THE GREAT MIGRATION AND SOCIAL INEQUALITY: A MONTE CARLO MARKOV CHAIN MODEL OF THE EFFECTS OF THE WAGE GAP IN NEW YORK CITY, CHICAGO, PHILADELPHIA AND DETROIT

Débora Mroczek
University of Houston
Honors U.S. History – Dr. Guenther
Spring 2017

I – Introduction:

The Great Migration was the most important event in the 1900's – it changed America's society and economy irreversibly and defined a dynamic that continues to draw racial lines today. Over those multiple decades, discrimination set a wage gap that determined how cities and states across the country developed. Some of the Northern locations that experienced a huge flux of African American migrants are now the most racially segregated in nation – a result of the *de facto* laws that dominated cities such as Detroit, New York City, Chicago, and Philadelphia.

A recent study by the Seven Pillars Institute (SPI) for Global Finance and Ethics concluded that there are four principal causes for social inequality: wage (which is influenced by a variety of factors, but mostly by supply and demand and skill level), education level, gender, and personal factors[1] (which accounts for any form discrimination.) Variables such as gender discrimination and personal factors are difficult to measure but wages and education level can be easily sized using statistical data. Therefore, since the U.S. Census Bureau has been collecting data since the 1800's, it is possible to take a close look into the cities affected by the multiple waves of the Great Migration and establish a connection between the racial inequalities that developed then and the social gap that exists today.

In this study, I use a Monte Carlo Markov Chain model to demonstrate that social inequality and wealth concentration are deeply rooted in the

1. May Leung, "The Causes of Economic Inequality," *Seven Pillars Institute (SPI) for Global Finance and Ethics*, 01/21/2015 http://sevenpillarsinstitute.org/case-studies/causes-economic-inequality

American socioeconomic system. The final results show that a society with equal work opportunities would still experience the effects of wealth concentration, producing statistics similar to the ones recorded recently by local and state governments in Detroit, New York City, Chicago, and Philadelphia.

II - Methods:

Mathematically, the model is simple – it incorporates each of the factors influencing social inequality in different ways and fixes the supply and demand factor to simulate a world in which everyone has a position in the job market. More importantly, it simulates a world in which everyone has equal chances of getting a position in the job market. That is, if a white person and an African American are competing for the same position, the final decision is random. In order to simulate such conditions, every factor must be modeled properly.

A 100x2 matrix D represents the population, and the entries are determined based on statistics drawn directly from the United States Census Bureau. The number "2" comes from the two independent variables in the model: race and level of education. The number "100" comes from the fact that looking at statistics in percentages makes data analysis a lot simpler (i.e. if 52% of the population is African American, there are 52 correspondent entries in the matrix.) Otherwise, the model would be analyzing thousands of entries and the matrix would have different sizes for different cities, which would result in an extremely inefficient program. On the other hand, with only 100 data points to choose from, the program runs efficiently.

I previously mentioned four factors that contribute to social inequality, but the model only accounts for two (race and level of education), what about gender and personal factors? Firstly, the main purpose of this model is to show that discrimination is rooted in the economic system. Therefore, it would be inconsistent to make gender or any personal factors independent variables in the calculations. Secondly, this simulation is a Monte Carlo Markov Chain model, meaning that the flow mechanisms are well defined by equations and known distributions. However, our understanding of how gender and personal factors affect the wage gap is very poor and could only possibly be modeled by a Hidden Markov Chain model. Given the purposes of this paper, in depth definitions of Monte Carlo Markov Chain (MCMC) and Hidden Markov models (HMM) are not necessary, but HMM's often require supercomputers and are extremely complex to analyze. Despite the fact that MCMC's exclude variables historians may consider critical, it is possible to obtain with a MCMC the same results a HMM would produce by choosing the independent variables carefully. Lastly, gender and personal factors influence the statistical data that is used to calculate wage gap. Therefore, since it is impossible to detach those variables from the wage gap, even if I attempted to make them independent variables in the study, the final distribution would consider the same factors twice and results would be skewed.

The data input has the following format:

$$D_{100x2} = [\ R_{i1} \quad E_{i2}\ ]$$

R – Race. 1 = White / 2 = African American
E – Level of Education. 1 = No education/Some education/High school but no college / 2 = College degree

It is important to notice that since the simulation uses qualitative information to generate quantitative results, all input must be coded in the form of "dummy" variables. This is a commonly used method in statistical analysis and it consists of assigning different numbers to the different characteristics that define the input. Therefore, every row is a combination of two numbers: 1 or 2 for race and 1 or 2 for education level (e.g. D_{ij} = [1 2] represents a white person with a college degree.)

Once the distribution is determined, the program randomly selects an input from the matrix D and runs it through the simulation once. It is imperative that the point is randomly selected since the model is based on the principle of equality. In other words, everyone has an equal chance of getting a job. Once the input is determined, the program "flips a coin" to decide which type of job the person, represented by the data point, is being offered; it is either a position that requires a college degree (dummy variable = 1) or a position that does not require a college degree (dummy variable = 0.) That is, if the point selected is D_{ij} = [1 1], a white person with no college degree, and the number randomly generated is 1, a position that requires a college degree, then that person would not get the job and the point would be returned to the matrix D to be randomly selected in future loops. An important point is that the first randomization of the simulation (the selection of the input) allows for repetition if the point is rejected at the second random step (job selection). That ensures that every input will be assigned to a position but it is impossible for more than one position to be assigned to the same input. Such event would be equivalent to a person working multiple jobs, which is unrealistic and would skew the results.

Assuming the input matches the requirements for the job, the program makes a "payment." Payments are numbers between 0 and 1 that represent proportional earnings. For this study, I consider the amount of money made by a white to be 1. Statistical data determines what proportion of 1 an African American would get for the same job. The program then adds all the payments made to whites and all the payments made to African Americans in separate columns of a final data matrix. The results are also broken down by education level. Since we are only interested in the absolute wage gap, all the information can be handled and analyzed in the form of proportions.

The final product of the simulation is a matrix containing the final values outputted by the model, which will invariably converge. Not all MCMC simulations converge but since this model is fairly simple and only flows in one direction, the simulation is finite. From those values, it is possible to look at the proportional income in four different categories: white with college degree, white with no college degree, African American with college degree, and African American with no college degree. The incomes can then be compared to more recent statistics from each city.

The full process is illustrated in figure 1.

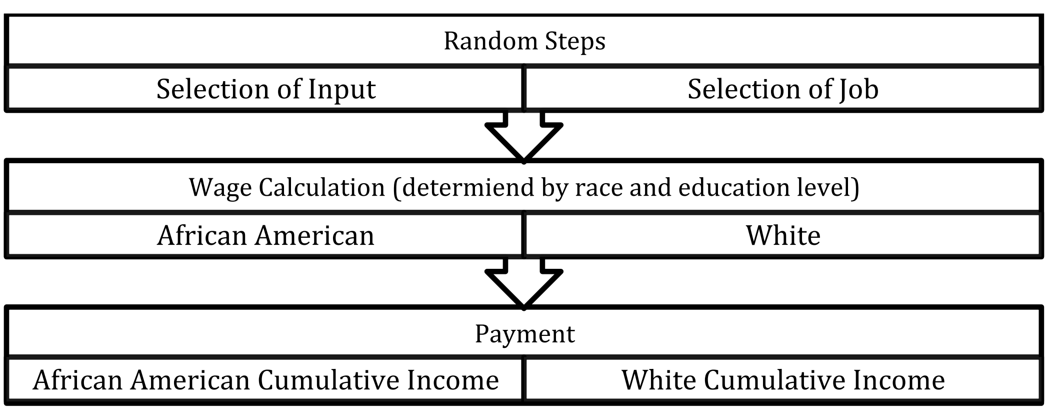


Figure 1 Illustration of MCMC simulation by steps

III – Data Selection:

The four cities selected for this study are among the locations that experienced the greatest increase in African American population between 1900 and 1990. All statistical data used was collected and analyzed by the United States Census Bureau and the United States Bureau of Labor Statistics.

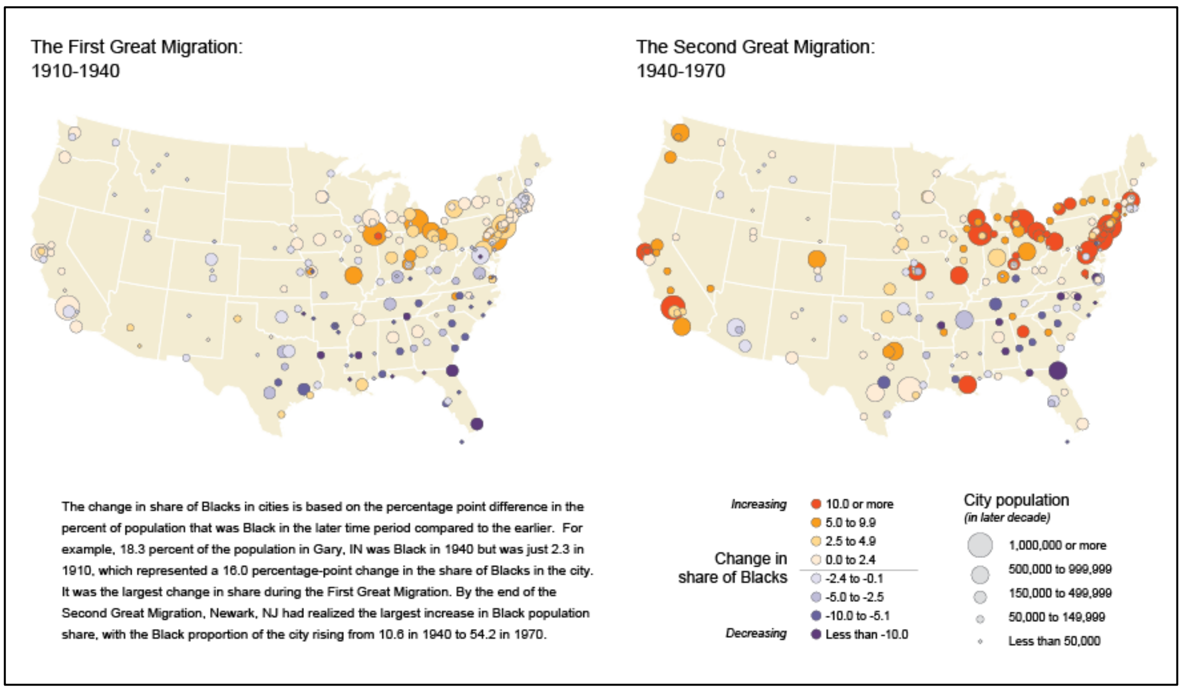


Figure 2 The Great Migration 1910-1970
Source: US Census Bureau, Data Visualization Gallery, https://www.census.gov/dataviz/visualizations/020/, February 2013.

As figure 2 illustrates, the number of African Americans moving away from the South increased sharply between 1910 and 1970. However, most of the migration took place during what is known as the Second Great Migration, between 1940 and 1970. In order to better reflect the wage gap that developed over the Great Migration, I used data from the year 1970. Tables 1 and 2 summarize the data used to model the demographics.

Table 1 Summary of Local Statistics

City	**Detroit**	**New York City**	**Chicago**	**Philadelphia**
African American Population in 1970 (%)[2]	43.7	21.1	32.7	33.6

Table 2 Summary of National Statistics

African American population with a college degree in 1970 (%)[3]	31.4
White population with a college degree in 1970 (%)[4]	54.5
Income ratio: African American to White in 1970[5]	.6296
Wage ratio: Machinery/Construction sectors (do not require a college degree) to Health/Education/Public Administration (require a college degree) in 1970[6]	.60003

2. Campbell Gibson "Historical Census Statistics On Population Totals By Race, 1790 to 1990, and By Hispanic Origin, 1970 to 1990, For Large Cities And Other Urban Places In The United States," *United States Census Bureau*, February 2005, https://www.census.gov/population/www/documentation/twps0076/twps0076.html

3. United States Census Bureau, "United States Statistical Abstract: 1999. Educational Attainment, by Race and Hispanic Origin: 1960 to 1998," Education, 10, 11/1999, https://www.census.gov/prod/99pubs/99statab/sec04.pdf.

4. Ibid.,10
5. Ibid., 6.
6. Ibid., 6.

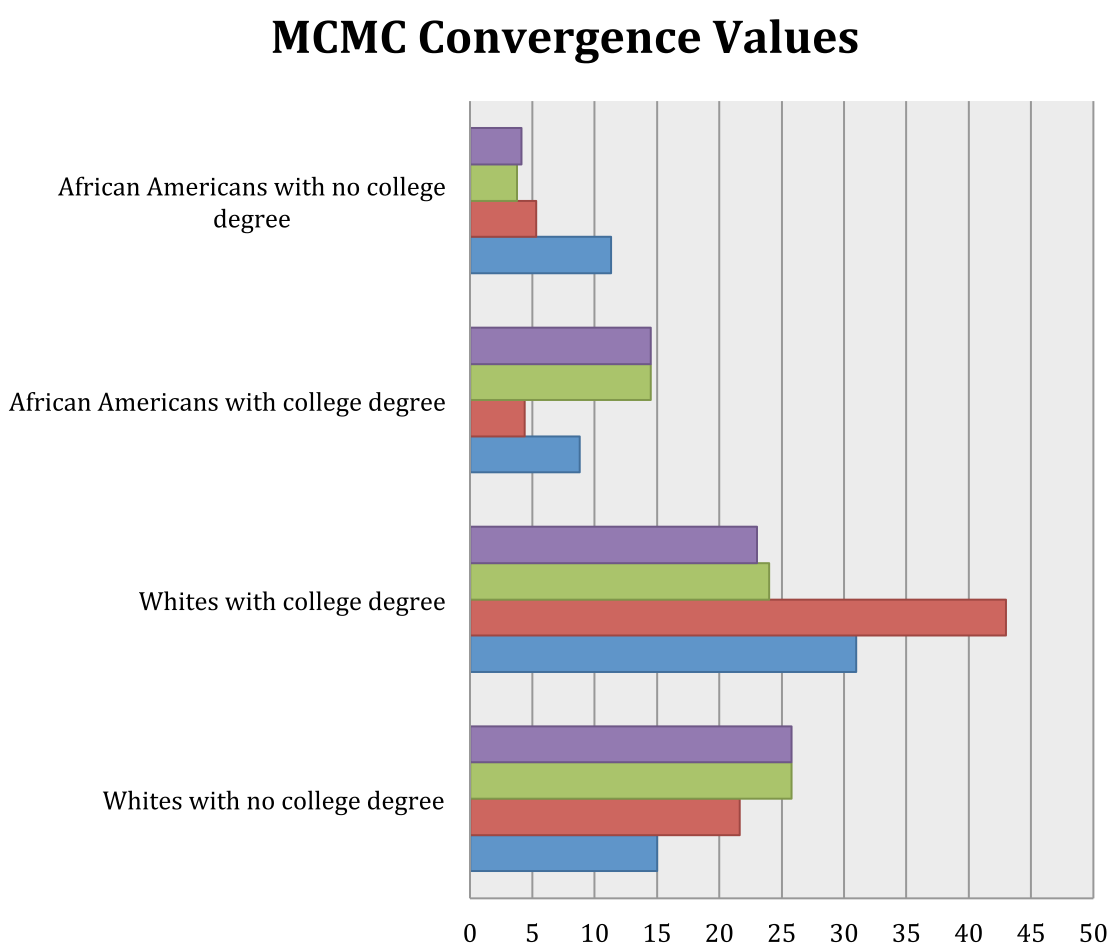
IV – Results:

As expected, due to the wage gap, there was a significant difference between the wealth accumulated by whites and the wealth accumulated by African Americans. Figure 3 details the convergence values yielded by the simulation. From those values it was possible to calculate the average income with respect to population by race using Equation 1:

$$I = \frac{C_1 + C_2}{P} \tag{1}$$

I: Average income with respect to population by race
C_1: Convergence value 1 (college degree)
C_2: Convergence value 2 (no college degree)
P: Percentage of population

Dividing the average income of whites by the average income of African Americans yields the income ratio, which allowed me to compare the income gap generated by the model to the actual income gap in all four cities. Not surprisingly, even when there exists equality of opportunity, the income gap is still significant. In fact, the percent differences indicate only a small disagreement between simulation outputs and recent statistical data, considering the simplicity of the model and the limited computational capability under which the simulation was conducted. Table 3 summarizes the results for income ratio.



	Whites with no college degree	Whites with college degree	African Americans with college degree	African Americans with no college degree
Philadelphia	25.8129	23	14.4808	4.1536
Chicago	25.8129	24	14.4808	3.776
NY City	21.6108	43	4.4072	5.2864
Detroit	15.0075	31	8.8144	11.328

Figure 3 Final Output

Table 3 Summary of income ratio

City	**MCMC Income Ratio: White to African American**	**2000's Income Ratio: White to African American**	**Percent Difference**
Detroit	1.794	1.331[7]	+18.9%
New York City	1.772	1.812[8]	-2.2%
Chicago	1.349	1.506[9]	-11.0%
Philadelphia	1.360	1.752[10]	+25.2%

7. New Detroit Coalition, "Metropolitan Detroit Race Equity Report," 03/2014, http://www.newdetroit.org/docs/press/MetropolitanDetroit_RaceEquity_Report_NewDetroit_V1.1.pdf
8. State of New York Department of Health, "New York City Health Indicators by Race/Ethnicity, 2012-2014," 08/2016, https://www.health.ny.gov/statistics/community/minority/county/newyorkcity.htm.
9. U.S. Census 2010, "Chicago Median Income by Race," 11/2010, http://www.usacityfacts.com/il/cook/chicago/economy/.
10. City Data, "Philadelphia, Pennsylvania (PA) income map, earnings map, and wages," 04/2016 http://www.city-data.com/income/income-Philadelphia-Pennsylvania.html#ixzz4fB99IkZx.

V – Conclusions:

The purpose of this study was to establish a correlation between racial inequality and economic inequality. A Monte Carlo Markov Chain model was used to simulate conditions under which personal discrimination and inequality of opportunity do not exist. As shown by the final results, racial discrimination is so profoundly rooted in how wealth flows in our economy that even under simulated equality we still observe massive income gaps develop.

The numbers in this study help further prove that America is far from being post-racial. In fact, using wage gap only, the model was able to recreate with high accuracy data recorded very recently. The small percent differences between the values produced by the MCMC simulation and recent statistical data highlight the correlation between a segregated society that began to develop in the 1900's and the social gap that continues to grow today.

Racial inequality must be addressed from multiple angles, attacking only one or two of its aspects makes no difference – neither in computer simulations nor in the real world. Unfortunately, race defines how we experience society and the economy, and it benefits ones more than it does others. The truth is, if America wants to call itself "post-racial," it has a long way to go. From education, to wage equality, to even opportunity across ethnicities, a lot needs to change – there is no such thing as equality when race regulates every aspect of social experience.

Bibliography:

City Data, "Philadelphia, Pennsylvania (PA) income map, earnings map, and wages," 04/2015, http://www.city-data.com/income/income-Philadelphia-Pennsylvania.html#ixzz4fB99IkZx.

Gibson, Campbell. "Historical Census Statistics On Population Totals By Race, 1790 to 1990, and By Hispanic Origin, 1970 to 1990, For Large Cities And Other Urban Places In The United States," *United States Census Bureau*, February 2005, https://www.census.gov/population/www/documentation/twps0076/twps0076.html.

Leung, May. "The Causes of Economic Inequality," *Seven Pillars Institute (SPI) for Global Finance and Ethics*, 01/21/2015, http://sevenpillarsinstitute.org/case-studies/causes-economic-inequality.

New Detroit Coalition, "Metropolitan Detroit Race Equity Report," 03/2014, http://www.newdetroit.org/docs/press/MetropolitanDetroit_RaceEquity_Report_NewDetroit_V1.1.pdf.

State of New York Department of Health, "New York City Health Indicators by Race/Ethnicity, 2012-2014," 08/2016, https://www.health.ny.gov/statistics/community/minority/county/newyorkcity.htm.

United States Census Bureau, "United States Statistical Abstract: 1999. Educational Attainment, by Race and Hispanic Origin: 1960 to 1998," Section 4: Education, 1999, https://www.census.gov/prod/99pubs/99statab/sec04.pdf.

U.S. Census 2010, "Chicago Median Income by Race," 11/2010, http://www.usacityfacts.com/il/cook/chicago/economy/.