

Examining the North-to-South Migration of Black and White people in the United States in the 1970s

Qiong Wu*

Karyne Brown†

December 19, 2023

This seems weird in the title since you have data from a long time period

Abstract

This paper explores the reasons behind the migration of three groups of people from the American North to the South in the 1970s: Black people, White people, and a "total" group that consists of everyone. Using census and survey data from 1950 to 2020, we attempt to determine a relationship between the median total personal income of a group of people and the percentage of that group of people who move from the North to the South. We find that the strongest correlation between median total personal income and migration rate exists among Black people, although further, more detailed analyses of the other factors underlying this migration pattern are needed to tell a more complete story.

1 Introduction

Migration is one phenomenon that has existed since people have existed. Studying patterns and motivations behind migration can be a rewarding process that reveals underlying economic occurrences and explains human behavior. Migration is an essential part of history; researching and understanding migration can help us to learn from the past. In the United States, specifically, migration has long been a widely discussed topic. Though much of these discussions center on migration between the US and other countries, it is important to take notice of migrations that have taken place within the country as well.

In this paper, we attempt to examine the migration of large numbers of White and Black people from the North to the South in the 1970s, a pattern that is particularly interesting given that the larger and much more well-known migration trend only a few decades earlier saw people moving in the complete opposite direction, from the South to the North. This larger, earlier trend, also known as the Great Migration, was a notable occurrence in Black American history; it is not at first apparent why significant numbers of Black people would move back to the South so shortly after.

*Syracuse University. Email: qwu102@syr.edu.

†Syracuse University. Email: kbrown65@syr.edu.

great way to set up the question! This is a perfect motivation!

Many researchers have attempted to explore the reasons behind this phenomenon, citing economic and social factors that may have made a contribution. We focus on one particular factor suggested by Hunt et al. (2008), suggesting that income may have played a role in the increasing rate of Black and White migration from the North to the South. Our data work attempts to determine whether or not there is a relationship between median income and migration rate for Black and White people, and we found that the relationship among median income and migration rate is stronger among Black people than among White people.

In this paper, we attempt to situate ourselves within the body of research that exists for this topic. We briefly introduce existing papers on the subject, discuss our own goals, and determine what could be done to improve our results. In our Literature Review section, we explore in more detail ideas from previous researchers who have suggested reasons for the observed North-to-South migration in the 1970s, and at its conclusion present our own hypothesis (briefly introduced above). In the Data section, we proceed to describe the data we needed to test our hypothesis, as well as the steps we took to access, extract, and work with the data. The Results section shows what we found in more detail, and the Discussion and Conclusion section acknowledges the gaps in our work and proposes some next steps a person could take if they were to build on it.

Great intro!

2 Literature Review

The Great Migration, which lasted from the early to mid-1900s (around 1910 to 1970), saw large numbers of people leaving the American South and moving to the North, in hopes of creating a better life. A significant number of these migrants were Black, which makes sense in historical context: in the early 1900s, the Jim Crow era was at its peak in the South, whereas the North was a somewhat less racially discriminatory place for people of color to live in.

In 1970s, however, this migration pattern seemed to reverse itself, as large numbers of people, a significant proportion of whom were Black, began to leave the Northern United States and move back to the South. According to Newman (1983), this is referred to as “the Sunbelt phenomenon” (pg. 76). Because it showed a complete turnaround in migration patterns from previous decades, the Sunbelt phenomenon has been studied by many researchers who have hoped to find the reasons behind this significant shift. In particular, Cromartie and Stack (1989) point out that a majority of the Black people who migrated from the North to the South as part of the Sunbelt phenomenon were return migrants (people who were born or have previously lived in an area) to the South or at least had a previous acquaintance with the region. Why would so many Black people suddenly move back into the region many of them had vacated only decades prior?

Hunt et al. (2008) suggest a general answer to this question: for Black people in the 1970s, the South simply became a better place for them to live while the North became worse in that regard. This idea is also supported by the work of other researchers.

Some of the ~~reasons and~~ factors underlying the improved economic and social prospects for Black people in the South in the 1970s included an increase in the demand for labor in the South, caused by growth in the service and manufacturing sectors (Cromartie & Stack, 1989). Proportionally more Black people in the 1970s were able to take advantage of this Southern economic boom due to “marked declines in the Black/White wage gap” (Hunt et al., 2008) and improvements in Black education (Smith & Welch, 1989). In addition, Newman (1983) claims that the state corporate income tax was lower in the South during the 1970s, creating a “more favorable business climate.”

On the other side of the country, the North was becoming a worse place for Black workers to make a living. Although the North had industrialized much sooner than the South, this perhaps became a weakness for the region in the 1970s, when globalization and deindustrialization began to take place (Hunt et al., 2008). These two forces contributed to the concentration of poverty in Northern cities (Hunt et al., 2008). In addition, issues such as redlining were also prevalent in the North during that period, leading to increased racial segregation. Eventually, according to Hunt et al. (2008), “de-facto racial segregation” became lower in the South than in other regions of the United States, such as the North. Combined, all of these factors contributed to the North becoming a worse environment for Black people to live in than the South. This decline, combined with perceptions of better economic opportunities in the South due to the reasons listed above contributed to the migration of Black people from the North to the South.

The above factors may contribute to the reason why people, especially Black people and workers, would have wanted to move to the South, but Greenwood & Hunt (1984) provide a reason for why this was possible for such a large number of people. According to Greenwood & Hunt (1984), migration is a “self-reinforcing phenomenon.” This means that as more and more people migrate, the destination (where people migrated to) becomes a more and more attractive place for even more migrants to move to. Greenwood & Hunt (1984) attribute this to the fact that migrants may influence prices and profitability, as well as “contribute to the growth of markets” and “cause increased investment” in their destinations. The idea that migration paves the way for further migration to take place may also partially explain why a larger number of Black people were able to move from the North to the South over time.

Our research examines one specific claim made by Hunt et al. (2008), that a possible cause of the Sunbelt phenomenon was the decrease in the racial wage gap between Black and White people. We decided to test the validity of this claim by examining whether or not a higher median income was related to a higher rate of North-South migration among Black people, White people, and a total group. We hypothesized that an increase in the median income of each group (i.e., Black people, White people, or total) would correspond with an overall increase in the percentage of that group that moved from the North to the South, then used [Stata](#) to test data to verify this claim.

It would be better to draw a more direct line from the wage gap to income.

The literature review is extraordinarily well done.

3 Data

Our data is drawn from IPUMS USA, a website that provides census and survey data from 1790 to the present. The data we collected from this website is anonymized and on an individual level. This means that we do not have the means to track specific people over time; each observation is only a snapshot of a person’s life in a specific year and does not tell their whole story. ~~Data downloaded from the IPUMS USA website was extracted then transported to Stata.~~

We used 8 data samples, one for each decade from 1950 to 2020: (1) the 1% sample for 1950, (2) the 5% sample for 1960, (3) the 1% metro fm1 sample for 1970, (4) the 1% metro sample for 1980, (5) the 1% metro sample for 1990, (6) the American Community Surveys (ACS) sample for 2000, (7) the ACS sample for 2010, and (8) the ACS sample for 2020.

In order to test the relationship between the median total personal income in a group of people and the tendency of members in that group to participate in North-South migration, we chose to use the following variables: race (RACE), birthplace (BPL), state or country of residence 5 years ago (MIGPLAC5), state or country of residence 1 year ago (MIGPLAC1), total personal income

It's good to give the IPUMS variable names once for people who are familiar with the data; after that, best practice is to refer to each variable by an easy-to-understand natural language name.

(INCTOT), and a consumer price index multiplier (CPI99). INCTOT displays an individual's total reported pre-tax income from any source over the year. CPI99 converts all of the dollar values across each year to 1999 dollars, therefore allowing one to directly compare dollar amounts from different years.

Awesome way to address the inflation issue!

Since our research attempts to determine a relationship between median total personal income and migration rates among Black and White people, the RACE variable was essential to our research. As stated in our Literature Review section, we focused on Black people, White people, and a "total" group.

We used the BPL, MIGPLAC5, and MIGPLAC1 variables to obtain a general idea of migration patterns in each decade. We only focused on people that our data showed migrated from the North to the South. This means that we were only interested in individuals who were born (BPL) in the North and ended up living in the South (MIGPLAC5 or MIGPLAC1). The North we defined as Maryland, Pennsylvania, New York, New Jersey, Connecticut, Rhode Island, Massachusetts, Vermont, New Hampshire, Maine, and Delaware; the South included Texas, Oklahoma, Arkansas, Louisiana, Mississippi, Alabama, Georgia, Florida, North Carolina, Tennessee, and Virginia.

Our original plan was to attempt to work with both return and nonreturn migration to the South. Return migration simply means that an individual was born in the South, moved to the North at some point in their life, then eventually migrated back to the South. Nonreturn migration describes individuals who were not born in and have not lived in the South, but who migrated to the region.

This level of analysis would require the use of all three of the variables BPL, MIGPLAC5, and MIGPLAC1. Identifying a return migrant using these variables would require an individual to be born (BPL) in the South, have lived in the North five years before a specific sample year (MIGPLAC5), then moved back to the South between one and five years before the sample year so that their place of residence one year ago (MIGPLAC1) would have been in the South. Using the same method, a nonreturn migrant would have to have been born (BPL) somewhere in the North, then lived in the South as five years or one year before the sample year (MIGPLAC5 or MIGPLAC1).

The obvious flaw in this approach, of course, was that since we could only observe individuals' place of residence in specific years, there could be many migrants who moved at different points in time that would go undetected. For example, we would not be able to identify a person who was born in the South, moved to the North eight years before a particular sample year, moved back to the South six years before the same sample year, and then stayed there. This is because their BPL, MIGPLAC5 and MIGPLAC1 would all show up as the South, indicating no migration at all.

I don't follow

The other flaw in this approach was that it was not possible to analyze at all, since only the BPL variable was available across all 8 of our samples. MIGPLAC5 was only offered in the 1970, 1980, and 1990 samples, and MIGPLAC1 was only available in the 1950, 2000, 2010, and 2020 samples. The 1960 sample does not have MIGPLAC5 or MIGPLAC1 at all.

Our data work, therefore, was confined to using only BPL to observe where a person was born and either MIGPLAC5 or MIGPLAC1 to see where they ended up. In other words, we are only able to observe migration if an individual was born in the North then lived in the South either one or five years prior to a sample year. Since MIGPLAC1 was more widely available across the samples than MIGPLAC5, we decided to use MIGPLAC1 as the "default" destination variable, and MIGPLAC5 as a "backup" when MIGPLAC1 wasn't available.

Since our goal was to determine how income affected migration rate among Black and White people, we needed to examine the median total personal income over time within each specific racial group (Black, White, and total). To accomplish this, we found that we needed to combine the INCTOT

and CPI99 variables by multiplying them together. This enabled us to look at adjusted income data without worrying about the effect of inflation.

Having thus determined the reasons for using each variable, we then needed to download the data. When the IPUMS website generates a data extract, it provides a data file and several options for command files. For each of the eight extracts we generated (one per sample year), we downloaded a data (.dat) file and a Stata command file (.do.txt).

The .dat files downloaded as compressed .zip files. To access them, we needed to use the 7-Zip application to extract the files. After the files were extracted, we renamed each of them by their sample year (e.g., “1950” for the 1950 sample) and saved them to a shared OneDrive folder (linked in our GitHub repository).

The .do.txt files could be opened directly in Stata, but the code they contained needed to be modified before they could be used to access the data. First, we added the “cd” command to the top of each do-file to change the working directory to ours. (We also added a blank line after the “cd” command for visual clarity, which means that this action added a total of two lines to the beginning of our do-file.) Then, we changed the name of .dat file in what was now line 26 of the code to correspond to the new names we gave to the .dat files. We then ran the entire do-file, which allowed us to view and begin working with the data.

We first ensured that the dollar amounts across all 8 sample years were equivalent and could be compared directly, by using the CPI99 variable:

```
gen realinc=inctot*cpi99
```

This line of code generates a new variable, “realinc,” which is the real value of the incomes, in 1999 dollars. This allows us to compare medians from all samples without needing to take any extra steps. We then eliminated the outliers from our income data. Since Stata is unable to display numbers greater than 9,999,999, we decided that all of those numbers would only provide an inaccurate picture if used in our calculations. (For example, 2,000,000,000 would have the same value as 10,000,001 since they are both greater than the maximum of 9,999,999.) We used the following code:

```
keep if realinc < 9999999
```

In order to find the individual income data, we first needed to sort by race.

I think dropping these worsens the measurement problem--now those people don't even exist in your data rather than just having their incomes reported as lower than they actually are.

We then used the “summarize” function to find the median incomes for each racial group (total, Black people, and White people):

```
sum realinc, detail
```

```
by race: sum
```

We identified the 50% percentile as the median value, which we felt was a more accurate portrayal of average than the mean in this case.

Having thus collected the income data, we needed to collect the migration rates. We wanted to separate out the people who migrated from the North to the South, so that we could find the migration rate for each population (Black, White, and total). At the end of our do-file, we added a few lines of code to generate dummy variables that would help us to achieve this.

The first dummy variable generated was called “north,” and was used to identify those individuals who had been born (BPL) in the North:

```
gen north=0
```

The next step was to make sure that the “north” code included each state that we identified as being part of the North. To achieve this, we used each state’s Federal Information Processing Standards

(FIPS) code:

```
replace north=1 if bpl==09
```

The FIPS code 09 corresponds to the State of Connecticut. We then repeated this line of code for each of the rest of the states in the North: Delaware (10), Maine (23), Maryland (24), Massachusetts (25), New Hampshire (33), New Jersey (34), New York (36), Pennsylvania (42), Rhode Island (44), and Vermont (50).

The second dummy variable generated was called “south,” and was used to identify those individuals who lived in the South right before the sample year (MIGPLAC1):

```
gen south=0
```

We then made sure that the “south” code included each state that we identified as being part of the South, using each state’s FIPS code:

```
replace south=1 if migplac1==01
```

The FIPS code 01 corresponds to the State of Alabama. We then repeated this line of code for each of the rest of the states in the South: Arkansas (05), Florida (12), Georgia (13), Louisiana (22), Mississippi (28), North Carolina (37), Oklahoma (40), Tennessee (47), Texas (48), and Virginia (51).

For samples that did not have the MIGPLAC1 variable available, we used MIGPLAC5 instead:

```
replace south=1 if migplac5==01
```

and so on and so forth.

Finally, we combined the “north” and “south” variables into one bigger variable, “migsouth,” that included everyone who was born in the North then lived in the South.

```
gen migsouth=0
```

```
replace migsouth=1 if north==1 & south==1
```

We then used the “tabulate” function to find the migration rates for each of the three racial groups we focused on (Black people, White people, and total):

```
tab migsouth
```

```
by race: tab migsouth
```

We used the percentage of “1s” as the percentage of each group that migrated from the North to the South. For example, if 0.16 percent of the observations for “migsouth” are 1s for Black people in the 1980 sample, then we said that 0.16 percent of Black people in 1980 had moved from the North to the South.

Having thus modified the do-files, we saved each of them to the same OneDrive folder that contains the data files and renamed them to match the names of the data files. In summary, each sample year has a .dat file and a .do file named after it (e.g., the sample year 1950 has a corresponding data file named “1950.dat” and a corresponding do-file named “1950.do.txt”).

4 Results

We hypothesized that a higher median income would lead to a higher rate of North-to-South migration, and that this would be true among three groups of people: Black people, White people, and a total group which includes every race. The reasoning was that a higher income gave people the financial ability to migrate in search of better economic opportunities. (A higher median income generally indicates that individuals within that population have higher incomes across the board.) In the 1970s especially, the South was economically a much better place to work than the North, due to a variety of reasons covered in our Literature Review section. As a result, our data work focused on attempting

to determine whether or not there was a relationship between the median income of Black and White people in the North and their inclination to migrate to the South. Our expectation was to find that a higher median income in each group would correspond to a higher rate of migration for that group of people.

Through our data work (described in our Data section), we found the median income and North-to-South migration rates for Black people, White people, and everyone. Our results are shown in the tables below:

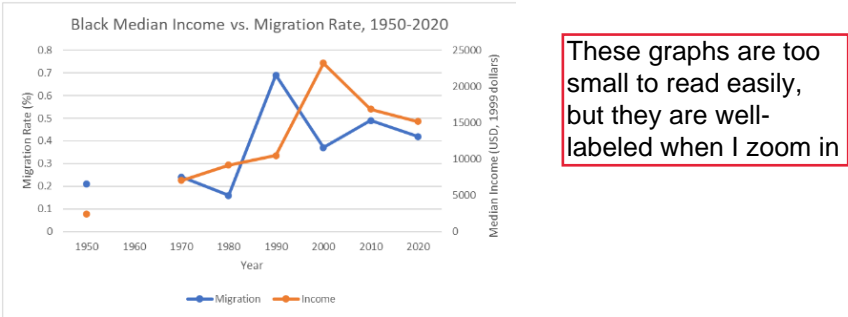


Figure 1: Black Median Income vs. Migration Rate, 1950-2020

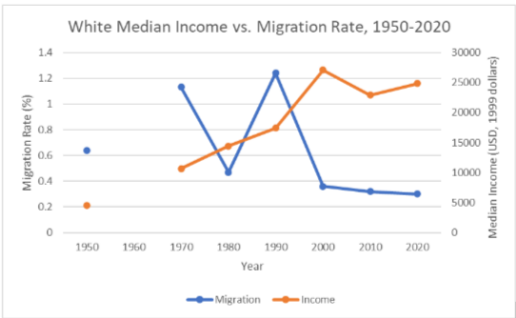


Figure 2: White Median Income vs. Migration Rate, 1950-2020

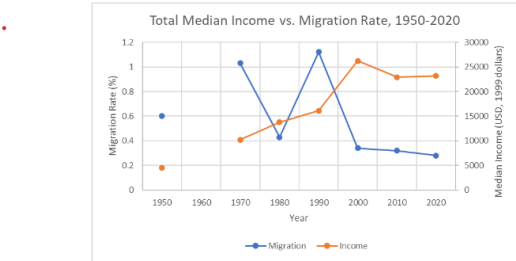


Figure 3: Total Median Income vs. Migration Rate, 1950-2020

We display our data in three different line graphs: one for the total population, one for just the White population, and one for the Black population. The left vertical axis shows the percentage of

each population that migrated from the North to the South, and the right vertical axis shows the median income for that population. The horizontal axis represents time and includes each of the sample years ~~we worked with throughout our research.~~

Our results showed that overall, there appears to be no direct relationship between the median income of a population and their tendency to move from the North to the South in search of better economic opportunity. As shown from each of the graphs above, the two lines that show migration rate and median income do not seem to display the same shape or pattern over time. This is especially true for the graphs corresponding to the total population and to only White people. This is mostly due to the drop in migration rates in 1980 and the higher-than-expected median income in 2000.

In Figure 1, however, which corresponds to only Black people, there appears to be a faintly recognizable pattern between the two lines that display migration rate and median income. From 1980 onward, it seems as though every time there is a change (i.e., a large increase or decrease) in the line for migration rate, the same change could be seen in the line for median income only a decade later. For example, from 1980 to 1990, the migration rate increased drastically, from less than 0.2% to nearly 0.7%. From 1990 to 2000, we see a drastic increase in the median income as well, from about \$10,000 to about \$22,000. There is then a significant, though not as drastic, decrease in the migration rate from 1990 to 2000, from nearly 0.7% to under 0.4%. A similar pattern can be observed in the median income from 1990 to 2000, as the median income falls from about \$22,000 to around \$16,000.

A possible implication of all these data is that median income is a more crucial factor in catalyzing the migration of Black people from the North to the South than for White people. A possible explanation for this is that it has historically been easier for White people than for Black people to gain financial freedom in any sense, and especially to accumulate enough wealth to migrate to a completely different region of the country.

Another possible implication is that migration may cause increases in median income, and not the other way around. If median income follows the same patterns as migration rates, but only a decade later, this may imply that people who were seeking better economic opportunities in the latter half of the 1900s through North-to-South migration were largely successful.

Though our data may not be very detailed, a few implications may actually arise from the incompleteness of our analysis. One possible explanation for the lack of a clear relationship between the two variables we examined is that migration is not driven purely by financial factors. After the majority of North-South migration took place in the 1970s, the number of people (in any given group or population) who wanted to migrate from the North to the South may have declined. Therefore, even if the median income for that group of people increased after the 1970s, they may still have chosen not to migrate. This would show up in the graphs as points with lower migration rates and higher incomes. (This is especially applicable to the White and total populations.) We may conclude from this analysis that a combination of factors catalyzed North-to-South migration in the 1970s, after which migration decreased, perhaps because the majority of people who intended to migrate already did so during that decade, and changes in any migration factors in the following decades will not display a strong correlation.

I would have liked to see correlation coefficients for each of the three groups.

The IPUMS data is very detailed; you just didn't choose to exploit most of the detail.

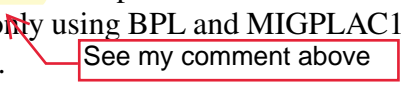
I don't understand. The migration rates is high in 1990 as well.

5 Discussion

Giving further thought to our data work process reveals that there are a few possible things that would make our research a lot more accurate and detailed.

The first step is to find a different, better source of data. The flaws with the IPUMS data that we used were that it was anonymized, did not have enough of the variables that we needed, was limited in its ability to display numbers higher than 9,999,999, and that the sampling was not consistent over time.

The major issue with anonymized data, as discussed in our Data section, is that it is impossible to track the migration of individual people over time. This causes our migration data to be inaccurate as we can't tell when any specific person migrated, which may lead to double-counting. For example, a person (let's call them Max) born in the North in 1920 who moved to the South in 1945 would have been included in our calculation for the migration rate in 1970 and in 1980, provided that their data was included in both the samples for 1970 and 1980. This is because, in 1970 and in 1980, Max's destination (MIGPLAC5 or MIGPLAC1) would have shown up in the "south" dummy variable. Even though Max only migrated once in their life, their migration would have been counted in both the 1970 migration rate and the 1980 migration rate.

The variables themselves were also unable to show a complete picture of the data we needed. To continue with the example given in the previous paragraph, we would have no way of telling when Max migrated just by using the variables BPL, MIGPLAC1, and MIGPLAC5. A different variable that showed exactly when a person migrated from the North to the South could be helpful in this regard. Another helpful variable would have been "current residence." It is possible that there are people who migrated one year before the sample year, in which case only using BPL and MIGPLAC1 (or MIGPLAC5) would not have been able to capture their migration. 

As previously discussed in the Data section, ~~Stata does not display values higher than 9,999,999,~~ instead treating them all as the same value. This results in inaccuracy in any average calculations of the data because, as previously stated, this system would view 10 million and 10 billion as the same value. On the other hand, removing all of these values, as we did in our research, is also not an optimal solution. One major reason is that it deletes many observations of our data (which is already a sample of the entire US population), and therefore results in an inaccurate migration rate. The other major reason is the more obvious concern that removing observations from a set of data would result in an inaccurate portrayal of the whole data set.

A fourth problem with the data from IPUMS is that the eight samples we collected (one for each year) are not consistent over time. For example, we used the 1% population sample for 1950, the 1% metro fm1 sample for 1970, and the ACS sample for 2000. Although the 1950 and 1970 samples both use data from 1% of the American population, the issue is that the 1950 sample uses 1% data from everyone, whereas the 1970 sample uses 1% data only from people who live in metropolitan areas. Comparing these two samples to the 2000 reveals even more inconsistency as the sample we used for the 2000 data is from an American Community Survey, which is not the same as the national census. Across different years, data collection and sampling methods are varied. In order to have the most accurate portrayal of the data possible, we would have needed to work with the entire US population. Since that is unrealistic, consistent sampling and data collection methods would at least have been preferable to what we did have.

Other issues with our data included that we had no means of telling whether migrants consisted mostly of individual working adults, or if working migrants brought dependents along with them (i.e.,

their children and non-working members of their families).

Further exploration of our research topic would first require obtaining a data set that fixes all of the above issues. Other steps include determining and examining other factors that affect North-to-South migration in the 1970s, as well as examining and comparing migration to and from other regions of the United States in other time periods. A comprehensive portrait of the causes and effects of North-to-South migration in the 1970s would include much more detailed analysis than we were able to accomplish.

6 Conclusion

Through our analysis of the median income and migration rates of Black people, White people, and the total American population over time, we found that there is no easily observable relationship between the two variables among White people and among the total population, but that when only Black people are considered, the two variables may be related. Unfortunately, our data and tools for analysis were very limited, and it is unclear which of the two variables influences the other, if such a relationship exists at all. We believe our work contributes to the existing body of research on North-to-South migration in the 1970s as it offers a possible topic for further research: the economic success of those individuals who participated in this migration, especially Black people. Our results may not have fully answered the question we posed, but they did reveal an interesting topic that may call for further, and more detailed, exploration.

Bibliography

Cromartie, J., & Stack, C. B. (1989). Reinterpretation of Black return and nonreturn migration to the South 1975-1980. *Geographical Review*, 79(3), 297-310. <https://www.jstor.org/stable/215574>

Greenwood, M. J., & Hunt, G. L. (1984). Migration and interregional employment redistribution in the United States. *The American Economic Review*, 74(5), 957-969. <https://www.jstor.com/stable/555>

Hunt, L. L., Hunt, M. O., & Falk, W. W. (2008). Who is headed South? U.S. migration trends in Black and White, 1970-2000. *Social Forces*, 87(1), 119-195. <https://api.semanticscholar.org/CorpusID:143467060>

Newman, R. J. (1983). Industry migration and growth in the South. *The Review of Economics and Statistics*, 65(1), 76-86. <https://www.jstor.org/stable/1924411>

Smith, J. P., & Welch, F. R. (1989). Black economic progress after Myrdal. *Journal of Economic Literature*, 27(2), 519-564. <https://www.jstor.org/stable/2726688>