped a friend, neighbor, or relative." WB16C10A asked respondents "Which of the following SOCIAL activities did you do yesterday? [MARK ALL THAT APPLY]--Socialized in person with friends or relatives." WB16C10B asked respondents "Which of the following SOCIAL activities did you do yesterday? [MARK ALL THAT APPLY]--Talked on the phone with friends or relatives." WB16C10C asked respondents "Which of the following SOCIAL activities did you do yesterday? [MARK ALL THAT APPLY]--Texted or emailed with friends or relatives." WB16C13E asked respondents "Who did you spend time with yesterday? [MARK ALL THAT APPLY]--My friends." All respondents responded with yes (coded as 1) or no (coded as 0). Then, I added all the values to construct the variable *Tie Index* that ranged from 0 to 5. Thus,

Table 4.2 Distribution of Respondent Characteristics by Tie Index

	0 (N=587)	1 (N=1582)	2 (N=2072)	3 (N=2029)	4 (N=1418)	5 (N=653)	Overall (N=8341)
Moved in 2 Years							
Mean (SD)	0.133 (0.340)	0.149 (0.356)	0.153 (0.360)	0.145 (0.352)	0.156 (0.363)	0.189 (0.392)	0.152 (0.359)
Median [Min, Max]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]
Missing	24 (4.1%)	64 (4.0%)	72 (3.5%)	73 (3.6%)	55 (3.9%)	25 (3.8%)	313 (3.8%)
Plan to Move							
Mean (SD)	0.275 (0.447)	0.256 (0.436)	0.280 (0.449)	0.270 (0.444)	0.268 (0.443)	0.276 (0.447)	0.270 (0.444)
Median [Min, Max]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]
Missing	46 (7.8%)	93 (5.9%)	115 (5.6%)	112 (5.5%)	80 (5.6%)	36 (5.5%)	482 (5.8%)
More Friends is Better							
Mean (SD)	0.257 (0.437)	0.267 (0.443)	0.224 (0.417)	0.215 (0.411)	0.183 (0.387)	0.177 (0.382)	0.222 (0.415)
Median [Min, Max]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]
Missing	15 (2.6%)	26 (1.6%)	38 (1.8%)	43 (2.1%)	20 (1.4%)	9 (1.4%)	151 (1.8%)
Male							
Mean (SD)	0.608 (0.489)	0.482 (0.500)	0.447 (0.497)	0.402 (0.490)	0.382 (0.486)	0.364 (0.482)	0.437 (0.496)
Median [Min, Max]	1.00 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]
White							
Mean (SD)	0.682 (0.466)	0.635 (0.482)	0.661 (0.474)	0.647 (0.478)	0.665 (0.472)	0.652 (0.477)	0.654 (0.476)
Median [Min, Max]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]
Missing	2 (0.3%)	10 (0.6%)	15 (0.7%)	16 (0.8%)	17 (1.2%)	6 (0.9%)	66 (0.8%)
Education (in Years)							
Mean (SD)	12.9 (3.95)	13.1 (3.80)	13.4 (3.75)	13.4 (3.66)	13.5 (3.71)	13.6 (3.65)	13.3 (3.74)
Median [Min, Max]	13.0 [0, 17.0]	13.0 [0, 17.0]	14.0 [0, 17.0]	14.0 [0, 17.0]	14.0 [0, 17.0]	14.0 [0, 17.0]	14.0 [0, 17.0]
Missing	9 (1.5%)	18 (1.1%)	20 (1.0%)	23 (1.1%)	14 (1.0%)	8 (1.2%)	92 (1.1%)
Lives in Metropolitan Area							
Mean (SD)	0.821 (0.384)	0.831 (0.375)	0.824 (0.381)	0.821 (0.384)	0.826 (0.379)	0.811 (0.392)	0.824 (0.381)
Median [Min, Max]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]
Missing	0 (0%)	7 (0.4%)	1 (0.0%)	3 (0.1%)	5 (0.4%)	3 (0.5%)	19 (0.2%)
Owns Residence							
Mean (SD)	0.687 (0.464)	0.701 (0.458)	0.702 (0.458)	0.683 (0.465)	0.699 (0.459)	0.684 (0.465)	0.694 (0.461)
Median [Min, Max]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]
Missing	31 (5.3%)	68 (4.3%)	64 (3.1%)	72 (3.5%)	48 (3.4%)	24 (3.7%)	307 (3.7%)
Employed							
Mean (SD)	0.641 (0.480)	0.623 (0.485)	0.681 (0.466)	0.673 (0.469)	0.666 (0.472)	0.666 (0.472)	0.661 (0.473)
Median [Min, Max]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]	1.00 [0, 1.00]
Missing	0 (0%)	2 (0.1%)	2 (0.1%)	1 (0.0%)	0 (0%)	0 (0%)	5 (0.1%)

this index can serve as a measure of tie strength between respondents and close ties by measuring respondent's interaction with friends and family the previous day.

Controls

This paper focuses on the tie strength between ego and alters (respondents and friend/relatives) as a perpetuating mechanism that sustains migration. Yet, the lack of migration may be confounded with the migration inhibitors as a perpetuation mechanism. Thus, to separate the two processes I controlled for social, demographic, and economic characteristics that can inhibit migration; that is, sex, race, race by class, education, living in a metropolitan area, owning a home, being employed, and if respondents had ever moved. Political and environmental factors can also initiate or prevent migration. However, given this data was collect in the US in 2016 and 2017, no extreme political instability or environmental disasters or other political and environmental barriers should have significantly affected internal migration pattern in the US.

Sex, race, race by class, living in a metropolitan area, owning a home, being employed, and if respondents had ever moved were all coded as binary variables. Sex was coded as 1 for male and 0 for female. Race was coded as 1 for white and 0 for non-white. I created two variables to measure race by class: blacks in the top 5% of earners in terms of household income (Black Top 5%) and blacks in the bottom 5% of incomes in terms of household income (Black Bottom 5%). Each category was coded as 1 for in their respected groups and 0 otherwise. Respondents living in a metropolitan area were coded as 1 and 0 otherwise. Respondents that owned a home, were in the process of buying a home, or were mobile homeowners renting lots were coded as 1 and coded as 0 if they paid rent or neither paid rent nor owned a home. Respondents that have lived in any other state or country besides where they were living at the time of the survey, including time spent abroad while in the armed forces were coded as 1 and 0 otherwise. Note that lived in means stayed

at least 3 months. Education was measured as the highest grade or year of school the respondent has completed. For example, a value of 8 would indicate that respondents had complete the eighth grade by the time of the 2017 interview. This variable ranged from 0 to 17. Table 2 shows the distribution and descriptive statistics of dependent and control variables by tie index.

Results

Estimating Internal Migration

Table 4.3 shows that respondents' tie strength significantly increased the log-odds of respondents moving in the past 2 years since 2017. Antithetical to my first hypothesis, the results indicate that stronger ties more than weaker ties promote movement rather than inhibiting movement. Respondent's tie index value approximates their daily interaction with friends and family, thus, measuring the closeness or tie strength between themselves and their interpersonal ties. Model 1 through 3 indicates that respondents with tie index values closer to 5 are at higher log-odds than respondents with tie index values closer to 0 of have moved in 2016 and 2017. Moreover, increasing respondent's tie index by 1 unit is associated with a 5% increase in their odds of migrating in 2016 or 2017, which is a small effect size.

Models 2 and 3 look at the base line effect of socioeconomic status of black respondents at the bottom 5 and top 5 percentiles of household income, respectively. That is, the general effects of being in a household of black excellence and being in a household at the bottom of socioeconomic ladder controlling for the tie strength of respondent's personal network. Model 2 indicates that blacks in the bottom 5% in terms of household income had 50% higher odds of migrating in 2016 or 2017 than blacks not in this group after controlling for tie index. Although blacks in the top 5% percent had similar odds of migrating, model 3 shows that this result was not

Table 4.3 Logistic Regression Estimating Migration in the US

	Dependent variable:				
	Moved in the Last 2 Years				
	Model 1	Model 2	Model 3	Model 4	Model 5
Tie Index	0.049**	0.043	0.048**		
	(0.023)	(0.028)	(0.023)		
Black (Bottom 5%)		0.400***		-0.004	
		(0.141)		(0.046)	
Black (Top 5%)			0.391		0.995***
			(1.007)		(0.295)
Male				-0.030	-0.031
				(0.044)	(0.044)
Education				-0.007	-0.008
				(0.006)	(0.006)
Lives in Metropolitan Area				0.058	0.052
				(0.063)	(0.062)
Owns Residence				-1.690***	-1.696***
				(0.046)	(0.046)
Employed				0.239***	0.235***
				(0.048)	(0.048)
Moved Ever				0.018	0.017
				(0.047)	(0.046)
Tie Index X Black (Bottom 5%)		0.004			
		(0.048)			
Tie Index X Black (Top 5%)			0.002		
			(0.321)		
Constant	-1.841***	-1.961***	-1.845***	-0.599***	-0.589***
	(0.066)	(0.081)	(0.066)	(0.100)	(0.097)
Observations	8,028	7,958	7,958	14,261	14,261
Log Likelihood	-3,420.241	-3,364.176	-3,382.585	-6,625.298	-6,620.277
AIC	6,844.482	6,736.351	6,773.170	13,266.600	13,256.550

Note: Standard errors are in paratheses. *p<0.05, **p<0.01, ***p<0.001

statistically significant. I also added interactions terms to both models to determine if black respondent's decision to migrate who were in the top and bottom 5% in terms of household income was affected by tie strength. In both models, not only were the results near zero but they were not statistically significant. Thus, I did not include tie strength of the interaction terms in the next two models. Models 2 and 3 serve as the base line results for migration with respect to tie strength and black excellence (as measured by household income).

Models 4 and 5 controls for respondent's social, demographic, and economic characteristics that tend to affect migration decisions, as well as respondent's previous migration history. Specifically, respondent's sex, race, education, residential location (living in a metropolitan or non-metropolitan area), home ownership status, employment status, and migration history. Model 4 indicates that the log-odds associated with black respondents at the bottom 5% effectively became 0 and was not statistically significant. This implies that respondents in the bottom 5% with similar social, demographic, and economic characteristic odds of migrating were equal with those not in that group. The log-odds associated with black respondents at the top 5% in model 5, however, significantly increased from model 3 after controlling for social, demographic, and economic characteristics that affect migration decisions. Model 5 indicates that blacks in the top 5% are 2.72 times the odds to have moved in 2016 or 2017 than black respondents not in this group. In fact, blacks in this group are the most likely to move among the groups being analyzed in this study.

Based on the AIC estimator of each model, model 5 loses the least amount of information compared to the other four models and, thus, is of the highest quality in the group. Controlling for respondent's prior movement history, and sociodemographic characteristics, Model 5 indicates that the largest predictor of migration in this study is being in a household of black excellence.

Predicting Migration

Table 4.4 indicates that respondent's tie strength does not significantly increase respondent's inclination towards moving in the coming years after 2017. Thus, I found no evidence that stronger ties more than weaker ties increased respondent's plans to migrate in the near future.

Model 6 serves as a baseline for estimating respondents' expectations of movement in the years following 2017 and shows that respondents tie index has nearly 0 association with their odds of anticipating movement. Model 7 and model 8 are similar to models 2 and 3, except the dependent variable is plans to move in the near future. Similar to model 2, blacks at the bottom 5% of the socioeconomic ladder have 30% higher odds of planning to move in the near future than blacks not in this group. I also added an interaction term to examine the if black respondents in the top and bottom 5% in terms of income anticipation of moving in the near future was affected by tie strength. Models 7 and 8 indicate that they are, more so for model 8 than 7.

After controlling for the interaction between being black and in the top % in terms of household income, model 8 indicates that black respondents in this income bracket had 83% lower odds to anticipate moving in the near future. However, when I looked at the moderating effects of tie strength, model 8 indicates that the odds double for this group the stronger their tie strengths. This implies that, black respondents in the top %5 in terms of household income highly anticipate moving in the near future the stronger their interpersonal network is. Tie strength seems to be an important predictor anticipated movement for black respondents in the top income brackets. One can compare this with the importance of tie strength for all races and statuses utilizing machine learning algorithms instead to predict migration (see Appendix A for this analysis)

Table 4.4 Logistic Regression Estimating Migration in Near Future

	Dependent variable:				
	Plans to Move in Near Future				
	Model 6	Model 7	Model 8	Model 9	Model 10
Tie Index	0.006	-0.027	0.002	-0.011	0.008
	(0.019)	(0.023)	(0.019)	(0.025)	(0.021)
Black (Bottom 5%)		0.262**		-0.091	
		(0.116)		(0.130)	
Black (Top 5%)			-1.781*		-1.141
			(1.077)		(1.103)
Male				0.007	0.001
				(0.058)	(0.058)
Education				-0.012	-0.014
				(0.011)	(0.011)
Lives in Metropolitan Area				0.250***	0.259***
				(0.080)	(0.080)
Owns Residence				-1.457***	-1.477***
				(0.059)	(0.058)
Employed				0.229***	0.229***
				(0.063)	(0.063)
Moved Ever				0.005	-0.001
				(0.059)	(0.058)
Tie Index X Black (Bottom 5%)		0.083**		0.068	
		(0.040)		(0.045)	
Tie Index X Black (Top 5%)			0.704**		0.600*
			(0.324)		(0.340)
Constant	-1.009***	-1.070***	-0.996***	-0.314*	-0.294*
	(0.053)	(0.064)	(0.053)	(0.175)	(0.165)
Observations	7,859	7,792	7,792	7,262	7,262
Log Likelihood	-4,585.831	-4,511.053	-4,546.762	-3,816.187	-3,815.471
AIC	9,175.661	9,030.106	9,101.524	7,652.373	7,650.941

Note: Standard errors are in paratheses. *p<0.05, **p<0.01, ***p<0.001

Models 9 and 10 control for respondent's social, demographic, and economic characteristics that tend to affect migration decisions, as well as respondent's previous migration history. Specifically, respondent's sex, race, education, residential location (living in a metropolitan or non-metropolitan area), home ownership status, employment status, and migration history. However, after controlling for these characteristics, tie index and being black in the bottom and top 5% in terms of household income had no effect on the odds of planning to move in the near future. In fact, the largest factor inhibiting respondent's plans to migrate in the near future after 2017 was owning a home, with owning a home decreasing respondent's odd of planning to migrate by 77%. Moreover, after the controls, the interaction term for black respondents in the top 5% in terms of household income and tie strength decreased slightly but still nearly doubled, at 1.82 times, the odds of anticipating migration in the near future.

Based on the AIC estimator of each model, model 10 loses the least amount of information compared to the other four models and, thus, explains the most amount variance. Controlling for respondent's prior movement history, and sociodemographic characteristics, model indicates that respondent's tie index increases their odds of anticipating movement in the coming years after 2017, with the magnitude of this change being higher than models' 4 and 5 estimations but was not significant. Thus, this paper finds little evidence that tie strength or black excellence increases the likelihood of migration inclination. On the contrary, this study finds being black at the bottom of the socioeconomic ladder increases migration inclination in general.

Discussion

Mass migration is intrinsically a complex process. However, belying this complexity is an axiomatically Marchovian process. That is, the process of mass migration always follows a series

of three independent steps: migratory initiation, perpetuation, and termination. The reversal of this process would be return migration. I called this *the axiom of migratory behavior*. Within each step of the migration process contains mechanisms that can be and are often times independent from one another. These mechanisms tend to initiate, perpetuate, or terminate migration. This paper focused on the perpetuating mechanism that is social networks.

Network theory asserts that social networks, specifically migrant networks increase the likelihood of movement by lowering the cost and risks associated with migration, while also increasing the expected net returns of movement (Kritz, Lim, and Zlotnik 1992). Garip and Asad (2016) identified three types of self-feeding social mechanisms that underlie network effects on migration: social facilitation (network peers provide information or help that reduces the cost or increases the benefits of migration), normative influence (network peers encourage of discourage migration through social rewards or sanctions), and network externalities (pooled resources common to prior migrants that increase the value or reduces the cost to potential migrants). Paul (2019) found that Filipino more than Indonesian migrant workers in Singapore and Hong Kong were more likely to have known existing migrants before their first migration, have significantly larger networks that are geographically dispersed, and have networks comprised more of white-collar contacts; these network differentials led to divergent migration experiences and aspirations.

Garip (2008) found that resources (information) from weakly tied prior migrants (village members) had a stronger effect than resources from strongly tied prior migrants (household members) on Thai migrant's internal migration decision. That is, Thai migrants relied more on resources from non-household members than household members when migrating within Thailand. Davis et al. (2002) and Palloni et al. (2001) find that for international migration, specifically Mexico to US migrants tend to rely on strong ties (family members) to migrate. These

studies operated under the assumption that family members were strong ties while non-family members were weak ties.

To contrast these findings this paper took a more general and direct approach to estimating whether strong ties compared to weak ties have a stronger effect on migration decisions by measuring tie strength directly, assuming the existence of interpersonal ties, considering family members and non-family members together (as strong ties can be between non-family members and family members), and considered the effect that non-prior migrants as well as prior migrants have on potential migrants.

This paper makes two important contributions to migration studies and network analysis. The first contribution of this paper is that the results imply that internal migrants in the US might rely more on resources from strong ties than weak ties in general. This finding is not congruent with previous findings in that they imply that internal migrants rely more on information from strong ties than weaker ties in the US. These findings also highlight the importance of stronger ties rather than weaker ties in internal migration decisions among US internal migrants. The second contribution of this paper is that it highlights the importance of race and class when it comes to internal migration in the US. Particularly for black people in the bottom 5% and top 5% of the socioeconomic ladder. In general, black folk in the bottom 5% are more likely to move and plan to move in the near future. This is mostly likely due to the inherent housing insecurity that comes with this income bracket. However, after controlling for factor inhibiting movement, blacks in the top 5% (black excellence) are significantly more likely to migrate compared to all respondents in this study.

The results of this paper also suggest that social networks in conjunction with race and class do in fact increase the likelihood of migration as well as the plans to migrate in the near future

through cumulative causation mechanisms described earlier. Black people in the top 5% plans to move in the near future are doubled and moderated by their tie strength. Given, that high earning blacks have the highest odds of having had migrated and when they have strong interpersonal network ties increases their plans to migrate in the near future, this result aligns with the current migration patterns of black people in the US. This is because blacks are leaving non-Southern metropolitan areas to reside in metropolitan areas at higher rates than any other race. Moreover, most black Americans have strong roots in the South. In conjunction with the cumulative causation mechanisms, these results support the idea of networks as perpetuating factors of migration.

However, a limitation of this paper is that it does not include qualitative analysis to ascertain to what extent potential migrant's migration decisions were influenced by their strong ties. Future researchers can build on these findings to include a qualitative component in order to ascertain to what extent potential migrants rely on strong ties to inform their migration decisions.

A second limitation of this paper is that it is unclear where migrants moved to in 2017 of 2016. Although it is clear that respondents moved somewhere in the US in 2017 or 2016, future research can use these results to compare US migrant's international and internal migratory patterns as it pertains to their tie strength between themselves and their interpersonal network. Prior research found that weak ties are more important than strong ties for internal migration and strong ties more than weak ties are more important for international migration. However, prior research did not measure tie strength directly. Future researchers can build on the results of this paper by distinguishing between internal and international migrants given a direct measure of migrant's tie strength with friends and family and prior and non-prior migrants.

References

Bansal, G., M. S. Yadav, and S. K. Kaur. 2020. "Predicting Migration Flows: A Machine Learning Approach." *Journal of Migration Studies* 33(4): 233-249.

Greenwood, M. J., L. T. Smith, and G. M. Thomas. 2019. "Economic Drivers of Internal Migration: A Machine Learning Analysis." *Urban Studies* 56(8): 1716-1731.

Kusumastuti, Y., R. K. Narayan, and S. R. Ward. 2021. "Analyzing Internal Migration: A Clustering Approach Using Machine Learning." *Journal of Population Research* 38(1): 59-74.

Ajrouch, K. J., T. C. Antonucci, and M. R. Janevic. 2001. “Social Networks among Blacks and Whites: The Interaction between Race and Age.” *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 56(2).

Belot, M. & Ederveen, S. (2011). Cultural Barriers in Migration between OECD Countries. *Journal of Population Economics*, 25(3),1077–1105.

Burt, R.S. (2004). Structural Holes and Good Ideas. *American Journal of Sociology*, 110(2),349–99.

Campbell, I. (2019). Harvest Labour Markets in Australia: Alleged Labour Shortages and Employer Demand for Temporary Migrant Workers. *Journal of Australian Political Economy*, (84),46–88.

Dolejš, M., Glorius, B., Hruška, V. (2016). Motives and Barriers of Migration to Saxony: The Case of Migrating Health Professionals from Czechia. *GeoScape*, 10(2),62–77.

Erdelen, W.R., & Richardson, J.G. (2020). Human Migration: Managing Its Increasing Complexity. *FORESIGHT,* 22(1),109–26.

Garip, F., & Asad, A.L. (2016). Network Effects in Mexico–U.S. Migration. *American Behavioral Scientist,* 60(10),1168–93.

Garip, F., (2008). Social Capital and Migration: How Do Similar Resources Lead to Divergent Outcomes? *Demography,* 45(3),591–617.

Granovetter, M.S., (1977). The Strength of Weak Ties. *Social Networks,* 347–67.

Kaplan, G., & Schulhofer-Wohl, S. (2017). Understanding The Long-Run Decline In Interstate Migration. *International Economic Review,* 58(1),57–94.

Katz, E. & Stark, O. (1986). Labor Migration and Risk Aversion in Less Developed Countries. *Journal of Labor Economics,* 4(1),134–49.

Kritz, M.M., Lim, L.L, Zlotnik, H. (1992). Migration Networks and the Shaping of Migration Systems. Pp. 150–76 in *International Migration Systems: A Global Approach*. Oxford: Clarendon Press.

Lewis, W.A. (1954). Economic Development with Unlimited Supplies of Labour. *The Manchester School,* 22(2),139–91.

Li, X., Huang, S., Chen, J., Chen, Q. (2020). Analysis of the Driving Factors of U.S. Domestic Population Mobility. *Physica A: Statistical Mechanics and Its Applications*, 539,122984.

Li, X., et al. (2016). Characterizing the International Migration Barriers with a Probabilistic Multilateral Migration Model. *Scientific Reports,* 6(1).

Lin, N. (2000). Inequality in Social Capital. *Contemporary Sociology,* 29,785–95.

———. (2001). *Social Capital: A Theory of Social Structure and Action*. Cambridge: Cambridge

Lizardo, Omar. 2006. "How Cultural Tastes Shape Personal Networks." *American Sociological Review* 71(5):778–807.

Massey, D.S., et al. (1993). Theories of International Migration: A Review and Appraisal. *Population and Development Review,* 19(3),431.

Myrdal, G. (1957). *Rich Lands and Poor: The Road to Prosperity*. New York, NY: Harper.

Paul, A.M. (2019). Unequal Networks: Comparing the Pre-Migration Overseas Networks of Indonesian and Filipino Migrant Domestic Workers. *Global Networks,* 19(1),44–65.

Piché, V., & Dutreuilh, C. (2013). Contemporary Migration Theories as Reflected in Their Founding Texts. *Population (English Edition),* 68(1),141.

Piore, M.J., (1979). *Birds of Passage: Migrant Labor and Industrial Societies*. Cambridge, UK: Cambridge University Press.

Sassen, S. (1988). *The Mobility of Labor and Capital: A Study in International Investment and Labor Flow*. Cambridge: Cambridge Univ. Press.

Sjaastad, L.A. (1962). The Costs and Returns of Human Migration. *Journal of Political Economy*, 70(5, Part 2),80–93.

Thai, H.C. (2014). *Insufficient Funds the Culture of Money in Low-Wage Transnational Families*. Stanford, CA: Stanford Univ. Press.

Valenta, M., Strabac, Z., Jakobsen, J., Reitz, J., Awad, M.A. (2017). Labour Migrations to Resource-Rich Countries: Comparative Perspectives on Migrants' Rights in Canada, Norway and the United Arab Emirates. *International Journal on Minority and Group Rights*, 24(2),150–73.

Zaman, M.R. (2020). Do Subnational Agreements Induce Interprovincial Migration? Empirical Evidence from Canada's Aggregate Migration Patterns. *Canadian Public Policy*, 46(2),279–89.

Zolberg, A.R., Suhrke, A., Aguayo, S. (1986). International Factors in the Formation of Refugee Movements. *International Migration Review*, 20(2),151.

Żołędowski, C. (2020). Poland in International Migrations: The Perspective of World Systems Theory. *Miscellanea Geographica*, 24(2),94–103.

Chapter 5

Main Discussion and Conclusion

Main Discussion

This study builds on the neoclassical macroeconomic theory of migration by introducing the concept of labor gradients, a framework that models labor flows in response to labor market differentials. Similar to the way temperature gradients drive the flow of heat, labor gradients describe how disparities in job opportunities induce migration patterns across regions. By measuring differences in labor market opportunities—operationalized through call-back rates—between two geographic areas, this study provides insight into how race and class interact to shape labor market outcomes for black and white applicants.

Mass migration is a complex process. The process of mass migration always follows a series of three independent steps: migratory initiation, perpetuation, and termination. The reversal of this process would be return migration. I called this *the axiom of migratory behavior*. Within each step of the migration process contains several mechanisms that can be and are often times independent from one another. These mechanisms tend to initiate, perpetuate, or terminate migration. This paper focused on the perpetuating mechanism that is social networks.

Network theory asserts that social networks, specifically migrant networks, increase the likelihood of movement by lowering the cost and risks associated with migration, while also increasing the expected net returns of movement (Kritz, Lim, and Zlotnik 1992). Garip and Asad (2016) identified three types of self-feeding social mechanisms that underlie network effects on migration: social facilitation (network peers provide information or help that reduces the cost or increases the benefits of migration), normative influence (network peers encourage of discourage migration through social rewards or sanctions), and network externalities (pooled resources common to prior migrants that increase the value or reduces the cost to potential migrants). Paul

(2019) found that Filipino more than Indonesian migrant workers in Singapore and Hong Kong were more likely to have known existing migrants before their first migration, have significantly larger networks that are geographically dispersed, and have networks comprised more of white-collar contacts; these network differentials led to divergent migration experiences and aspirations.

To contrast these findings this study took a more general and direct approach to estimating whether strong ties compared to weak ties have a stronger effect on migration decisions by measuring tie strength directly, assuming the existence of interpersonal ties, considering family members and non-family members together (as strong ties can be between non-family members and family members), and considered the effect that non-prior migrants as well as prior migrants have on potential migrants.

This study also directly analyzes the causal structure of the effect of childhood SES on one's current household income. Specifically, this study looked at respondents who grew up in black households of excellence, where respondent's parents either held a college degree or graduate degree. There were several key findings on both fronts. An analysis of the path diagrams showed that the path coefficients for the effect of childhood health (youth health) on educational aspirations, the effect of childhood health on educational attainment, and the correlation between current health and current household income varied by race and the educational attainment of black parents.

Key Findings and Their Implications

The findings suggest that the labor gradient for black applicants residing in affluent neighborhoods was the largest among all applicants, whereas the labor gradient for white applicants in affluent neighborhoods was the smallest. Although this result was not statistically

significant, it aligns with prior research by Frey (2004), which highlights the migration of high SES black people to the South while the migration of whites into the region has been less pronounced. This suggests that black excellence may be a key push and pull factor of black migration to the South. Moreover, class appears to have a differential effect on labor gradients for black and white applicants, with affluent blacks experiencing the most labor market opportunities in the South, whereas affluent whites experience the least.

The call-back rates further contextualize these findings. Quillian et al. (2017) reported that hiring discrimination against African Americans in the U.S. has remained constant since 1989, with an increase in discrimination in the Northeast. While this study does not confirm or refute these claims due to a lack of statistically significant differences, the call-back data offer some suggestive patterns. In Atlanta, Black and white applicants received equal overall call-back rates, whereas in New York City, Black applicants received 20% fewer call-backs than their white counterparts. When class was included, affluent Black applicants in Atlanta received 20% more call-backs than affluent white applicants, while in New York City, affluent Black applicants received 50% fewer call-backs than affluent whites and 20% fewer call-backs than poor Black applicants. These findings suggest that Black excellence is rewarded in Atlanta but penalized in New York City, reinforcing the idea that regional variations in labor market discrimination may influence Black migration patterns.

While these results were not statistically significant, they point to an important theoretical consideration: Black excellence may be differently perceived and valued depending on regional context. Historically, the South has served as the cultural epicenter of Black excellence, emerging from the Civil Rights Movement. If the celebration of Black success is more ingrained in the South than in other regions, this could explain why affluent Black applicants experience greater

economic opportunities in Atlanta than in New York City. Additionally, it is possible that white employers are supportive of Black socioeconomic advancement only to the extent that it does not surpass their own status. Future research should further explore the potential negative effects of Black excellence, as the literature has not yet fully addressed this growing concern.

This study also found that the better the childhood health of the respondents the less educational aspirations they had. However, this effect is reversed in black households and more than doubled in households of black excellence. However, the effect of childhood health on actual educational attainment for households of black excellence is negative and double the average, which is expected as poorer health as a child would necessarily limit one's potential for higher education.

For respondents who grew up with black parents or guardians with graduate degrees the correlation between their current health and current household income is double the correlation of the average household and over triple that of white household in terms of magnitude. The path coefficient for the effect of respondent's parent's education on current household income was similarly close to zero for all models except for respondent's with at least one black parent with at least a bachelor's degree. This result supports the "stickiness" theory of SES. That is, many people are not upwardly or downwardly mobile but instead spend most of their lives in the class they were born in. However, the path coefficient for respondents who grew up in households of black excellence, specifically with parents or guardians holding at least a bachelor's degrees saw the largest gain in household income compared to all other respondents including black respondents in general. These findings suggest that respondents with black parents holding at least a bachelor's degree were most likely out of all respondents to see upward mobility.

An analysis of the structural equation for the total effect of childhood SES on current household income also shows that the path through education is the most important by being the largest contributor to the total effect. However, according to the results, this path is more important for respondents in white households than black households, with this pathway contributing 59 percent to the total effect for white households and 47 percent for black households. The contribution is even less for respondents with black parents with graduate degrees at 30 percent.

Additionally, when educational aspirations are included into the pathway mentioned above, the contribution to the total effect becomes greatest for black households than white households but not by much. This may suggest that the educational aspirations of respondents who grew up in black households of excellence were less important to their financial future when compared to other households. It is beyond the scope of this study to reach any conclusion as to why this may be the case, but future researcher should focus on this line of inquiry.

Moreover, this study makes two important contributions to migration studies and network analysis. The first contribution of this study is that the results imply that internal migrants in the US might rely more on resources from strong ties than weak ties in general. These findings also highlight the importance of stronger ties rather than weaker ties in internal migration decisions among US internal migrants. The second contribution of this paper is that it highlights the importance of race and class when it comes to internal migration in the US. Particularly for black people in the bottom 5% and top 5% of the socioeconomic ladder. In general, black folk in the bottom 5% are more likely to move and plan to move in the near future. This is mostly likely due to the inherent housing insecurity that comes with this income bracket. However, after controlling for factor inhibiting movement, blacks in the top 5% (black excellence) are significantly more likely to migrate compared to all respondents in this study.

The results of this paper also suggest that race and class do in fact increase the likelihood of migration and in conjunction with social networks the plans to migrate in the near future through cumulative causation mechanisms described earlier. Black people in the top 5% plans to move in the near future are doubled and moderated by their tie strength. Given, that high earning blacks have the highest odds of having had migrated and when they have strong interpersonal network ties increases their plans to migrate in the near future, this result aligns with the current migration patterns of black people in the US. This is because blacks are leaving non-Southern metropolitan areas to reside in metropolitan areas at higher rates than any other race. Moreover, most black Americans have strong roots in the South. In conjunction with the cumulative causation mechanisms, these results support the idea of networks as a perpetuating factor of migration.

Limitations and Directions for Future Research

Several limitations of this study should be acknowledged. First, employer recognition of neighborhood characteristics in Atlanta posed a challenge. Unlike cities with clearly defined neighborhood distinctions, Atlanta's street addresses do not always indicate a specific neighborhood, which could affect how employers perceive applicants' backgrounds. Second, while this study establishes regional differences in economic opportunities for Black applicants, it does not identify the underlying reasons why employers might perceive Black applicants as more employable in Atlanta than in New York City. The economic development of the South since the 1970s, coupled with an improving racial climate for middle-class Black individuals, may partly explain these differences, but additional research is needed to confirm these mechanisms.

Furthermore, this study does not address why the Great Migration has reversed, nor does it measure the broader economic, labor market, and housing implications of this demographic shift. Migration is inherently complex, involving numerous factors beyond labor market differentials.

Future research should incorporate ethnographic and survey methodologies to better understand the motivations behind Black migration to the South and its impact on urban economies.

For the causal structural analysis portion of this study, there is no algorithm for calculating total effects using calculus in sociology. This means that, to implement this method, researchers may have to do all calculations by hand. For a small path diagram such as the one in this study, calculating the total effect for long pathways merely take approximately 3 to 4 pages per pathway. However, for larger path diagrams, calculations can easily balloon to 10 pages or more for the analysis of just one pathway. Researchers should focus on developing an algorithm for this method. Moreover, this algorithm should also be developed for computer software such as STATA (the main program most researchers utilize to calculate total effects). This would mean that researchers need not be well versed in calculus to perform these calculations. However, the author highly recommends a basic understanding of calculus to correctly implement this method and interpret subsequent results.

Moreover, data may be biased in two ways. First, respondents may be limited in regard to their memory. However, the bias here is likely to be negligible given that questions focused on significant and generalized collections of moments in their childhood and adulthood. That is, their health as a child, mother's and father's education, educational aspirations, educational attainment, household income, and current health status.

For the migration analysis portion of this study, the limitation of this study is that it does not include qualitative analysis to ascertain to what extent potential migrant's migration decisions were influenced by their strong ties. Future researchers can build on these findings to include a qualitative component in order to ascertain to what extent potential migrants rely on strong ties to inform their migration decisions.

Additionally, it is unclear where migrants moved to in 2017 of 2016. Although it is clear that respondents moved somewhere in the US in 2017 or 2016, future research can use these results to compare US migrant's international and internal migratory patterns as it pertains to their tie strength between themselves and their interpersonal network. Prior research found that weak ties are more important than strong ties for internal migration and strong ties more than weak ties are more important for international migration. However, prior research did not measure tie strength directly. Future researchers can build on the results of this paper by distinguishing between internal and international migrants given a direct measure of migrant's tie strength with friends and family and prior and non-prior migrants.

Theoretical Contributions and Future Research Directions

Despite the complexities of migration, this study proposes that migration can be conceptualized as a physical process modeled using thermodynamic principles. The notion of migration gradients—labor gradients, network gradients, and security gradients—provides a systematic approach to studying migration. Labor gradients reflect differences in labor supply relative to capital, network gradients capture variations in social ties, and security gradients account for disparities in crime and safety. This framework offers a structured way to analyze migration patterns and could contribute to the development of a unified theory of migration.

This study also contributes to understanding the causal structure of socioeconomic mobility, specifically the relationship between childhood socioeconomic status (SES) and adult household income. By analyzing structural equations and path diagrams, the study uncovers key racial differences in the effects of childhood SES. Notably, black people who grew up in households of Black excellence—where parents held at least a bachelor's degree—experienced

the greatest upward mobility. This contrasts with the relative SES "stickiness" observed among white individuals, where childhood SES had a stronger direct effect on adult income.

An important finding from the path analysis is that childhood health negatively influenced educational aspirations on average, but this effect was reversed by black excellence. However, childhood health had a negative impact on actual educational attainment among black individuals with highly educated parents, suggesting that poor health still constrains upward mobility despite high aspirations. Additionally, for individuals raised by black parents with graduate degrees, the correlation between current health and income was the strongest among all groups, indicating that health and economic well-being are more closely linked for this demographic.

The analysis of total effects further supports these findings. Compared to white households, black households exhibited less SES stickiness, meaning that black individuals had greater potential for mobility. However, among black individuals, those raised in households of black excellence experienced the most significant gains in SES over their lifetime. The educational pathway was the most critical determinant of economic mobility, though its influence was more pronounced for white individuals than for black individuals. The reduced importance of educational aspirations in black households suggests a potential area for future research on how aspirations translate into economic outcomes across racial groups.

For migration studies, black excellence should be considered a major push and pull factor for black migrants. Within the context of the New York City and Atlanta labor markets, my findings suggest that black people experience a pull towards Southern metropolitan areas and push away from non-Southern metropolitan areas. This suggests that, after a certain amount of success, black folk do not see the same returns on their human capital in non-Southern areas. Future

research should focus on other cities such as Chicago, Detroit, and other prominent non-Southern black metropolitan areas to see if these results hold true around the country.

Conclusion

This study highlights the complex interplay between race, class, and regional labor market opportunities. This study offers preliminary evidence that black excellence is an important factor in the New Great Migration. Additionally, the migration of highly educated black individuals to the South aligns with broader historical and economic trends, suggesting that regional cultural and economic factors shape labor market outcomes.

The findings underscore the need for further research on the effects of black excellence beyond academic success, regional variations in labor market dynamics, and the broader implications of the New Great Migration. Future studies should explore these dynamics using mixed methods approaches, including ethnographic and longitudinal analyses, to provide a more comprehensive understanding of black migration and socioeconomic mobility. By conceptualizing migration through the lens of gradients, black excellence, and social networks, this study offers a novel framework for analyzing the mechanisms driving labor market dynamics and mobility patterns across different geographic regions.

Appendices

Appendix A: Machine Learning Approach to Migration Prediction

Machine learning has proven to be a valuable tool for understanding internal migration. This section explores how machine learning can be applied to internal migration studies, focusing on the prediction of migration and analysis of migration drivers. Researchers apply ML models like decision trees, random forests, and gradient boosting to identify the factors that drive migration within countries. These models can predict where individuals are likely to move based on variables such as employment opportunities, housing prices, educational institutions, and access to social services (Bansal et al. 2020). For instance, studies have used machine learning to predict the migration of young adults from rural areas to urban centers, driven by factors such as job availability and education (Greenwood et al. 2019). By incorporating demographic data, economic conditions, and urbanization trends, machine learning models can provide valuable insights into future movement, which can help policymakers and urban planners prepare for changes in local populations and infrastructure demands.

Understanding the socio-economic factors that drive internal migration is another area where machine learning has been applied. Traditional theories of internal migration emphasize economic factors—such as income disparities, employment opportunities, and housing availability—as key drivers. Machine learning may allow researchers to move beyond correlations, uncovering more nuanced patterns in the data that may not be clear through conventional statistical methods. For example, by applying clustering algorithms to demographic and economic data, researchers can identify regions with similar migration trends, helping to categorize areas based on common socio-economic characteristics and migration behaviors (Kusumastuti et al. 2021).

Machine learning has become a powerful tool in the study of internal migration, providing sociologists with new ways to analyze migration patterns, identify drivers of migration, and assess the socio-economic impacts of migration flows. This study uses a random forest and logistic regression to identify the key features in predicting migration.

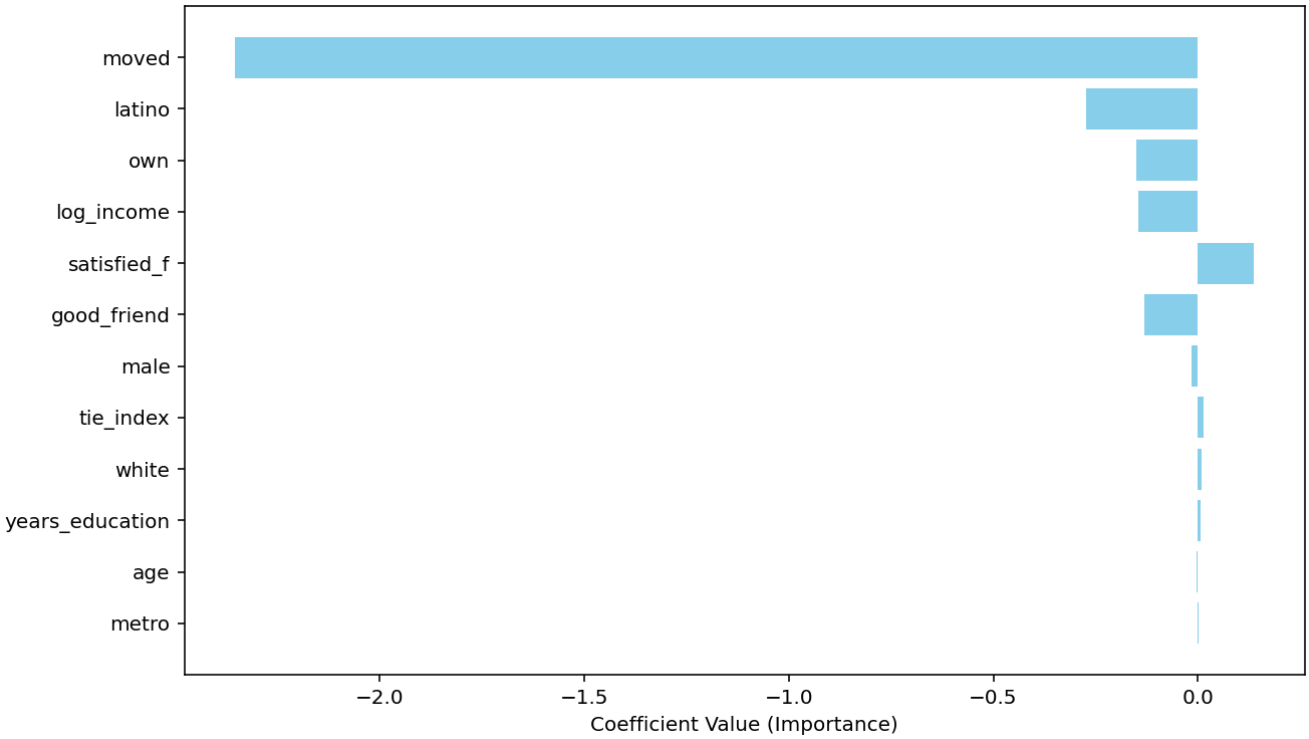
Random Forest is an ensemble learning algorithm used for both classification and regression tasks. It builds multiple decision trees and aggregates their individual predictions to improve accuracy and reduce overfitting. The algorithm works by creating decision trees through bootstrap sampling, where each tree is trained on a random subset of the data, and by selecting a random subset of features at each split, which ensures that trees are diverse and not overly reliant on any single feature. For classification tasks (migrated or did not migrate), the final prediction is determined by a majority vote from all the trees, while for regression tasks, the prediction is the average of the individual tree predictions. This approach makes random forests highly robust and less prone to overfitting compared to single decision trees. One of the key advantages of random forests is their ability to handle missing data and model complex, non-linear relationships between variables. Additionally, random forests can provide insights into *feature importance*, making it useful for feature selection.

However, random forests come with some limitations, including a lack of interpretability—while individual trees are easy to understand, the ensemble of many trees can become a "black-box" model. Furthermore, training a large number of trees can be computationally expensive and may require significant memory, especially with large datasets. Logistic regression can also serve as a machine learning algorithm to analysis feature importance. This study uses both random forest and logistic regression to gain insight on the key factors influencing US internal migration (See appendix C. for code).

Feature Importance

Feature importance can help determine which variables significantly influence model predictions. In Random Forest, importance is typically measured using Mean Decrease in Impurity (MDI) and Permutation Importance (MDA). MDI calculates how much each feature reduces impurity (e.g., Gini impurity) when used for splitting in decision trees. MDA, a more robust method, shuffles feature values and measures the drop in model performance, capturing both individual effects and feature interactions. This paper uses scikit-learn library which utilizes MDI for Random Forest feature importance estimations.

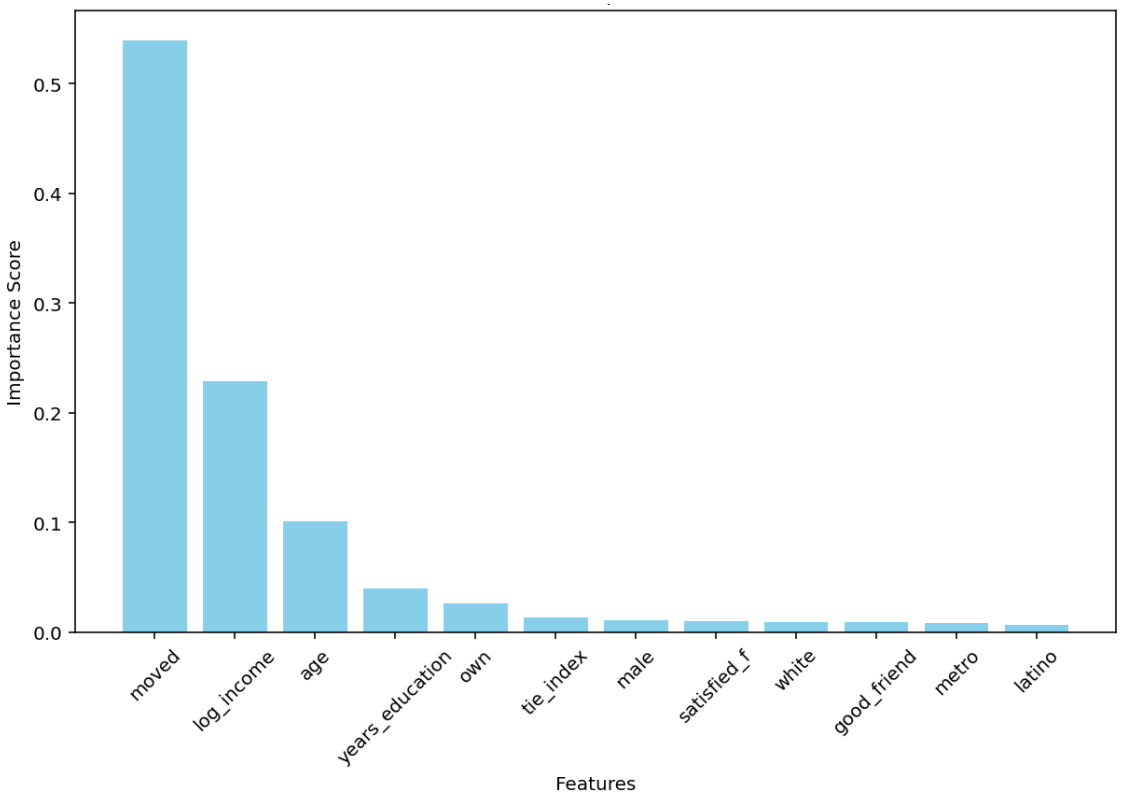
Figure 1A Feature Importance Using Logistic Regression



In Logistic Regression, feature importance is derived from the magnitude of coefficients. Larger absolute values indicate stronger influence on the predicted probability. However, since

coefficients depend on the scale of the input features, standardizing them allows for better comparison.

Figure 2A Feature Importance for Random Forest

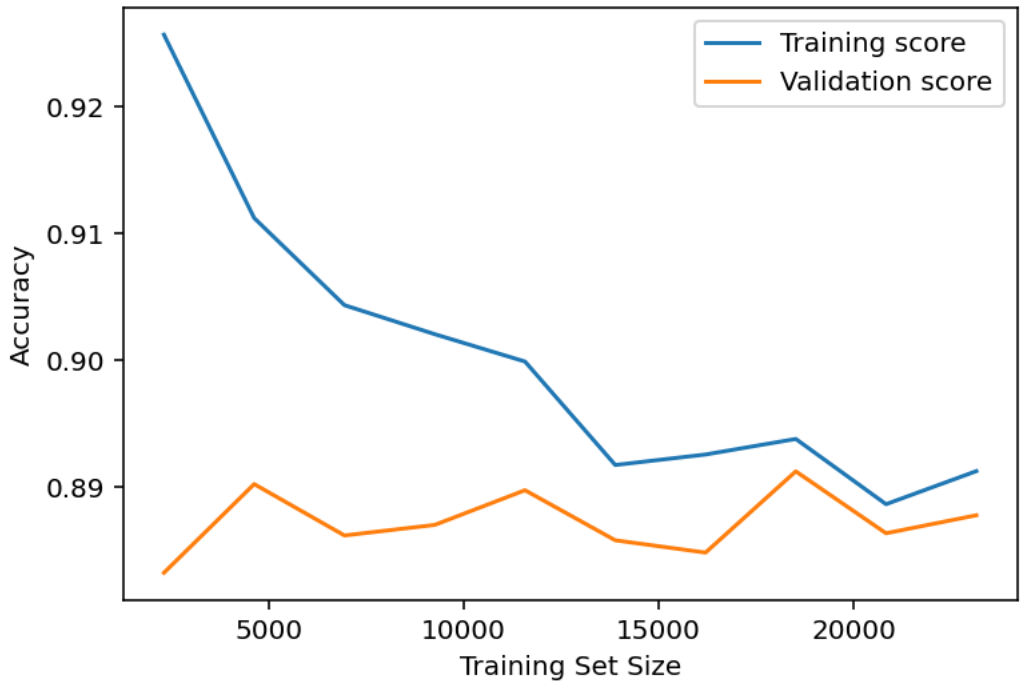


Figures 1A and 2A indicate that having moved before is the best predictor of migration. However, the difference is in directionality when analyzing both models. With a Random Forest algorithm, the feature importance in Figure 2A only illustrates the magnitude of the importance of each variable in the model. With a Logistic Regression algorithm, the feature importance in Figure 1A illustrates directionality in terms of predicting migration. Looking at Figure 1A, having had moved significantly reduces the odds of migrating because the feature importance of that variable is negative. What is interesting is that being satisfied with current friendships predicts migration in the positive direction. Figure 1A indicates that the 5 most important features to predict migration

is having had moved, being Latino, owning a residence, income, and satisfaction with friends. Figure 2A indicates that the 5 most important features to predict migration is having had moved, income, age, years of education, and owning your residence.

Machine learning models are prone to overfitting. In order to test if these models learn the training data too well by capturing noise and random fluctuations rather than the underlying pattern, I used the learning curve for each model. A learning curve is a graph that shows how a model's performance (e.g., accuracy) improves as the size of the training dataset increases. The x-axis represents the number of training samples, while the y-axis shows the performance on either the training set or a validation/test set. Typically, there are two curves: one for training accuracy and another for validation accuracy.

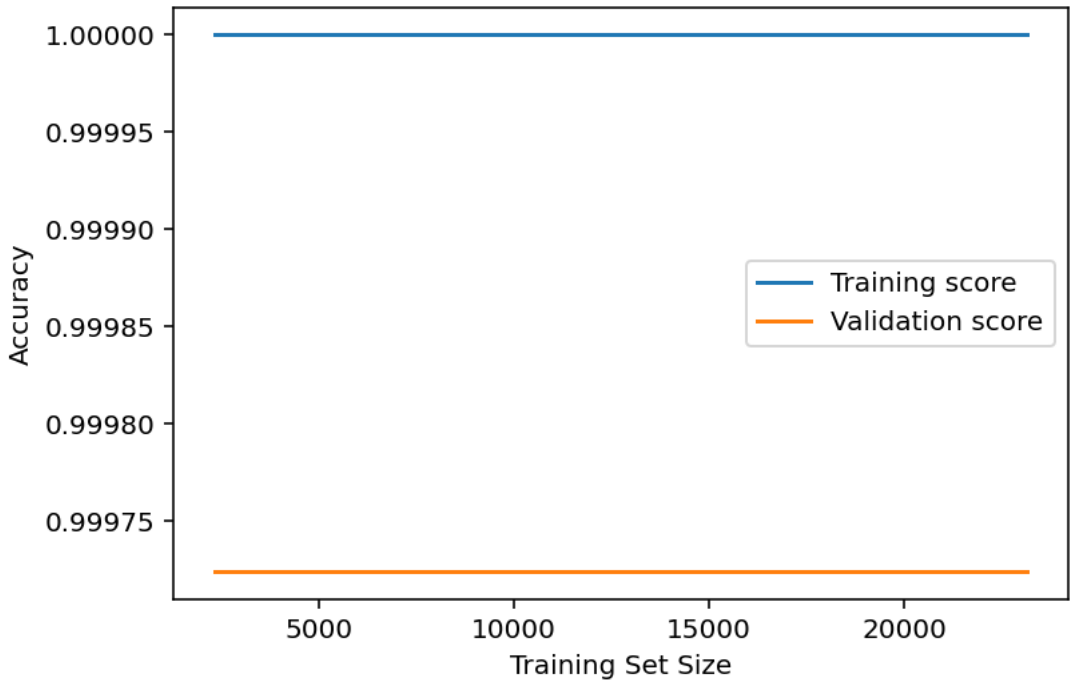
Figure 3A Learning Curve for Logistic Regression



If both training and validation accuracy are low, it suggests the model is underfitting and may need a more complex approach. If training accuracy is high but validation accuracy is much lower, the model is likely overfitting and failing to generalize well to new data. A good model is

reflected in converging curves, where both training and validation accuracy improve together as more data is added, showing that the model is learning to generalize effectively.

Figure 4A Learning Curve for Random Forest



Figures 3A and 4A indicate that the models are not overfitted. Figure 3A shows the training score starts out higher than the validation score, but both are around 90% accuracy and converge as more training data is used. This is a clear indication that the model is not overfitted. Moreover, Figure 4A indicates that both training and validation scores started and remained high, nearing 100% accuracy as the training set size increased. To further determine which model is better for predicting and understanding internal migration in the US I used a classification report to examine the accuracy, precision, recall, f1-score and macro and weighted average for each score of the Logistic Regression Model and Random Forest Model.

Accuracy is used to evaluate the performance of a classification model. It measures the proportion of correct predictions made by the model, including both true positives (correctly predicted positive cases) and true negatives (correctly predicted negative cases), out of all predictions made. Accuracy gives a straightforward measure of how often the model makes correct predictions overall. However, it can be misleading in certain situations, especially when dealing with imbalanced datasets. Datasets where classes (the number of people who moved or did not move) are imbalanced – not 50/50 split. In cases where one class is much more frequent than another, a model that simply predicts the majority class most of the time can still achieve high accuracy, even if it performs poorly on the minority class. As a result, it's important to use additional metrics like precision, recall, and F1-score to gain a more complete understanding of a model's performance, particularly when dealing with class imbalances.

Precision measures the accuracy of positive predictions. That is, it is calculated as the proportion of true positive predictions (correctly predicted positives) out of all instances predicted as positive, which includes both true positives and false positives. High precision indicates that the model does not frequently label negative instances as positive.

Recall, also referred to as sensitivity or true positive rate, measures the proportion of true positive predictions out of all actual positive instances in the dataset. It helps evaluate how well the model identifies positive instances and avoids false negatives. A high recall value means the model does not miss many positive cases.

The F1-score is the harmonic mean of precision and recall. It combines both precision and recall into a single metric, providing a balance between them. The F1-score is especially useful when there is a need to balance the trade-off between precision and recall. In cases where one metric is prioritized over the other (e.g., precision over recall or vice versa), the F1-score gives a

more comprehensive evaluation by considering both aspects. A high F1-score indicates that the model performs well in both precision and recall.

Support refers to the number of true instances for each class in the dataset. It tells you how many samples belong to each class and helps contextualize the precision, recall, and F1-score for that class. A higher support means more data is available for that class, making the metrics for that class more reliable. It's important to consider support when comparing performance across classes, especially in imbalanced datasets where some classes might have fewer instances.

When aggregating performance across multiple classes in a multiclass classification task, macro averages provide an overall evaluation. These averages calculate the mean of precision, recall, and F1-score for each class, treating each class equally regardless of the number of instances in each class. This is useful in situations where the goal is to ensure that the model performs well across all classes, without being biased by the class distribution. Macro averages are especially helpful for imbalanced datasets, as they highlight how the model handles smaller classes without being overshadowed by larger ones.

Table 1A Classification Report for Logistic Regression

	Precision	**Recall**	**F1-score**	**Support**
0	1	0.87	0.93	4460
1	0.69	1	0.82	1328
Accuracy	0.9	0.9	0.9	0.9
Macro Avg	0.85	0.93	0.87	5788
Weighted Avg	0.93	0.9	0.9	5788

Table 2A Classification Report for Random Forest

	Precision	Recall	F1-score	Support
0	0.95	0.9	0.93	4460
1	0.72	0.84	0.77	1328
Accuracy	0.89	0.89	0.89	0.89
Macro Avg	0.83	0.87	0.85	5788
Weighted Avg	0.9	0.89	0.89	5788

Tables 1A and 2A are the Classification Reports for the Logistic Regression and Random Forest models respectively. The tables indicate that both models are generally better in terms of precision, recall, and F1-score at predicting people who do not migrate, achieving scores of near 90% or higher. The Logistic Regression model seems to perform better at balancing the trade off between precision and recall when predicting internal US migration, achieving a higher F-1 score of .82 compared to the Random Forest F-1 score of .77. Moreover, the Logistic Regression model achieved a higher accuracy score as well but only by 1%. Where the Logistic Regression model dominates the Random Forest in this study is in the recall. That is, the Logistic Regression model perfectly identifies instances of migration and avoids false negatives 100% of the time where the Random Forest model does this 84% of the time. Thus, for the purposes of this study, the Logistic Regression model is superior to the Random Forest model when predicting US internal Migration in its ability to identify all the people that moved. This implies that it is more likely to be capturing which features are most relevant to predicting migration than the Random Forest model. Python and R code for this dissertation can be found at https://github.com/robersonbeauvile/Beauvile-Dissertation.