Summer 8-1-2011

Revisiting the Dimensions of Residential Segregation

Harry Sharp
Georgia State University

REVISITING THE DIMENSIONS OF RESIDENTIAL SEGREGATION

by

HARRY F. SHARP, III

Under the Direction of Dr. Yixin Fang

ABSTRACT

The first major work to analyze the dimensions of segregation, done in the late 1980s by Massey and Denton, found five dimensions which explained the phenomenon of segregation. Since the original work was done in 1988 it seems relevant to revisit the issue with new data. Massey and Denton used the technique of factor analysis to identify the latent structure underlying the phenomenon. In this research their methodology is applied to a more complete data set from the 1980 Census to confirm their results and extend the methodology. Due to problems identified during the analysis confirmation was not possible. However, a simpler structure was identified which is comprised of only two factors. This structure is replicated when the methodology is applied to the 1990 and 2000 Census data thereby proving the robustness of the methodology.

INDEX WORDS: Thesis, Georgia State University, Segregation, Factor analysis, African-
                 American

REVISITING THE DIMENSIONS OF RESIDENTIAL SEGREGATION

by

HARRY F. SHARP, III

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2011

REVISITING THE DIMENSIONS OF RESIDENTIAL SEGREGATION

by

HARRY F. SHARP, III

Committee Chair:	Dr. Yixin Fang	
Committee:	Dr. Gengsheng Qin	
	Dr. Satish Nargundkar	

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2011

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION AND PURPOSE

1.1 Introduction

Segregation by race is generally looked upon as a myopic and barbaric practice in the modern era. Efforts have been made to eliminate the stereotypes of race and promote good relations between peoples of various races and ethnicity. Barriers which were erected to separate persons of different races have been struck down by the courts and, in some instances, mixing of the races has been mandated by the courts.

Historically, race relations have been strained. Blacks were not immigrants by choice but were rather brought to the American colonies by force as slave labor. Differing opinions between the North and the South concerning the use of slaves was a contributing factor to the Civil War. However, it should not be assumed that the Civil War resolved many of the questions of race relations.

Even after emancipation, African Americans in this country were not given their full rights as equal citizens until the federal courts mandated it. Race relations continue to be a cause of great turmoil in this country, as evidenced by the beating of motorist Rodney King in Los Angeles, the resulting riots after the acquittal of his attackers, the arrest of Dr. Henry Louis Gates in Cambridge, Mass., by a white officer, and most noticeably the debates on race during the 2008 election of President Barack Obama. Today, the treatment of African Americans in this country raises questions of social justice and equity. The importance of amicable relations between races makes it vital to study and understand the causes of conflict, one of the most obvious being segregation. Understanding residential segregation is therefore critical to understanding race relations in the United States.

1.2 Purpose

In an effort to better explain residential segregation, this research compares and contrasts the existing measures of residential segregation. Some indices of residential segregation have been validated over time while others have been discarded, but none have been accepted without reservation, or left unchallenged ([1]; [2]). The concept of segregation as a multidimensional phenomenon was advanced by the research community during the past decade, but no one has yet proposed a single index to measure these dimensions, although several indices make claim to this effect. In fact, most urban researchers simply continue to use individual indices, which measure only part of the phenomenon, while ignoring others ([2]; [3]). If progress toward understanding and quantifying segregation is to be made, then efforts toward validating its multidimensional quality and constructing an index to measure multiple dimensions must be attempted. A multidimensional index provides a better measure of the phenomenon than a unidimensional index.

Unlike a physical process that can be observed and measured directly, segregation is a phenomenon that must be measured indirectly. To analyze the causes and effects of segregation it is necessary to identify where it is occurring [4]. It can be sensed and known when it is seen, but it is not something that can be measured with a ruler or a meter of any sort. It can only be measured by constructing an index from observed data. In this regard segregation is an ideal application of the technique known as *factor analysis*.

1.3 Statement of Problem

Residential segregation of African Americans[1] and whites is mainly an urban phenomenon; moreover, it is one that did not come into wide-scale practice until after the end of the Civil War, when white Southerners reacted to the removal of the master-slave relationship by establishing a new social order with regards to racial proximity and race relations [5]. This trend, with significant implications for black/white racial relationships, was exacerbated by the large scale migration of southern blacks to northern cities which began at the turn of the twentieth century [6].

African American residences often clustered near the central part of the city in areas that were left by whites fleeing the influx of black immigrants to the inner city ([7]; [8]). Often the houses were of inferior quality due to absentee landlords, often white, who subdivided and resubdivided the remaining housing stock to house increasing numbers of poor, black rural immigrants [8]. The northern cities followed the

---

[1]Throughout this document, the terms "African American" and "Black" will be used interchangeably.

classic invasion-succession scenario, postulated by Burgess [9], where new immigrants to an urban area move first to older, centrally located housing in the central city [7].

In northern cities, blacks were forced into low-cost housing by economic hardship. White-occupied blocks that bordered the black areas converted quickly to all-black once the percentage of black families reached a "tipping point." This process was abetted by real estate brokers who played on white fears of black neighbors to make quick sales, and often windfall profits [8]. New devices were created to prevent black infiltration of white areas, such as exclusionary zoning, restrictive covenants, and deceitful practices by real estate agents ([7]; [8]). As a result, a higher degree of concentration of blacks was found to exist in northern cities (especially Philadelphia, New York, Chicago, and St. Louis) by the 1960s than in southern cities [8].

While much progress has been made in correcting the injustices of the past, segregation in housing still exists and actually has increased. In 1970, a surprisingly small proportion of the population lived in what could be called "integrated" neighborhoods (12 % of whites, 14 % of blacks) [10]. Further, a high percentage of the population in 1970, both black and white, lived in highly segregated neighborhoods, with an increasing tendency for black areas to cluster together. During the 1970s, residential segregation decreased in many Metropolitan Statistical Areas (MSAs) across the country, but southern MSAs actually experienced an increase in residential segregation [11]. This was attributed to the relative youth of the MSAs and the disappearance of traditional race relations ("backyard" segregation)[11]. Southern MSAs are rather young in comparison to northern MSAs as they owe their growth to the migration of population to the Sunbelt. Many of these cities, such as Atlanta, experienced major growth during the 1970s and 1980s (20% to 50% population growth in urban counties) [12]. This influx of new population caused rapid expansion of the urban area and caused a massive restructuring of the urban landscape, often displacing established black communities. This new growth often was manifested as new middle-class suburbs, which were commonly all white. Blacks were then forced to migrate into the central city to find housing. The concentration of blacks in the central districts of one Sunbelt city, Atlanta, was found to have increased dramatically during the past thirty years due to black in-migration from the rural areas to the central city and white out-migration to the suburbs, the formerly rural areas [13]. This rearrangement of the population eliminated the "backyard" segregation that was common in older Southern cities which relied on social separation rather than spatial separation [11].

The effects of segregation are greater than simply separating races. Because of *de facto* and *de jure* segregation, the issue of school desegregation was still being debated and fought over in the courts until the late 1970s even though it had been banned by the courts in *Brown vs. Board of Education* [14]. People are denied equal access to jobs, income differentials increase, and poverty becomes more centralized in our central cities, all due to segregation [15]. This persistent separation and stigmatization results in negative psychological impacts on the black population and prevents them from becoming full, productive members of society ([8]; [14]; [15]). Segregation in housing leads to segregation in other aspects of everyday life, leading to alienation of an entire segment of the population [16] and has been linked to the increase in black-on-black homicides [17].

1.4   Research Questions

The following research questions are investigated :

1. Can the results of Massey and Denton be replicated with a complete dataset (1980), thereby validating the methodology;

2. Can the results of Massey and Denton be replicated with current data (1990, 2000); and

3. Can the dimensionality of the results be simplified.

1.5   Significance and Justification

The ultimate goal of this research is to provide a more parsimonious description of the phenomenon of segregation. It is expected this will be accomplished by taking the results of Massey and Denton to the next level. If segregation is indeed a multidimensional phenomenon as is widely believed then to use a single statistic that describes only one dimension is to ignore the others. To adequately explain a phenomenon of this type, it is incumbent upon the researcher to use a measure that explains as much of the phenomenon as possible. Currently only univariate measures are available, but the results of this research should provide measures that explain multiple dimensions of the problem.

1.6   Organization of Thesis

The organization of this thesis will be as follows :

- Chapter 1 will be the problem introduction and statement;

- Chapter 2 will be a review of literature relevant to the problem and statistical technique;

- Chapter 3 will present the analysis to be done;

- Chapter 4 will be a discussion of the results from the analysis; and

- Chapter 5 will discuss the relevance of the results and directions for future research.

<center>**Chapter 2**</center>

<center>**REVIEW OF THE LITERATURE**</center>

2.1 Organization of Chapter

This chapter constitutes the review of the literature relevant to the problem under investigation. It is divided into two main parts, with each part further subdivided. The first part of the chapter reviews the qualitative aspects of segregation, including the processes which gave rise to it and the social and economic consequences of it. The second part of the chapter reviews the quantitative aspects of measuring segregation. The development of segregation measurement is traced with specific mention of significant indices or research. It is in this section where the multidimensional qualities of segregation are also discussed.

2.2 Qualitative Background

2.2.1 Defining Segregation

The literature on race relations and segregation is voluminous. Solutions proposed to address one aspect of the problem become issues in and of themselves. As an example, the concept of busing was proposed and implemented to reduce *de facto* school segregation which was a direct result of residential segregation. The controversy that erupted over busing spawned an entire branch of race relations literature. This review is limited to issues of segregation and its causes, segregation as an element of urban structure, and the economic and social ramifications for society from segregation. Before a phenomenon can be studied, it must be defined. The working definition of segregation for this research shall be :

> "...behavior that denies members of a racial group the rights or opportunities given to other groups, regardless of the formal qualifications of that group for those rights or opportunities."
> ([18, p. 430])

<center>6</center>

This definition encompasses all forms of segregation including residential, organizational, and educational. The most apparent form of segregation involves residential segregation between whites and African Americans although there are regional variations (white/Hispanics in the Southwest, white/Asians on the West coast). For the purposes of this research, all discussions of segregation are confined to those issues involving whites and African Americans, as this is the most pervasive and recalcitrant form of segregation in most of the United States.

2.2.2 Processes of Segregation

Segregation was not always an established facet of urban life. Prior to the migration of rural blacks to the North, blacks and whites lived together, often as equals. In Chicago, black professionals attended white churches and were prominent members of society[6]. This situation began to change with the influx of poorly educated southern blacks. The demise of sharecropping in the South and the lure of jobs in the North drew many poor blacks to the cities of the North[19]. Once there, they not only occupied historically black areas but any other accessible and affordable residential areas. These areas were often run-down and less desirable, which led to the propagation of stereotypes. White fear and distrust of the new immigrants led to the imposition of rules, laws, and behaviors which led to the concentration of blacks in highly segregated neighborhoods. Consequently, the northern blacks were shunned along with the southern immigrants due to "guilt by association"[6].

As a reaction to the massive influx, legal efforts were made in Northern cities to restrict blacks from certain areas. This effort, called racial zoning, took the form of zoning restrictions which forbade residence by blacks in areas for whites, and vice versa. Racial zoning did not survive legal challenge, and was deemed unconstitutional by the Supreme Court in 1917 in *Buchanan v. Warley* ([20]; [21]; [6]). A new instrument, the racial covenant, was developed and employed by whites and real-estate agents. These covenants specifically forbade the selling of property to anyone who was a member of a specified group, usually black. Actively promoted by all parties, including homeowners' association, boards of realtors, and even the F ederal Housing Administration, covenants were popularly employed until they were outlawed in 1948 in *Shelley v. Kraemer* ([20];[22];[23]).

A widely recognized process contributing to the problem of segregation, especially of the residential form, is "white-flight." This process was a continuation of the institutionalized residential segregation that existed before World War II. Whites were extremely resistent to integration of neighborhoods, and to a large

extent still are ([8]; [24]; [23]). In this process, white homeowners held a "front line" against expansion of the black ghetto. Once a black homeowner succeeded in making a "beachhead" position on an all-white block, a process of invasion-succession began. This is not invasion in the common sense of the word, as blacks do not have a plan to take over entire blocks, thereby expanding the ghetto. Rather, it is invasion as seen from the white homeowners' point of view; a breach of the bulkhead, as it were. The reaction of white homeowners, often aided by real estate agents, was to quickly put their homes up for sale due to fear of devaluation. This often became a self-fulfilling prophecy, at least in the short-run ([25]; [8]). In the long-term, however, property values stabilized. The decline in value was only during the periods of "invasion" ([26]; [27]). Neighborhoods quickly changed from all-white to all-black, with the white residents leaving for the suburbs and "safety." This resulted in the expansion of the ghetto and the increasing concentration of blacks in the central city ([21]; [28]; [29]; [10]; [23]). Real-estate agents aided and abetted the concentration of blacks into segregated areas. In 1924, the National Association of Real Estate Boards adopted Article 34 of their Code of Ethics. This article stated

> "A Realtor should never be instrumental in introducing into a neighborhood a character of property or occupancy, members of any race or nationality, or any individuals whose presence will clearly be detrimental to property values in that neighborhood." [21]

This article institutionalized the belief that blacks caused the property values in a neighborhood to decline, despite a lack of supporting evidence ([6]; [8]; [30]), or evidence to the contrary ([31]). This article was strictly enforced, to the point of expulsion from the National Board if a realtor were to sell to a black ([32]). Pressure was also brought to bear on realtors by communities to prevent any abrogation of this belief. Realtors often feared reprisals, either by loss of business or physical violence if they sold property in white areas to blacks ([21]; [8]). Although officially discouraged as a practice, realtors still direct blacks away from predominantly white areas through subtle means such as "steering", showing prospective buyers houses only in neighborhoods with like race residents, requiring more information, and being unavailable to show houses ([29]; [24]; [33]).

Realtors helped promote the exclusion of blacks from white areas, and received indirect assistance in this from the Federal government. Beginning in the 1930s, the Federal government instituted programs to promote home ownership. This was declared to be part of the American dream, a way out of the Depression, and a benefit to society. Housing has always held a special place in the American psyche. In contrast to

cities in the Old World, American cities are sprawling entities composed of single family dwellings on large lots ([34]). Homeownership was long associated with affluence and with power in American society, but it also conferred a sense of respectability upon the owner ([35]). Society also believes that homeownership is a merit good, one which people should be encouraged to buy, even if they are not otherwise inclined to do so. A merit good is something that society places a higher value upon than the individual might (housing, education, medical care). This good often confers some benefit to society at large, so the society is willing to subsidize the acquisition of this good ([36]). But what benefit does society derive from private homeownership and its subsidization? During the Depression, the goal of the Federal programs was to put people back in houses or on farms where they would have adequate housing. The creation of new housing after World War II provided jobs for returning servicemen as well as a boost to the economy while it shifted from a military emphasis to a consumer emphasis ([34]). The social benefits were increased housing stock and an increased tax base.

There are also other, less tangible, benefits to society from homeownership. The intangible benefits are the more controversial aspects of subsidization of homeownership. Through homeownership, it is believed, people improve their standard of living and concomitantly their expectations and aspirations. Improvements in aspects of personal life make people into better members of society and bring them closer to the body politic in thinking. Subsidization of homeownership is also believed to result in a reduction in negative externalities of poor-quality housing, such as incidence of fire, communicable disease, crime, and other social disorders (e.g. delinquency) ([36]). It was to achieve these goals, both tangible and intangible, that the Federal government instituted programs to assist home buyers.

The first of these programs was the Home Owners' Loan Corporation (HOLC), whose purpose was to refinance mortgages in danger of foreclosure and assist farmers in reclaiming their land through low-interest loans. Unfortunately, the HOLC formulated the practice of "redlining," where areas were graded as to their suitability for loans, with the lowest rating being assigned the color red. Areas that fell into this category were colored red on housing maps, and loans were prohibited in these areas, or areas next to them ([23]). As many of these areas were established ghettos, this led to the expansion of the ghetto areas, due to white fears of loss of property value, and the decline of the housing stock located within and near these areas. This practice was adopted by the successors to the HOLC, the Federal Housing Administration (FHA) and the Veterans' Administration. The practice was institutionalized by the banks and lending institutions that dealt with the Federal agencies. The FHA was also guilty of promoting restrictive covenants as a method

of insuring the safety of loans guaranteed by the agency ([6]). In fact, the official language in federal documents mentioned "infiltration of business and industrial uses, lower class occupancy, and inharmonious racial groups" ([6]).

All aspects of segregation are derived, ultimately, from the belief that separation of races is a desired goal of both races. Sometimes the belief is that blacks, as well as whites, are more comfortable and happier with their own. Other times it is because of the belief that the races are incompatible. These attitudes are partially true for both blacks and whites, although more so for whites, who view the situation from a more traditional viewpoint ([7]; [21]; [6]; [37]).

Although the American creed holds all people to be created equal, there are still distinct delineations in American society. Often there are mentions of classes in society (lower, middle, and upper). But to maintain this distinction it is necessary that some classes of people be rated below others. In the minds of many whites, blacks occupy a class below them, regardless of socioeconomic status ([7]; [32]; [6]; [38]; [39]; [40]). The attitudes of whites toward blacks were reinforced by residential separation. With little or no interaction between the races and discriminatory policies that forced blacks to live in geographically small, overcrowded areas, various social ills afflicted blacks disproportionately (unsanitary living conditions, overcrowding, crime, concentration of poverty, dysfunctional families) ([41]; [3]; [42]; [43]), reinforcing the stereotypes ([6]; [23]).

This separation is also having a negative affect on the cohesiveness of American society. Many blacks see the seemingly insurmountable barriers to assimilation and decide to not assimilate, but to self-segregate. Blacks decide to pull their money from white banks and put it in minority owned banks. They begin to patronize minority owned businesses rather than the white owned businesses. They begin to demand classes in African history and culture to reinforce their separateness. In some cases, the "separate but equal" philosophy is being resurrected, but this time by the blacks themselves. Special voting districts are created to insure election of a representative who is of the same color as those who elected him, resulting in racial "homelands" within the U. S. ([43]; [44]; [45]; [46]). In cases where blacks have managed to escape the inner city ghettos they live in black suburbs rather than in the white suburbs ([47]; [48]). This speaks to the failure of American society to assimilate the African Americans. While other immigrant groups have advanced socioeconomically and dispersed residentially, the same cannot be said for African Americans. This lack of socioeconomic advancement is an impediment to the assimilation of African Americans into society as a whole ([49]; [48]).

2.2.3 The Economic Consequences of Segregation

As urban areas have grown outward, usually due to improved transportation networks, jobs have followed the population and gravitated towards cheaper costs which are associated with the suburbs. This is referred to as "white-flight", discussed earlier, or sometimes as the "pull" of the suburbs ([24]). This has lead to the decline of the Central Business District and the loss of many sources of inner-city employment. This emigration of jobs has left people without adequate transportation effectively stranded in the central city, unable to profit from the growth of the city, and unable to take advantage of those economic opportunities that would raise their standard of living ([50]; [51]; [52]). In isolated instances where the public transportation system is well established this is not necessarily the case, but the misallocation of jobs is still a serious problem ([53]). What jobs have been created in the central cities are primarily white collar, and not available to the African American worker. Those workers who succeed in obtaining a job in the suburbs often see their earnings adversely impacted by commuting costs ([24]), or the high cost of automobile maintenance ([51]). One result has been high concentrations of poverty in inner city neighborhoods ([54]). The option of moving to the suburbs is often only available to upper- and middle-class whites, resulting in a concentration of middle- and lower-income blacks in the central city ([43]). This form of *de facto* segregation has been observed since the 1960s ([28]; [55]; [10]).

The result is a closed system, which Fusfeld ([41]) describes. In this system, racial attitudes have kept blacks in the ghetto and deserted the city, leaving it impoverished and unable to provide the social services necessary to prepare blacks for jobs. Even those who succeed in escaping the ghetto are only exacerbating the problem by removing positive role models and possible entrepreneurship which might help raise the living standard of the ghetto. The poverty cycle in the ghetto therefore has four self-reinforcing influences :

1. "Poverty breeds a style of life which reinforces the conditions that lead to poverty.

2. Resources that might lead to betterment and development are drained out.

3. Lack of political power has brought public programs that are often harmful to the ghetto economy (i.e. welfare).

4. White attitudes toward race have kept most of the ghetto residents from moving out. ([41])"

This isolation is exacerbated by the creation of all-white suburban "paper towns" that ring the central cities and prevent expansion. These suburban towns are home to workers who have white-collar jobs in

the central cities. These workers make use of city services and roads but take their paychecks back to the suburbs, making no contribution to the maintenance of the city or its tax base ([36]). Where the annexation laws are strict and prevent expansion of the cities to include these suburbs, the central cities are impoverished ([42]).

The end result is deleterious to the cities. The increase in previously mentioned social ills is paralleled by an increase in demand for social and protective services. People whose main concern is simply living do not have the time to be concerned with the upkeep of their housing. This leads to deterioration of the housing stock and the creation of slums, further depressing the tax base. This weakened fiscal condition of the central cities results in a decreasing availability and quality of social services, providing more impetus for middle- to upper-income families, both black and white, to leave for the suburbs ([24]). The cycle feeds upon itself.

2.3   Quantitative Background

2.3.1   Measuring Segregation

Early work - 1930 to 1970

The majority of the work done in measuring segregation and developing indices to measure segregation was accomplished by sociologists. Some work was done by geographers, but only since the 1980s. The development of segregation indices began, however, with an economist.

The measurement of segregation can be traced back to a paper by Hoover ([56]) who proposed a statistic to measure the distribution of industrial locations around a city. Hoover was interested in the distribution of certain types of industries within the urban area and how the population was distributed in relation to each industry. While his paper was economic in nature, he did note the applicability of his measurement technique to the distribution of populations, specifically the black population. In a later paper Hoover ([57]) applied his index to the measurement of black and white populations and their distributions. He called his measurement "coefficient of redistribution" as it could be interpreted as the percent of population that would have to move to make the distribution even.

Hoover set the precedent of measuring against what was assumed to be the absence of a phenomenon (localization, segregation). He used the concept of a Lorenz curve to measure the deviation from what was assumed to be a perfect distribution (Figure 2.1). The Lorenz curve was originally developed to measure

the distribution of wealth among a population ([58]). The perfect distribution was a line, drawn from the origin, at a 45 degree angle. This line arose from the assumption that, in an equal distribution, the percent of a resource controlled by a segment of the population would be equal to the percent of the population which that segment constituted. The Lorenz curve is produced by ordering the individual points being studied (e.g. census tracts, census blocks) from lowest to highest in value. The area under the diagonal is considered the deviation from the perfect distribution, or as Lorenz stated, "the rule of interpretation will be, as the bow is bent, concentration increases" ([58, p. 219]). Restated, this means the closer a Lorenz curve comes to the diagonal, the less severe the inequality ([59]). For an in depth analysis of the mathematical properties of the Lorenz curve and related measures, see Gastwirth ([60]). The Lorenz curve has become "the most commonly accepted measure of inequality" ([59, p. 270]) and has been used to study income differentials in the United States ([61]; [62]) and China ([63]), agricultural policy in South Africa ([64]), the effect of land reform in Pakistan ([65]) and the tourist industry in South Korea ([66]).
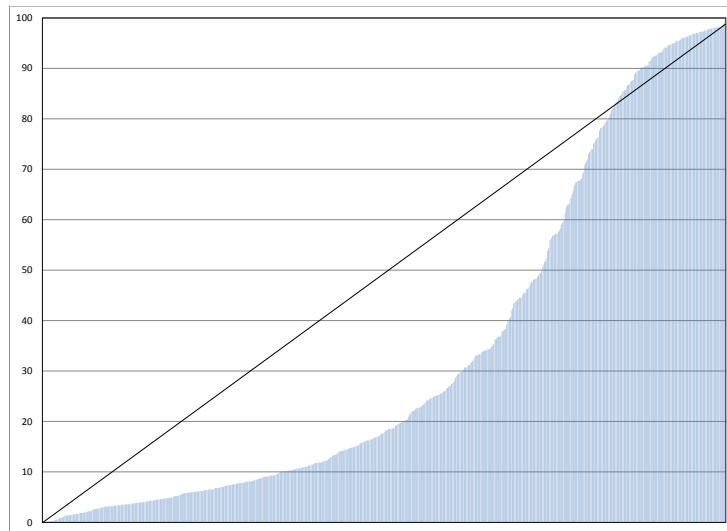
Figure 2.1: Lorenz Curve for Percentage Black by Tract for Atlanta (2000 Census)

In the 1940s, a group of sociologists, Julius Jahn, Calvin F. Schmid, and Clarence Schrag, published the first paper to specifically discuss the measurement of segregation ([67]). They termed it "ecological segregation," but the issue under consideration was segregation of races. In this paper they presented four indices that measured various facets of the segregation problem selected from a claimed multitude of derived

indices. Three notable contributions to the study of segregation were made in this paper: bounds on an index, criteria for acceptability of an index, and the mathematical formulation of Index Four.

The first contribution was the placing of bounds on an index. Jahn, Schmid, and Schrag made the observation that a satisfactory index of segregation must meet two basic criteria. The first criterion stipulated that in situations where there was no segregation, the value of the index should be zero (0). The second criterion stated that in situations where total segregation existed, the value of the index should be one hundred (100). This standardized the interpretation of segregation scores and made them more understandable. The value of the index could be interpreted as a percentage of its own maximal possible value. The second contribution was a catalog of five criteria used for evaluating the utility of an index of segregation. The five criteria state that a satisfactory measure of segregation should :

1. be expressed as a single quantitative value so as to facilitate such statistical procedures as comparison, classification, and correlation;

2. be relatively easy to compute;

3. not be distorted by the size of the total population, the proportion of Negroes, or the area of a city;

4. be generally applicable to all cities; and

5. differentiate degrees of segregation in such a way that the distribution of intermediate scores cover most of the possible range between the extremes of 0 and 100. ([67, p. 294])

The third contribution of the paper was the formulation of Index Four. This index was in actuality a formalization of Hoover's measure. Hoover never formally stated his measure, only alluded to it and demonstrated its use. This index was later found to be the most comprehensive of the four developed by Jahn *et. al* ([68]), and still later was elevated to a place of preeminence in the measure of segregation as the Index of Dissimilarity ([1]).

In his critique of Jahn *et. al* , Hornseth attacked the individuality of the four indices proposed. He showed mathematically that Index Three and Index Four were the same. Jahn *et. al* had found them to be highly correlated. Hornseth also showed that indices one and two were mathematically the same. Jahn *et. al* had found them to be highly correlated as well. Hornseth's final proof showed that Index One was in reality an "incomplete measure," or subset, of Index Four, leaving only Index Four as a useful index ([68]).
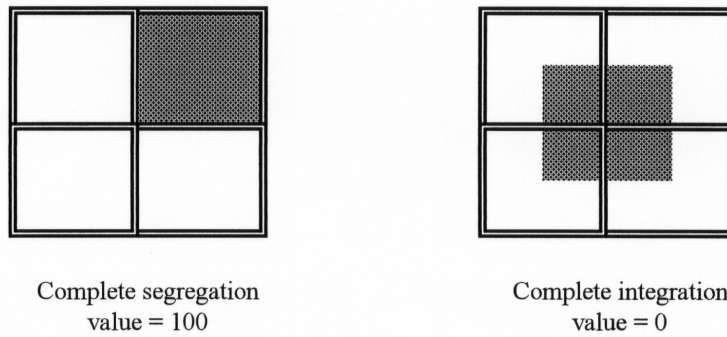
Figure 2.2: Cowgills' Illustration of Weakness in Jahn's Index
(Source : Cowgill and Cowgill, 1951)

A second critique of the work of Jahn *et. al* was done by Williams in 1948. In her paper, she proposed that measures based on Chi-square distributions were preferable, and that Jahn *et. al* had in fact derived Chi-square measures in their work. She successfully showed that Index One and Index Four were, in fact, Chi-square measures, and simpler to compute than other established Chi-square measures ([69]). This provided more support for the indices.

Jahn proposed yet another index to measure racial segregation, but this time he based his methodology on the criterion of reproducibility postulated by Louis Guttman ([70]). Reproducibility is a concept from scale design, which is concerned with the percent of all individuals in a defined population who can be correctly classified with respect to an original set of variables. A high reproducibility score indicates a well constructed scale. Jahn applied this concept to the measurement of segregation and developed his "index of segregation." He tested his index on the population of the urban area of Seattle, Washington. Jahn made no claims for this index, other than to state it was simple to derive and compute. Jahn further proposed his methodology might be applied to other problems of index construction.

This approach to measuring segregation was not without its critics. Cowgill and Cowgill ([71]) raised the first serious flaw with the indices proposed thus far. They noted the effect of areal unit choice on the resulting values produced by the indices. Most of the indices proposed to this point were based on census tracts, and assumed each tract was heterogeneous in population. Cowgill and Cowgill provided an example which showed two instances where a hypothetical city of four sectors is either perfectly segregated, or perfectly integrated, simply by moving the location of the minority population (Figure 2.2). This weakness was actually illustrated by Jahn himself when computing values for residential segregation in Seattle. The value

was lower for census tracts (0.17) than when the index was based on census blocks (0.63) ([70]). To bolster their conclusion, Cowgill and Cowgill derived an index based upon census blocks and then computed scores for 187 cities. They found high segregation (> 0.75) in cities of the South Atlantic region, intermediate segregation (0.75−0.25) in the other southern states and the north central states, and low segregation (> 0.25) in cities in New England, the Mountain, and Pacific regions. The Cowgills did not derive any substantive conclusions from their results, but did hypothesize the reason for high segregation in some areas was due to a change in social patterns and adherence to the most conservative forms of racial etiquette. They did note, however, their results mirrored those of Jahn *et. al* . Another index derived from the economic literature is the Gini index. It is based on the Lorenz curve, like Hoover's, and explains the area between the Lorenz curve and the diagonal, expressed as a fraction of all area under the diagonal ([72]). It is related geometrically to the Index of Dissimilarity and, like the index of dissimilarity, it varies between zero (0) and one (1). It is more sensitive to transfers of minority members between all areas, not just those between areas of over- and under-representation ([72]; [2]). The Gini index is not used often in segregation analysis due to its tedious calculation ([72]). This should change, however, with the faster computers available today ([60]; [73]).

The paper by Duncan and Duncan ([1]) put all the discussion of segregation indices to rest for almost 20 years. In their paper, they found many of the prevailing indices (e.g. Gini (Index Three, [67]), Cowgill's, the "Nonwhite Ghetto" index (Index One, [67]), Jahn's "Reproducibility" index) could actually be stated in terms of Index Four, which they termed the index of dissimilarity, or D. While they did not explicitly state so, this paper was used to promote D as the statistic of choice when measuring segregation ([74]; [72]; [2]).

The Index of Dissimilarity

Originally proposed in the 1940s as Index Four ([67]), D was established as the predominant statistic by the Taeubers in the 1960s with the publication of their landmark book ([75]). The statistic has many aspects which recommend it, among those being ease of computation and comprehension (see Appendix I for a computational formula). The values on the index range from zero (0), for no segregation, to one (1), for maximum segregation. This index measures the evenness of the racial groups in the city, and represents the number of people who have to move from all areal units being studied to bring the racial composition of all areas into line with the city at large ([76]). D can therefore be thought of as the ratio between the number of people of group X who must move (M^x) and the total population of group X in the study area, or $\frac{M^x}{X}$.
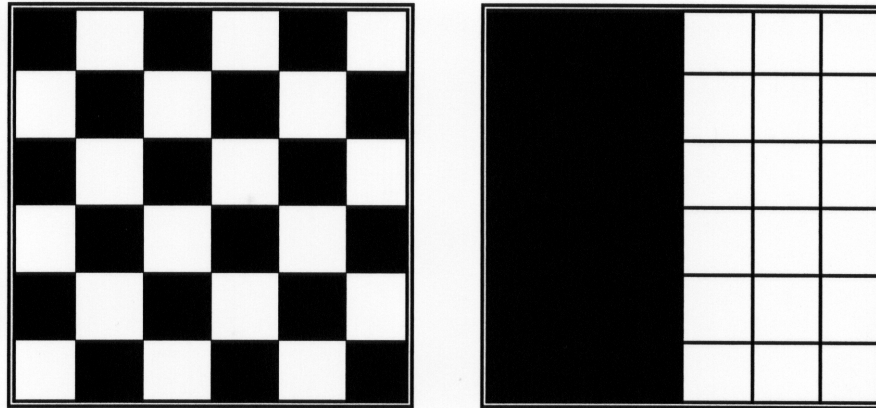
Both patterns have D = 0

Figure 2.3: Graphical Representation of Insensitivity of D

Problems with D

With the publication of a critique of D by Cortese, Falk, and Cohen ([77]), the general consensus on the acceptance of D as the normative measure of segregation was abrogated. Its faults are numerous, the most damaging of them being its sensitivity to the size of the areal unit being used in the computation ([71]). Another flaw is its insensitivity to spatial distribution of the computational units which allows the same value to be computed for drastically different distributions (Figure 2.3).

Some questions have been raised about the correct interpretation of the statistic ([78]; [4]), but those questions have been addressed and resolved ([79]). In their article, Cortese, Falk, and Cohen listed four major objections to D, which were :

1. "The expectation of "evenness" as the opposite of segregation is not as useful in most cases as the concept of "randomness."

2. D is affected by differences in the proportion of the minority in the population, thus preventing inter-city comparisons.

3. D is affected by the size (number of households) of the areal unit of analysis.

4. The present interpretation of D as the proportion of nonwhites who would have to change their tract of residence to make the distribution of the minority even throughout the city is misleading since it does not include the concept of replacement of the relocated minority." ([77, p. 631])

Despite all the attempts to replace D, it has withstood all assaults ([2]). Several attempts were made to modify D to make it less susceptible to the change in areal unit size and to incorporate more of the spatial aspects of the phenomenon, but none were adopted by the research community ([80]; [74]; [81]; [82]; [83]). Some suggested standardizing D to reduce its variability ([77]; [84]). Standardization is a statistical technique where the mean value for a statistic is subtracted from the actual value and the difference divided by the variance of the statistic. The end result is a number that measures the degree of departure from the mean. The standardized value is computed using the established formula,

$$Z = \frac{D - \mu_D}{\sigma_D}$$

where D is the overall value for the Index of Dissimilarity; μ_D is the mean of the values for D for a city; and σ_D is the standard deviation for the values of D for a city. This reformulation of D, they felt, would answer the objections to D and allow intercity comparisons. A similar modification was proposed by Winship to test the deviation of an established pattern from random segregation ([4]).

Massey ([85]) disagreed with this conclusion. He showed there was in fact no benefit to using the standardized form of D. Further, the standardized form lacks interpretability, as Massey makes clear, since deviations from the mean, which standardized values are measured in, are meaningless in relation to segregation ([85]; [86]).

The Index of Dissimilarity has been both hailed and reviled ([87]). Having been sanctioned as the statistic of choice for measuring evenness of population distribution ([2]), however, it appears that D will continue to be the primary statistic measuring segregation.

Spatial Modifications to D

Beginning in the 1980s, several spatially modified versions of D were proposed in the literature. These indices attempted to resolve one of the flaws in D which was its spatial insensitivity. These indices also measured more than one aspect of the segregation problem, a first step toward a multidimensional view of segregation. The spatial indices were appealing because they were "dimensionless statistics derived from the ratio of measured segregation in two geographically identical settings" ([81, P. 1243]). This made them comparable among cities, addressing another of D's flaws.

In 1981, Jakubs proposed a spatially modified version of D that he called DBI, or Distance Based Index of Dissimilarity ([80]). In formulating DBI, Jakubs used two important concepts, each of which is the antithesis of the other: uniformity and exclusivity. Uniformity, or zero-level segregation, is a condition which exists when population proportions by group are constant across areal units. Exclusivity, or maximal segregation, exists when each areal unit is occupied solely by members of one group ([79]; [80]). DBI is then conceptualized as the ratio between the minimum distance that minority members would have to move to achieve uniformity from the current state (Z) over the minimum distance required to achieve uniformity from the maximally segregated state (Z^*), or $\frac{Z}{Z^*}$. Effort is defined as distance traveled, in person-miles, by an individual or household ([80]; [2]).

The computation of DBI uses a form of linear programming known as the transportation problem. This is a special case of the standard linear programming problems used in management science. Linear programming was used extensively in geographical research, so this is not a new technique. It is, however, an application of linear programming to an area previously unexplored.

In the transportation problem, a matrix is formed of supply centers and demand centers. Items flow from supply centers to demand centers. In DBI, the supply centers are tracts with an excess of white population, and the demand centers are tracts with white population deficits. Once the matrix is formed, it can be solved by standard linear programming, such as simplex method, and the result is the optimal solution that minimizes the overall cost. In this case the cost is the total distance traveled to achieve an optimal distribution of whites and blacks in an urban area.

The only major critique of DBI was done by Morgan ([76]). His concerns were more with the procedure for computing DBI than with its mathematical properties. In fact, Morgan affirmed that DBI was less sensitive to the change in areal unit size than D and a valuable contribution to the field of segregation studies. This decreased sensitivity to areal unit size answered a complaint with D dating back to the Cowgills ([71]). Morgan also commented on the ability of DBI to measure multiple aspects of segregation (evenness, racial homogeneity, clustering, and location) ([76]). The procedural questions involve the placing of the initial white and non-white tracts and the selection of the maximally segregated pattern.

Morgan proposed a competing spatial measure that he called the modified distance based index (MDBI) (Morgan, 1983). This was a modified form of DBI which he had developed independently of Jakubs. The major difference between DBI and MDBI is the concept of maximal segregation. Morgan believes this can be accomplished in a different manner than that proposed by Jakubs. By using Morgan's methodology,

the effect of the location of the black ghetto is removed, to which DBI is susceptible. Morgan uses a hypothetical black ghetto that has its mean center within a specified distance from the actual mean center of black population in the city. This yields results that are more accurate. MDBI is not without flaws, however. MDBI is more sensitive to tract size than DBI, but Morgan places less emphasis on this and focuses instead on the inclusiveness of the index (multiple dimensions of the problem). Morgan makes the same observation (inclusivity rather than insensitivity) about DBI.

White ([74]) proposed an index based on the Cartesian distance formula for distance between two points, derived from a series of four equations measuring the mean proximity between and among members of two racial groups. These were then combined with population counts for each group. This statistic, which he called P, ranges from one (1.0) for no racial clustering, and has a value higher than one when there is evidence of clustering. It is more complex to compute than D, requires more data than D, and does not conform to the accepted standards for indices in that it does not have an upper bound.

A later addition to the spatial modifications fold was proposed by Wong ([83]). He formulated and proposed a measure of spatial evenness that attempted to account for the length of the border of the areal unit inhabited. Wong postulated that a longer border would provide more opportunities for interaction with other populations, and therefore should be incorporated into the measurement of spatial separation. Wong did not, however, consider the permeability of the border. As was noted in the studies of spatial diffusion in the 1960s, not all barriers to travel, or dissemination, are impermeable. Some are more permeable than others, while still others are transparent. If the borders between two groups are permeable, then Wong's measure is applicable. However, if the border is not permeable (e. g. a river, a chasm), then the length of the border is irrelevant to interaction. Some borders, while not physical, may be socially impermeable, which is an equal deterrent to social interaction and an equally strong means of segregation.

2.3.2   The Multidimensional Qualities of Segregation

The Hypothesis of Multidimensionality

Beginning with Duncan and Duncan ([1]), some mention was made as to the possibility that segregation was not measurable by one index alone as there were multiple aspects to the problem. Duncan and Duncan specifically stated,

20

Evenness	the differential distribution of two social groups among areal units in a city
Exposure	the extent to which minority and majority members physically confront one another by virtue of sharing a common residential area
Concentration	the relative amounts of physical space occupied by a minority group in the urban environment
Centralization	the degree to which a group is spatially located near the center of an urban area
Clustering	the extent to which areal units inhabited by minority members adjoin one another in space

Table 2.1: Dimensions as defined by Massey and Denton

"As we have suggested, it may be that no single index will be sufficient, because of the complexity of the notion of segregation, involving as it does considerations of spatial pattern, unevenness of distribution, relative size of the segregated group, and homogeneity of sub-areas, among others." ([1])

Jahn *et. al* had originally made the claim of numerous indices, each measuring a different facet of segregation, but they had not made claim to a single index for all purposes ([67]). The concept of segregation was acknowleged to be complex and multifacted, but most researchers had accepted the "emasculation" of the phenomenon by abstraction to allow measurement ([88]). Beginning with the spatially modified indices, claims were made as to the abilities of the indices to measure multiple aspects of segregation, such as evenness and clustering ([80]; [76]; [81]; [89]). These claims were not seriously investigated, however, most likely due to the complexity of computation of the indices.

In 1988, Massey and Denton undertook to identify the underlying dimensions of segregation and specific statistics to measure each one ([2]). In this paper, they took twenty common indices (Appendix I), computed values for each of them for 60 MSAs, and then factor analyzed the data to identify the underlying structure. Massey and Denton *a priori* identified five spatial aspects of segregation : evenness, exposure, concentration, centralization, and clustering (Table 2.1). Statistics were chosen that measured each of these aspects, or dimensions. From the results, Massey and Denton then identified the statistic which they felt best measured each dimension (see following Section). This paper has achieved some degree of stature in the literature, having been cited more than five-hundred times since its publication.

Some questions still linger, however. The multidimensional quality of segregation seems to be accepted by the research community as evidenced by the numerous citations of the Massey and Denton work, long

standing consideration of the concept ([1]), and no evidence of challenge to the Massey and Denton work. Some of the results from Massey and Denton would seem to indicate there may yet be fewer dimensions to the problem. Some of the factors derived from the data were highly correlated and showed evidence of collapsing to only two, or possibly one, factor ([82]). Perchance a single index cannot measure all aspects of segregation, but could a composite index made up of available statistics measure the phenomenon? Or could the spatial indices truly be multidimensional? They have not adequately been examined due to their complex nature and difficulty in generating values. It would therefore seem that Massey and Denton have brought us only half-way, and there remains work yet to be done to completely resolve the issue.

Measuring the Dimensions of Segregation

As previously mentioned, Massey and Denton chose a single index from each of the factors to represent that dimension of segregation. Each of these indices was chosen to be the "single best indicator" ([2], p. 282) for its dimension. This approach is known as "surrogate variables" wherein a single variable from a factor is chosen to represent the entire factor. The variable chosen does not always have to be the one which loads highest on a factor. The variable is chosen more for its descriptive and illustrative properties.

For example, D was chosen as the statistic of choice to measure evenness even though it was not the highest item to load on the factor (Atkinson index with b=0.9 loaded at 0.99 versus D at 0.84). It was chosen due to its long use in the field, its simplicity of computation and interpretation, and again because it was shown to be highly correlated with other similar measures. No clear choice emerged for exposure, as the interaction index is the reciprocal of the isolation index, and they both loaded equally on the factor (interaction at 0.82, isolation at -0.82). The methodology is basically sound, however, and their assumptions clearly stated in their conclusions.

The first dimension deals with the distribution of groups among areal units in a city. The zero condition is an even distribution among all areal units, such that each unit has the same proportion. This dimension was therefore named evenness. Figure 2.4 shows the zero case of high evenness. If every unit (tract) had the same proportion of minority and majority, then the MSA would be considered "gray." Figure 2.5 shows the opposite condition, low evenness. In this instance, each unit (tract) is either totally minority or totally majority. The result is a checkerboard MSA.

The Index of Dissimilarity (D) was chosen as the representative statistic for this dimension. The method of computing D, its properties and disadvantages were discussed in a previous section. The values for this

statistic range from zero (0), when all units have the same relative number (%) of minority and majority members as the city as a whole, to one (1), when no minority members share a common area of residence with a majority member.
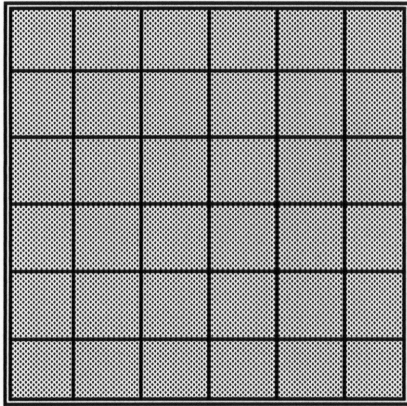
Figure 2.4: High Evenness
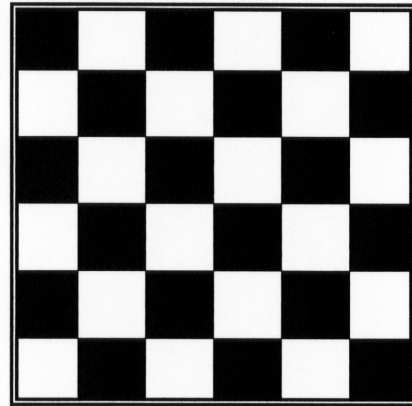
Figure 2.5: Low Evenness

The second dimension, named exposure, deals with interaction between groups and the degree of potential contact (possibility of interaction) between members of minority and majority groups. This dimension came directly from the sociological literature. The concept is rather simple: How likely is a member of one group to come into contact with a member from another group? In Figure 2.6, an ideal situation is shown. If there are four minority tracts in our MSA, then locating them at four separate points, surrounded by majority tracts, gives maximum probability of interaction between the two groups. Every unit has the same probability of interaction. In Figure 2.7 the converse is shown. If all minority tracts are located in a
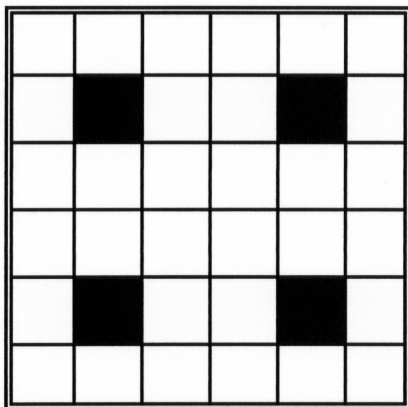
Figure 2.6: High Exposure
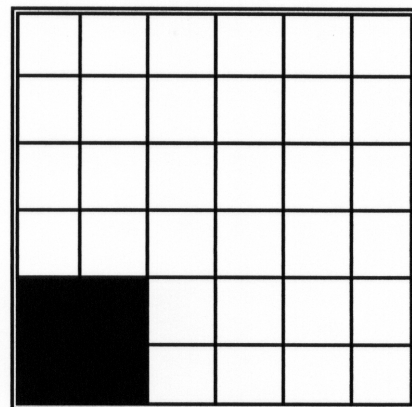
Figure 2.7: Low Exposure

confined spatial area, then the probability of interaction between the two groups is uneven across the MSA, lowest for the tracts bordering the minority enclave to highest for those tracts in the northeast corner.

Bell's interaction/isolation indices ([90]) were chosen as the representative statistic. The interaction index is the minority-weighted average of each spatial unit's majority proportion and measures the extent to which members of minority group *X* are exposed to members of majority group *Y*. The converse of this measure is the isolation index, which measures the extent to which minority members are exposed only to one another, rather than to majority members. It is computed as the minority-weighted average of each unit's minority proportion. Both may be interpreted as the probability a randomly drawn *X*-member shares an area with a member of *Y*, or with another *X*-member. The values range from zero (0), which is the minimum probability (none), to one (1), the maximum probability. In a two group case, the sum of the two measures equals one (1) ([2]).

If the first two dimensions are examined closely, the high degree of similarity between the two can be seen. The ideal case of evenness requires a uniform proportion minority across all tracts. The ideal case of exposure requires an equal probability of minority/majority interaction for all tracts. If one of these conditions is achieved, the other has a high probability of also being achieved. The results from Massey and Denton bear this out. The degree of intercorrelation between the two factors is 0.54 ([2]), indicating a strong positive relationship between the two.

Note, however, that while the concepts of evenness and exposure are inherently spatial (see Figures 2.4 and 2.6) the indices chosen to measure these dimensions are aspatial. In fact, this is a charge leveled at D many times in the past. It has been shown previously that two distinct spatial patterns can yield the same value for D and yet be radically different social situations (Figure 2.3). The same can be shown for Bell's measures of exposure.

The third dimension, concentration, deals with the relative amount of physical space occupied by a minority group. A minority group restricted to a defined geographical area is considered more segregated than one dispersed across the entire area. A high degree of concentration would therefore indicate a high degree of segregation. This concept is related to evenness. A group which is highly concentrated would not be evenly distributed across the urban area. In Figure 2.9 the ideal case of low concentration is shown. Note the low density of the minority population as indicated by the gray tracts, each of which are 50% minority

24
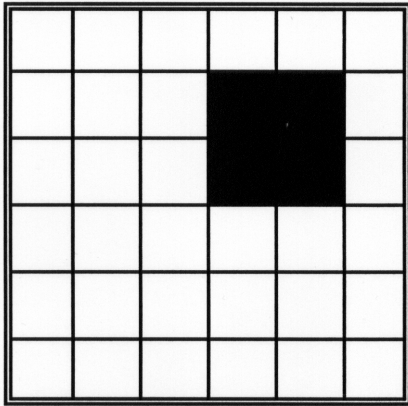
Figure 2.8: High Concentration
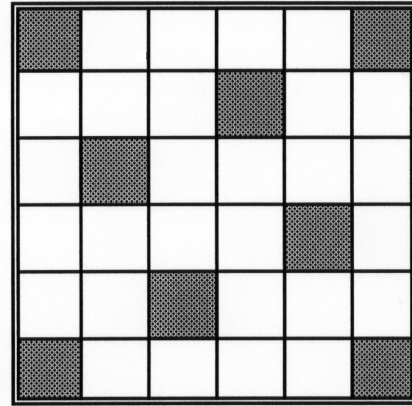
Figure 2.9: Low Concentration

and 50% majority. The opposite situation is shown in Figure 2.8, where the minority population is isolated in only four tracts, each of which is 100% minority.

The relative concentration index (RCO) was chosen as the representative statistic for this dimension. This index measures the geographic concentration of minority group X in an absolute sense, regardless of how majority group Y is distributed. It compares ratios of X to Y concentrations with the maximum possible ratio if X were maximally concentrated and Y minimally concentrated. The values for the index range from -1.0 (Y more concentrated than X) to 1.0 (X more concentrated than Y), with 0 meaning the two groups are equally concentrated.

This is more easily understood if the tracts are allowed to differ in size, rather than be regular as in the illustration. Assume the two population groups (X and Y) are equal in size and exclusivity is the norm (no mixed tracts). If population X lives in tracts which are one half the size of the tracts in which population Y lives, then group X is much more concentrated than group Y. The same number of people are being forced to live in half the space. This is true regardless of where the tracts are located.

Centralization, the fourth dimension, considers the degree to which a group is spatially located near the center of an urban area. Minority groups have traditionally occupied central city areas in the US. This is where the lowest quality housing is, and where most new immigrants will initially locate (according to Burgess' zonal theory) ([9]). This dimension is related to concentration, but it is not an exact copy. Minority groups may be confined spatially, but the location does not necessarily have to be the central city. In some places, gentrification has displaced the minority groups and scattered them across the urban area, or displaced them to a new confined area on the outskirts of the MSA. In the developing countries, this is

25
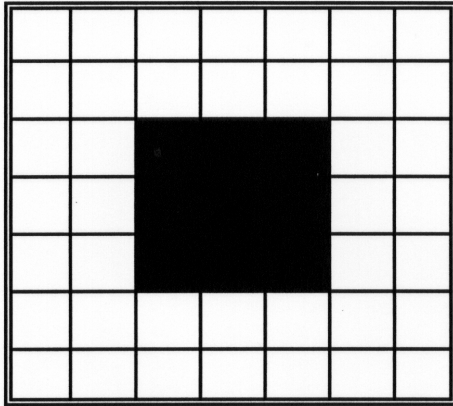
Figure 2.10: High Centralization   Figure 2.11: Low Centralization
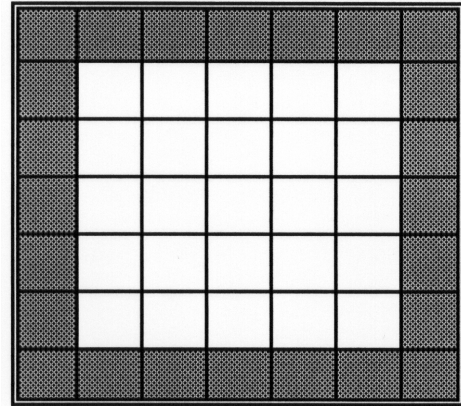
the case, with squatter settlements surrounding the city. In Figure 2.10, the case of high centralization, as is common in the US, is shown. The opposite case of low centraliztion, as in developing countries, is shown in Figure 2.11. Note the decrease in density as the area occupied is increased.

The absolute centralization index (ACE) derived from Duncan and Duncan's relative concentration index ([91]) was chosen as the representative statistic for this dimension. In computing this index, the tracts are ordered in increasing distance from the city center (CBD). The zeroth tract, then, is the most centrally located. This measure gives a group's absolute distribution in urban space; that is, a group's spatial distribution compared to the distribution of land area around the city center. The index ranges from negative one to positive one. Positive values indicate a more centralized distribution of group members while negative values indicate a more suburban distribution. Zero indicates a uniform distribution throughout the urban space.

The final dimension, clustering, refers to the extent to which minority areal units adjoin one another. If a high degree of clustering is evident, then is it highly probable the minority population is concentrated in a single ethnic enclave. A low level of clustering would indicate multiple minority tracts, or neighborhoods, throughout the MSA. It is related to evenness and concentration, and often associated with centralization and concentration, although distinct conceptually ([2]). An urban area with a single ethnic enclave would be considered more segregated than one with the same minority population scattered across several tracts, even if no minority member shared a common residential area with a majority population. In Figure 2.12 the case of high clustering is shown, with two ethnic enclaves. In Figure 2.13, the case of low clustering is shown with the same minority population scattered across the MSA in various tracts.

26
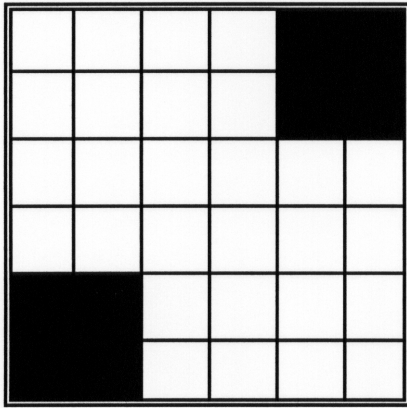
Figure 2.12: High Clustering
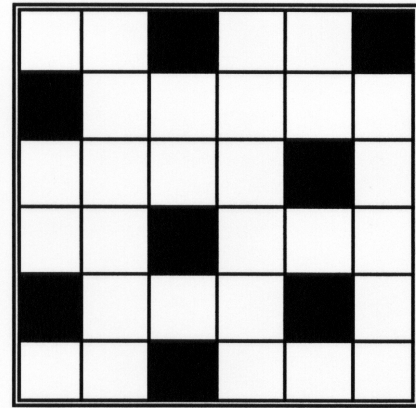
Figure 2.13: Low Clustering

White's index of spatial proximity (SP) ([74]) was selected as this dimension's representative statistic. The statistic equals one (1.0) if no clustering is evident and greater than one when members of the same group tend to live together. A value less than one indicates a form of segregation wherein members of a group live closer to members of the other group than to their own ([74]), which is considered highly unusual.

For the first time, a spatial component has been added to an index ($f(d_{ij})$). This is the function which measures how social interaction changes with distance. This function is not explicitly defined, however, and must be derived from urban sociological theory. There is no absolute definition. In his article, White examined four possible choices for $f(d_{ij})$: $\exp(-d_{ij})$, d_{ij}, $\frac{1}{d_{ij}^2}$, and $\exp(-2d_{ij})$. White chose $\exp(-d_{ij})$ as the "best" approximation for the function ([74]).

2.4 Summary

Segregation is a complex and persistent urban phenomenon whose roots go back to the forced immigration of African Americans in the colonial period. Though they were liberated from slavery and declared free and equal, African Americans have had to fight for the rights granted to other ethnic minorities. White society made many efforts to keep a wall between the races using such tactics as racial zoning, racial covenants, and steering by real estate agents. Policies which were never intended for separating the races, such as redlining, were used to further denigrate the African American. When the courts struck down these legalistic tactics, the whites simply left the cities for the suburbs.

This refusal of the whites to live next to African Americans has had a negative effect on society. African Americans are unfairly stereotyped by the media, and the lack of interaction between the races only perpetuates these stereotypes. African Americans live in conditions which are well below societal norms, and when they succeed in leaving the ghetto, often the only places they can go are extensions of the ghetto. This forced isolation feeds a growing sense of anomie and separatism on the part of the African Americans.

Such a persistent and pervasive a phenomenon has attracted the attention of social scientists. Attempts to adequately measure segregation began in the 1930s and continue until the present time. Over the past sixty years, many indices have been proposed to measure segregation, but none have been received without reservation. Each new index has claimed to measure segregation better than other indices, but each has focused on a single aspect of segregation independent of all others. In the past thirty-one years, the concept of segregation as a multidimensional phenomenon has gained wider acceptance and the new indices proposed now make claim to multidimensionality though the claims have yet to be verified.

Chapter 3

RESEARCH DESIGN AND METHODOLOGY

3.1 Introduction

The purpose of this study is to examine and expand the methodology used by Massey and Denton in studying residential segregation. Massey and Denton used first order factor analysis to analyze common indices of segregation and to group them into dimensions. The same methodology will be applied using different data to confirm the results.

3.2 Description of Data

Massey and Denton used a subset of data from the 1980 Census. Massey and Denton showed the results were consistent for the 1970 Census and several different groupings of the 1980 Census data. As a further test of their results, their methodology will be applied to the complete 1980 Census data.[1] Since some of the variations in Massey and Denton findings may have resulted from using only 60 MSAs and including multiple types of segregation (white vs. black, white vs. Latin, white vs. Asian), an effort will be made to reduce this variation by including data from all MSAs identified in the 1980 Census and using only a single type (white vs. black) of segregation in the analysis. If the methodology developed by Massey and Denton is robust, then the same dimensions should be evident in the 1990 and 2000 data.

Upon examination of the data files it was noted there were no values for People in Central City (PCC). This reduced the number of centralization indices to two. This was not deemed a significant loss, as PCC is defined as "the number of people in a given group that live within the bounds of the central city, expressed

---

[1]The data for this research were obtained from the 1980, 1990, and 2000 Censuses of Population and Housing conducted by the U.S. Bureau of the Census. The indices were computed by the Bureau and are available from `http://www.census.gov/hhes/www/housing/housing_patterns/housing_patterns.html`.

Index	Mean	Std Dev	Minimum	Maximum
Atkinson Index at 0.01	0.16289	0.08375	0.003	0.458
Atkinson Index at 0.05	0.50784	0.18977	0.017	0.921
Atkinson Index at 0.09	0.69222	0.21188	0.029	0.988
Absolute Centralization Index	0.69888	0.2241	-0.308	0.969
Absolute Clustering Index	0.16386	0.16188	-0.016	0.671
Absolute Concentration Index	0.88815	0.12784	0.284	0.997
Correlation Index	0.29073	0.22005	0	0.813
Dissimilarity Index	0.58694	0.14585	0.123	0.908
Distance Decay Interaction	0.24749	0.19647	0.001	0.719
Distance Decay Isolation	0.75192	0.19634	0.281	0.999
Delta	0.80935	0.12102	0.165	0.974
Entropy Index	0.33005	0.1798	0.005	0.782
Gini Index	0.73072	0.14736	0.123	0.974
Isolation Index	0.35027	0.24669	0.001	0.855
Interaction Index	0.64973	0.24669	0.145	0.999
Relative Centralization Index	0.22627	0.23703	-0.535	0.739
Relative Concentration Index	0.52522	0.45106	-3.108	0.965
Relative Clustering Index	1.35638	2.50348	-0.657	29.362
Spatial Proximity	1.16167	0.16936	1	2.054

Table 3.1: Simple statistics for Segregation Indices

as a proportion of the total number in the entire metropolitan area" ([2, p. 292]). Massey and Denton recognized PCC as not the best measure and listed several drawbacks, among those being

- the artificiality of central city boundaries (political, not natural);

- the differences in city structure due to age and growth patterns; and

- the insensitivity of the measure to a groups distribution in space.

Simple univariate statistics for each index are in Table 3.1

3.3 Analytical Design

The goal will be to revalidate the methodology proposed by Massey and Denton. To prove the methodology is robust it is necessary to be able to reproduce the results using a different data set. To accomplish this, the data for all MSAs in the 1980 Census will be used to confirm the 1980 results from Massey and Denton while the 1990 and 2000 Census data will be used to derive more current results. The hypothesis to be tested, in two parts, is therefore:

Following the methodology established by Massey and Denton, the same dimensions can be derived from (1) a more comprehensive data set and (2) more current data sets.

3.3.1   Discussion

The analysis follows that of Massey and Denton ([2]). Briefly, they computed values for all twenty segregation statistics (Appendix I) for the three groupings (white vs. black, white vs. Latin, white vs. Asian) which yielded 180 values (60 MSAs x 3 groupings). Two sets of simple univariate statistics (mean and variance) were calculated for these values, one set weighted by minority population, the other set unweighted. A matrix of intercorrelation values was constructed and analyzed for *a priori* groupings of statistics into dimensions. The correlations of the weighted statistics were then factor analyzed to identify the underlying dimensions in the data[2]. Principal axis factoring was used to extract the initial factor loadings, or the initial solution. Following the extraction, a varimax (orthogonal) rotation was performed to rotate the initial factors so that they would better explain the variance of the data space. In order to improve the significance of the concentration, centralization, and clustering factors, a second rotation was performed. The method used was promax, an orthogonal/oblique method, which allows the factors to be correlated. The independence constraint of the varimax rotation resulted in a very low percentage of explained variance for these factors (0.11, 0.10, 0.07 respectively). The promax rotation provides both an orthogonal (independent factors) and an oblique rotation (correlated factors). This ensures that orthogonality of the final results is due to the data and not to the rotation method used. The promax rotation relaxed the independence constraint and increased the explained variance to 0.32, 0.22, 0.27 respectively. Conclusions were made using factor loadings from the resulting pattern matrix. From each of the five resulting factors, a single variable was selected to represent each dimension of segregation. In their results, Massey and Denton noted the first two factors (evenness and exposure) were positively correlated. Further, these two dimensions explained a large part of the variance while the remaining three (concentration, centrality, and clustering) were empirically insignificant.

Massey and Dentons methodology will be followed to test the first hypothesis. The test for rejection of the hypothesis will be the replication of the factors. If the same factors are extracted, then the hypothesis is confirmed. There is some deviance from the exact methodology, however. The data are of a single type of segregation (white vs. African American) and the values are unweighted, as Massey and Denton found no

---

[2]For a brief explanation of factor analysis and related concepts, see Kime and Mueller[92], Kim and Mueller[93], or Rummel[94]

difference between the weighted and unweighted statistics in their analysis. To determine if a variable loads on a factor, it must have a factor loading of greater than 0.40 for this research[3].

Using the results from the first part of the analysis, data sets from the 1990 and 2000 Census will be used to attempt to replicate the results with more current data. Ideally the factors derived from the analysis will appear in the later data sets thereby providing a method of tracking changes in segregation in cities over time.

3.4 Rationale for Factor Analysis

Throughout this research, use is made of the technique of factor analysis. Factor analysis is a technique derived from a combination of psychology and mathematics and has been called the "calculus of social science" ([94, p. 4]). The fundamental principle behind factor analysis is, "...that some underlying factors, which are smaller in number than the number of observed variables, are responsible for the covariation among the observed variables" ([92, p. 12]). The goal is to reduce the complexity of a problem to a more manageable size.

Factor analysis is applicable to the problem of explaining the dimensionality of segregation because it, "...permits the distillation of the variance within a large set of variables to a small set of underlying patterns" ([95, p. 40]). In this case, the distillation of a set of 19 indices of segregation to a more manageable set of only five dimensions. These factors may explain all the variance among the variables with no interaction among themselves, or they may be correlated, as appears in Massey and Denton's results. Uncorrelated factors indicate a simple structure composed of distinct and independent dimensions. Correlated factors indicate a simple structure composed of distinct dimensions that interact with each other. It has been argued that an intercorrelated structure is more representative of the real world than an uncorrelated structure ([94, p. 388]).

To explain interfactor correlations requires second order, or hierarchical, factor analysis. Second order factor analysis takes the factor intercorrelation matrix from the initial analysis and runs the same procedure again. The hope is that by doing this a further reduction in the number of factors may be obtained([96]). This reduction explains the interaction of the factors and highlights structures not readily apparent in the first

---

[3]This is a rule of thumb only. Other researchers have used values as low as 0.30 and as high as 0.50. Variables which load at zero or near-zero on a factor are considered unrelated to the factor, but at what point a variable becomes "significant" to the factor is a decision left to the individual researcher. For more on this see Rummel[94, pp. 472-489]

order solution. A good description of the application of second order factor analysis in the social sciences is given by Rummel. Many social theories, he states, have mental, or intuitive, higher order factors which relate seemingly disparate concepts together in a global domain. Marx, for example, explains the social and political dimensions of society by use of a second order factor, the economy ([94], p. 427).

<center>Chapter 4</center>

<center>**RESULTS OF ANALYSIS**</center>

4.1 Introduction

This chapter discusses the results obtained from the analysis of the data. Each research question and hypothesis is presented along with the relevant results. Each hypothesis is then either confirmed or refuted.

4.2 Revalidation of Methodology

4.2.1 Purpose

The first goal of the research is to revalidate the methodology used by Massey and Denton. A methodology which is tied to a specific dataset is of limited use. Replication of results using a different data set would prove the robustness of the methodology. Massey and Denton used a sample of 60 MSAs from the 1980 Census and computed index values for blacks, Hispanics, and Asians, resulting in a heterogeneous data set. For this research data from all MSAs identified in the 1980 Census ($n = 330$) will be used and only index values for blacks. Using the same indices as Massey and Denton (Table 4.1)[1], less People in Central City, the results of this revalidation should confirm those of the original study (Table 4.2). Further validation is accomplished by replicating their results using 1990 and 2000 Census data. The hypothesis to be tested is:

> *Following the methodology established by Massey and Denton, the same dimensions can be derived from a more comprehensive data set.*

---

[1]The computational formulae and relevant citations for all 19 indices are located in the Appendix.

Evenness	Concentration
Index of Dissimilarity (D)	Delta
Gini Coefficient	Absolute Concentration Index (ACO)
Entropy Index	Relative Concentration Index (RCO)
Atkinson Index ($b = 0.10$)	
Atkinson Index ($b = 0.50$)	**Centralization**
Atkinson Index ($b = 0.90$)	Relative Centralization Index (RCE)
	Absolute Centralization Index (ACE)

Exposure	Clustering
Bell's Isolation Index (xPx)	Absolute Clustering Index (ACL)
Bell's Interaction Index (xPy)	Relative Clustering Index (RCL)
Correlation Ratio (Eta)	Distance Decay Interaction (DPxy)
	Distance Decay Isolation (DPxx)
	Spatial Proximity (SP)

Table 4.1: Segregation Indices Used in Analysis

	Evenness	Exposure	Concentration	Centralization	Clustering
D	0.84	0.17	0.06	0.00	0.04
GINI	0.91	0.12	0.06	-0.02	-0.03
ENTROPY	0.81	0.18	0.03	0.00	0.09
ATKIN1	0.78	-0.07	-0.03	0.05	0.30
ATKIN5	0.93	0.05	0.02	0.01	0.06
ATKIN9	0.99	0.08	-0.04	0.03	-0.10
PXY	-0.30	-0.82	0.01	0.02	0.05
PXX	0.30	0.82	-0.01	-0.02	-0.05
ETA	0.71	0.36	0.02	-0.02	0.06
DEL	0.26	-0.04	0.27	0.48	0.05
ACO	0.30	-0.52	0.62	-0.07	0.11
RCO	0.03	0.19	0.86	-0.03	-0.04
PCC	0.10	0.35	0.04	0.37	-0.16
ACE	0.00	-0.06	-0.11	0.83	0.08
RCE	0.25	0.25	0.32	0.26	0.03
ACL	0.08	0.78	0.03	-0.07	0.34
SP	0.10	0.55	0.04	0.05	0.57
RCL	0.34	-0.20	0.04	0.04	0.60
DPXY	0.06	-1.04	0.00	0.00	0.05
DPXX	-0.06	1.04	0.00	0.00	-0.05

Table 4.2: Original Results from Massey and Denton

4.2.2 First Attempt

A first order factor analysis following the methodology as outlined by Massey and Denton was performed. The factor extraction method was set to principal component analysis. Initial estimates for communalities were set to squared multiple correlations. Squared multiple correlations (SMC) were used as initial communality estimates as this method takes into account all the interrelationships between variables and is considered a more stable approach ([94]; [97]). The communality of a variable (X) is that part of the variance of X shared in common with the other $n - 1$ variables in the analysis ([94]). SMC are not an exact replacement for the communalities of the variables, but they are related to the communalities and are often only slightly below the theoretical true communalities ([98]). The number of factors was not specified nor was a retention criterion for factors defined.

The results of the initial attempt identified a problem. The correlation matrix used in the procedure was singular, meaning it was not invertible[2]. A singular matrix indicates the rows and columns are not independent. That is, one or more rows are linear combinations of others ([94]; [99]). SAS printed an error message and reset all communality estimates to 1.0.

Upon investigation, it seemed likely this flaw resulted from two sets of indices. Two pairs of the indices, the Bell interaction/isolation indices (P_{xx}, P_{xy}) and the distance modified Bell indices (D_xP_x, D_xP_y), had perfect, or near perfect, negative correlations (-1.00). The perfect correlations arose from the mathematical relationship of the index pairs. Each index in a pair is the other's inverse. The standardized scoring coeffiecient for P_{xx} is set to zero in the results, implicating this index as the source of the singularity.

To rectify the matrix singularity problem, the Isolation index (P_{xy}) was removed. Following the same logic the distance modified Isolation index (D_xP_y) was also removed. It would be equally appropriate to remove the Interaction. The analysis was repeated with the reduced set of indices.

4.2.3 Second Attempt

This time the procedure completed successfully. Six factors were retained based on the cumulative proportion criterion. The results are shown in Table 4.3.

Examination of the rotated factor pattern showed only three definitive factors (1, 2, and 5) and two singleton factors (3 and 4) (Table 4.4). Factor one seems to be a "super factor" with ten indices loading

---

[2]The complete correlation matrix is in the Appendix.

36

Factor	Eigenvalue	Difference	Proportion	Cumulative
1	9.21392985	5.69910929	0.6044	0.6044
2	3.51482056	2.34335310	0.2306	0.8350
3	1.17146746	0.47639663	0.0768	0.9119
4	0.69507084	0.34751542	0.0456	0.9575
5	0.34755541	0.04125705	0.0228	0.9803
6	0.30629836	0.17677813	0.0201	1.0004

Table 4.3: Cumulative proportion of explained variance from first pass

Index	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Pxx	0.9730	-0.0944	0.0702	-0.0686	0.0434	0.1439
Eta	0.9623	0.0554	0.1487	0.0021	0.0602	0.1598
DPxx	0.9570	-0.2139	0.0248	-0.0791	0.0093	-0.0119
ACL	0.9254	-0.0843	0.1380	0.1953	0.0542	-0.1692
SP	0.8885	-0.0223	0.1350	0.3151	0.0688	-0.2000
Entropy	0.8849	0.1732	0.2918	0.1132	0.0810	0.2606
Atkin9	0.7461	0.2274	0.3382	0.1323	0.0854	0.4792
Atkin5	0.7279	0.2631	0.4972	0.1728	0.0999	0.3372
D	0.7078	0.2916	0.4483	0.2116	0.0815	0.3747
Gini	0.6996	0.2915	0.4241	0.1753	0.0846	0.4369
ACO	-0.1107	0.9220	0.1043	-0.0832	0.1964	0.0330
DEL	0.0466	0.8554	0.1516	0.1408	0.2564	0.0899
RCO	0.0425	0.7136	0.1720	-0.4606	0.2875	0.0857
Atkin1	0.4010	0.2454	0.8101	0.1809	0.1307	0.0469
RCL	0.1560	-0.0632	0.1673	0.8195	-0.0137	0.0641
RCE	0.1409	0.3157	0.1675	-0.0680	0.6462	0.0545
ACE	0.0156	0.4703	-0.0457	0.0302	0.6415	-0.0187

Table 4.4: Initial rotated factor pattern extraction

indices since its only significant loading was on factor four. The Atkinson index was retained as it loaded significantly on two factors (1 and 3) and might load on a single factor in the next iteration.

4.2.4 Third Attempt

Again, six factors were retained based upon the cumulative proportion criterion (Table 4.5).

Factor	Eigenvalue	Difference	Proportion	Cumulative
1	9.12029099	5.63964132	0.6315	0.6315
2	3.48064967	2.63548044	0.2410	0.8725
3	0.84516923	0.37494524	0.0585	0.9310
4	0.47022399	0.14412395	0.0326	0.9636
5	0.32610004	0.07428580	0.0226	0.9862
6	0.25181425	0.15343452	0.0174	1.0036

Table 4.5: Cumulative proportion of explained variance from second pass

Better structure emerged from this pass (Table 4.6). The "super-factor" (Factor one) has broken into two factors. Factor one now contains indices from the Clustering and Exposure groups along with a single Evenness index (Entropy). Factor two contains only Evenness indices and it is noted that Atkin1 did increase its loading to be significant. There is noticeable crossloading between factors one and two for the Gini, Atkin9, D, and Atkin5 indices but the loadings on factor one are markedly lower than on factor two. The cross loading of Atkin1 is more problematic since the loadings are still relatively close (0.6779 vs. 0.5991). More improvement could be accomplished, possibly, by reducing the number of factors extracted.

4.2.5 Fourth Attempt

In the next iteration, the number of factors was specified as five. With the number of factors predetermined the cumulative proportion criterion is no longer in effect. The results from the rotation are shown in Table 4.7.

Looking at the rotated factor pattern it can be seen that Atkin1 is now loading strongly on factor two (0.8259). It no longers loads on any other factor significantly. Further examination of the factor pattern reveals that no other loading on factor five was significant. This indicates the factor is not a true common factor and should be eliminated ([97]). The next logical step is to reduce the number of factors extracted to four and run the analysis again.

Index	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
DPxx	0.9118	0.2733	-0.1900	0.0039	-0.0410	0.1105
ACL	0.8978	0.2921	-0.0997	0.0487	0.1823	-0.1376
Pxx	0.8758	0.4231	-0.0823	0.0341	-0.0910	0.1623
SP	0.8648	0.2914	-0.0551	0.0614	0.2149	-0.2193
Eta	0.8415	0.5077	0.0501	0.0492	-0.0282	0.1065
Entropy	0.7101	0.6735	0.1340	0.0750	0.0413	0.0476
Gini	0.4510	0.8531	0.2100	0.0950	0.0444	-0.0189
Atkin9	0.5014	0.8309	0.1553	0.0945	-0.0547	0.0290
D	0.4747	0.8281	0.2126	0.0871	0.1103	-0.0308
Atkin5	0.5003	0.8200	0.1957	0.1037	0.1654	0.0142
Atkin1	0.2274	0.6779	0.1935	0.1357	0.5991	-0.0001
ACO	-0.1601	0.1261	0.9233	0.1716	0.0473	-0.0151
DEL	-0.0475	0.3052	0.7985	0.2543	0.0521	-0.2410
RCO	-0.0196	0.1320	0.7759	0.2589	0.0160	0.3575
ACE	0.0133	0.0244	0.4684	0.6364	-0.0331	-0.1441
RCE	0.1021	0.1673	0.3445	0.6276	0.0923	0.1350

Table 4.6: Rotated factor pattern from reduced data set

Index	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
DPxx	0.9330	0.1967	-0.1921	0.0337	-0.0746
Pxx	0.9189	0.3222	-0.0751	0.0447	-0.1786
ACL	0.8914	0.2912	-0.1026	0.0626	0.2455
Eta	0.8848	0.4248	0.0579	0.0429	-0.1174
SP	0.8542	0.3036	-0.0547	0.0618	0.3241
Entropy	0.7624	0.6109	0.1480	0.0510	-0.0790
Atkin1	0.2138	0.8259	0.1713	0.1816	0.2534
Atkin5	0.5538	0.8009	0.2103	0.0772	-0.0301
Gini	0.5241	0.7944	0.2384	0.0391	-0.0934
D	0.5362	0.7918	0.2334	0.0420	-0.0354
Atkin9	0.5850	0.7376	0.1894	0.0373	-0.1815
ACO	-0.1537	0.1414	0.9340	0.0737	0.0171
DEL	-0.0243	0.3070	0.8392	0.1207	0.1583
RCO	-0.0080	0.1262	0.7787	0.2315	-0.2673
ACE	0.0250	0.0082	0.5543	0.5384	0.0989
RCE	0.1113	0.1782	0.4062	0.6039	-0.0549

Table 4.7: Rotated factor pattern extraction from five factor model

4.2.6 Fifth Attempt

A similiar structure emerged, but changes were noted (Table 4.8). Factors one and two were still evident but factor four had collapsed into factor three. Factor four now has no significant loadings which by the previously used argument means it is not a true factor.

At this point it must be considered that the true dimensions are far fewer than originally thought. Massey and Denton found five dimensions (factors), but at this point the number of dimensions (factors) that can be accepted is now three. Could the true number of dimensions be fewer?

Support for this hypothesis can be found in the table of eigenvalues and proportion of explained common variance (Table 4.9). To this point the number of factors retained for the rotation step has been determined either by the cumulative proportion of variance being greater than one or explicitly specified. Another approach to identifying the number of factors to retain is the minimum eigenvalue criterion. A factor with an eigenvalue of less than one contributes little to the analysis and is usually discarded. Plotting eigenvalue by factor number produces a scree plot showing the contribution of each factor. Looking at the scree plot (Figure 4.1) for the model with Isolation index (P_{xy}) and Distance modified Isolation index ($D_x P_y$) removed it can be seen that the amount of variance being explained by the factors decreases markedly after the third

Index	Factor 1	Factor 2	Factor 3	Factor 4
DPxx	0.9479	0.1786	-0.1476	-0.0223
Pxx	0.9358	0.3023	-0.0409	-0.1324
Eta	0.8885	0.4186	0.0764	-0.1068
ACL	0.8834	0.2967	-0.0511	0.2535
SP	0.8373	0.3183	-0.0083	0.3122
Entropy	0.7633	0.6086	0.1588	-0.0830
Atkin1	0.2237	0.8159	0.2385	0.2758
Atkin5	0.5580	0.7965	0.2243	-0.0339
D	0.5335	0.7947	0.2283	-0.0585
Gini	0.5249	0.7938	0.2304	-0.1114
Atkin9	0.5936	0.7287	0.1857	-0.1810
ACO	-0.2192	0.2072	0.8607	-0.1629
DEL	-0.0829	0.3655	0.8010	0.0041
RCO	-0.0194	0.1347	0.7932	-0.3144
ACE	0.0499	-0.0239	0.7417	0.1833
RCE	0.1721	0.1092	0.6387	0.1125

Table 4.8: Rotated factor pattern extraction from four factor model
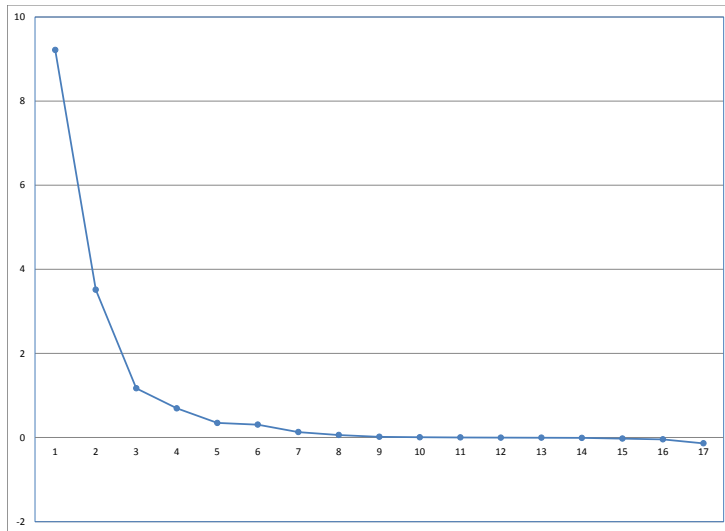
Figure 4.1: Scree plot for 17 variable original model

Factor	Eigenvalue	Difference	Proportion	Cumulative
1	9.21392985	5.69910929	0.6044	0.6044
2	3.51482056	2.34335310	0.2306	0.8350
3	1.17146746	0.47639663	0.0768	0.9119
4	0.69507084	0.34751542	0.0456	0.9575
5	0.34755541	0.04125705	0.0228	0.9803
6	0.30629836	0.17677813	0.0201	1.0004
7	0.12952023	0.06918536	0.0085	1.0089
8	0.06033487	0.04362236	0.0040	1.0128
9	0.01671251	0.00969026	0.0011	1.0139
10	0.00702225	0.00484355	0.0005	1.0144
11	0.00217870	0.00460576	0.0001	1.0145
12	-.00242705	0.00111987	-0.0002	1.0144
13	-.00354692	0.00496808	-0.0002	1.0141
14	-.00851500	0.01676743	-0.0006	1.0136
15	-.02528242	0.01780672	-0.0017	1.0119
16	-.04308914	0.09530864	-0.0028	1.0091
17	-.13839778		-0.0091	1.0000

Table 4.9: Cumulative proportion of explained variance from 17 variable model

factor. If a line is drawn at one on the vertical axis (the minimum eigenvalue) it shows that only three factors are worth retaining.

To test this another analysis is done, this time setting the option MINEIGEN to be one (1.0). The varimax rotation is still done on the initial extraction to clarify the structure. The results of this are shown in Table 4.10.

Index	Factor 1	Factor 2	Factor 3
Pxx	0.9898	-0.0395	-0.0395
Eta	0.9800	0.0961	0.0767
DPxx	0.9575	-0.1832	-0.1154
Entropy	0.9226	0.2242	0.2767
ACL	0.9055	-0.1156	0.1411
SP	0.8561	-0.0877	0.2368
Atkin9	0.8119	0.3122	0.3763
Atkin5	0.7983	0.3504	0.4615
D	0.7705	0.3545	0.4823
Gini	0.7678	0.3683	0.4545
ACO	-0.1306	0.9068	0.0771
RCO	0.0737	0.8703	-0.2218
DEL	0.0311	0.8395	0.2771
ACE	0.0139	0.6511	-0.0192
RCE	0.1897	0.5933	0.0042
RCL	0.1403	-0.2034	0.7561
Atkin1	0.4948	0.3753	0.5698

Table 4.10: Rotated factor pattern extraction from MINEIGEN=1

From this it can be seen that the "super-factor" has returned. Factor one contains indices from Evenness, Exposure, and Clustering. Factor two contains indices from Concentration and Centralization. Factor three contains the two problematic indices, RCL and Atkin1. Again it is seen that Atkin1 loads on two factors (one and three) while RCL only loads on factor three. The decision is again to discard RCL and rerun the analysis with the hope that Atkin1 will load on one of the other two factors.

4.2.7 Sixth Attempt

The eigenvalue table now shows, clearly, that only two factors should be retained for the rotation (only the first three rows are shown in Table 4.11). Looking at the rotated factor pattern shows that two factors are

Factor	Eigenvalue	Difference	Proportion	Cumulative
1	9.12029099	5.63964132	0.6315	0.6315
2	3.48064967	2.63548044	0.2410	0.8725
3	0.84516923	0.37494524	0.0585	0.9310

Table 4.11: Cumulative proportion of explained variance from 17 variable model

clearly identified (Table 4.12). The first factor contains all indices for Evenness, Exposure, and Clustering. The second factor contains all indices for Concentration and Centralization. Since the analysis has reached a stopping point, the task at hand is now to name the factors which have emerged.

Index	Factor 1	Factor 2
Eta	0.9684	0.0446
Entropy	0.9662	0.2112
Pxx	0.9465	-0.1110
ACL	0.9069	-0.1438
Atkin5	0.8968	0.3762
DPxx	0.8935	-0.2638
Atkin9	0.8876	0.3219
SP	0.8804	-0.0927
D	0.8748	0.3862
Gini	0.8661	0.3941
Atkin1	0.6333	0.4391
ACO	-0.0935	0.9135
DEL	0.1117	0.8739
RCO	0.0326	0.7978
ACE	0.0141	0.6383
RCE	0.1903	0.5732

Table 4.12: Rotated factor pattern extraction from reduced MINEIGEN=1 model

Factor one contains indices which measure exposure of persons of different race as well as those which measure the interaction of persons in a given area. It seems most appropriate to label this factor **Interaction**, as all the indices which comprise this factor deal with interaction between races or neighborhood composition (racial homogeneity). This factor closely resembles the first dimension proposed by Yinger ([18]) which he called "degree."

The second factor consists of indices from Concentration and Centrality. These indices measure the distribution of the minority ethnic group in the urban area, so it therefore seems logical to label this second order factor **Agglomeration**. This factor matches the definition of Yinger's other dimension, "pattern."

4.3 Application of Methodology

The next step in the analysis is to apply the methodology previously derived and apply it to more current data. The data here are the same indices but computed for the 1990 and 2000 Census. The goal here is to reproduce the same two factors from different data sets and thereby validate this as a useful methodology for studying change over time.

The data were analyzed as proposed and the results are shown in Table 4.13. The same two factors emerged in both instances. The order of the variables which loaded on the factors did change, but the constituent indices for each factor did not change. These results confirm the existence of the two factors as well as validate the new methodology.

Results : 1990				Results : 2000		
Index	Factor 1	Factor 2		Index	Factor 1	Factor 2
Entropy	0.9728	0.1977		Eta	0.9816	0.0424
Eta	0.9719	0.0383		Entropy	0.9802	0.1699
Pxx	0.9387	-0.1488		Atkin5	0.9487	0.2532
Atkin5	0.9187	0.3322		Atkin9	0.9407	0.1875
Atkin9	0.9103	0.2591		Pxx	0.9406	-0.1634
ACL	0.8963	-0.1408		D	0.925	0.2609
D	0.8944	0.3427		Gini	0.9223	0.2451
Gini	0.893	0.333		Atkin1	0.9221	0.2975
SP	0.8848	-0.0916		SP	0.9117	-0.0209
DPxx	0.8829	-0.2884		ACL	0.8999	-0.0577
Atkin1	0.7822	0.4054		DPxx	0.8936	-0.2638
ACO	-0.1235	0.9104		ACO	-0.1512	0.8985
DEL	0.0652	0.8585		DEL	0.0256	0.8346
RCO	0.1119	0.7655		RCO	0.2085	0.7956
ACE	-0.012	0.6591		ACE	-0.0176	0.6484
RCE	0.2488	0.5783		RCE	0.3117	0.5498

Table 4.13: Rotated factor pattern extraction from 2000 data

A test of the usefulness of the factors derived from a factor analysis is the measure of variance explained by each factor. For each of the three years these values were calculated by the software and are reported in Table 4.14. The total variance to be explained is calculated as 1.0 times the number of variables being used in the analysis. In this case, we had 16 indices so the total variance to be explained is 16. The percentage

explained in each year was at least 78% of the total and the percentage increased each year. This is further proof of the "goodness" of the factors that were derived.

Year	Factor 1	Factor 2	Total	Percent
1980	8.7302	3.8707	12.6009	78.76
1990	9.1137	3.6648	12.7785	79.87
2000	9.7545	3.3062	13.0607	81.63

Table 4.14: Variance Explained by Each Factor

4.4 Summary

Replicating Massey and Denton's original methodology presents several challenges. Since the selected cities in the subset were not listed, the data set used by Massey and Denton was irreproducible. This was corrected by using the entire 1980 Census. Methodological flaws were corrected by making better beginning assumptions regarding the communalities and removing indices which were causing matrix singularity. Using the amended methodology, two underlying factors were identified. These two factors combine the dimensions of Evenness, Exposure, and Clustering into one factor (**Interaction**) and Centrality and Concentration into a second factor (**Agglomeration**).

Chapter 5

CONCLUSIONS

5.1   Restatement of Purpose

The purpose of this research was to investigate in detail the multidimensional aspects of residential segrega-
tion focusing on African-American segregation as it is the most pervasive form. Following the methodology
established by Massey and Denton, an attempt was made to reproduce the five factor model they derived and
then to simlify their results by using second order factor analysis. In this chapter the findings are summarized
and the implications examined. Recommendations for future research in this area are also presented.

5.2   Summary of Findings

As the starting point for this research the premise of segregation as a multidimensional phenomenon was
accepted. The first stage of this research was devoted to a replication of Massey and Denton's 1988 study.
The data used for this research contains 330 observations, one for each MSA defined at the time of the
Census, and is homogeneous in type (white-black only). It was not possible to recreate the same data set
used by Massey and Denton since they did not list all 60 of the MSAs used in their analysis. The top fifty
in size were used, which are easily identifiable from the 1980 Census, but in addition, ten were chosen at
random. Not knowing these ten it was impossible to recreate the sample. Inclusion of more data points was
also motivated by a desire to increase the adequacy of the sample. The rule of thumb for sampling adequacy
is ten observations for each variable. In Massey and Denton's analysis there were only 180 observations for
twenty indices (9:1) while in this analysis the ratio was nearly doubled (17:1). This increase in sample size
improves the initial correlation matrix and the subsequent analysis.

When the initial attempt at replication was made a problem was discovered. In the initial correlation
matrix there were two pairs of perfect, or near perfect, inverse correlations (-1.00). These two pairs of indices

were the Bell's Isolation and Interaction indices and the Distance Decay Isolation and Interaction indices. The inverse correlation results from the fact that the Isolation index is a linear combination of the Interaction index, and vice versa. These perfect correlations caused an error in the factor analysis procedure due to a singular correlation matrix. After reviewing Massey and Denton's methodology, it was discovered they had patched this problem by using an estimator for the initial communalties (Maximal Absolute Correlations) which allowed the procedure to complete.

The removal of redundant entries in the correlation matrix (xPy, D_xP_y) and the subsequent use of squared multiple correlation (SMC) instead of maximal absolute correlations for the communalities also altered the end results of the analysis. These modifications are improvements to the original methodology that make it more conservative.

The results from this part of the research confirmed the multidimensional hypothesis, but differ from those of Massey and Denton, most likely due to methodological corrections. Where Massey and Denton found five dimensions, the results from this analysis only support the existence of two. Modifications to the methodology originally proposed by Massey and Denton make it a viable methodology for the study of residential segregation.

The two dimensions (**Interaction** and **Agglomeration**) closely parallel the two dimensions mentioned by Yinger ([18]). These two dimensions make the problem of segregation simpler to understand and easier to measure. These results are consistent with the results of Massey and Denton. In the original research, they found a high degree of correlation among the factors for Evenness and Exposure, and a similar degree of correlation among the factors for Centrality, Clustering, and Concentration.

5.3   Implications of Findings

The dimensions of segregation can be known quantitatively. Using this knowledge, it should be possible to construct a scale from the two dimensions which would constitute a more comprehensive measure of segregation. A summative scale could be developed from the factor scores derived from these that would measure the phenomenon of segregation in its multidimensional form. The greatest difficulty in deriving this scale would be in objectively assigning weights to the dimensions.

It must also be considered that the previous models of four and five dimensions may be overly complex. It seems that the phenomenon of segregation may be explainable in only two dimensions. While this is may

seem to be overly simplifying the problem, this also makes understanding possible. Analyzing inherently qualitative phenomena is difficult. Abstraction makes the real world manageable and analyzable. A concept such as segregation cannot easily be measured using quantitative means. Since it involves people, it becomes a humanistic phenomenon. As such, it can be studied, and theories derived from observation for explanatory purposes, but there is no single positivist approach to the problem. It is only through statistical methods such as factor analysis that phenomena such as segregation can be measured at all.

5.4 Recommendations for Future Research

This research focused exclusively on African-American residential segregation. The data exist for other ethnic groups (Asian, Black Hispanic, White Hispanic), so further analysis is neither impossible nor difficult. Examination of the phenomenon of residential segregation among other ethnic groups is called for, especially in light of the growth in the Asian and Hispanic sectors. An analysis of all ethnic groups combined would give an insight to the state of residential segregation in the United States. Either of these analysis paths might produce findings which differ from those of this research, indicating a variance among treatment accorded to different ethnic groups in different parts of the country.

An avenue to explore would be to expand the set of indices. Measures which concentrate on socioeconomic status would be one addition. More spatial measures, such as measuring the number of minorities within a defined radius from the center of the MSA, would provide a more geographic, rather than sociologic, explanation of the phenomenon. Alternatively, the indices could be kept and the method of analysis changed. Factor analysis has many critics. Other structure seeking methods, such as cluster analysis or multidimensional scaling, could be applied to the data. The results may be more satisfying than those derived from factor analysis.

REFERENCES

[1] O. D. Duncan and B. Duncan, "A methodological analysis of segregation indexes," *American Sociological Review*, vol. 20, pp. 210–217, 1955.

[2] D. S. Massey and N. A. Denton, "The dimensions of residential segregation," *Social Forces*, vol. 67, no. 3, pp. 281–315, 1988.

[3] ——, "Hypersegregation in U. S. Metropolitan areas: Black and Hispanic segregation along five dimensions," *Demography*, vol. 26, no. 3, pp. 373–391, 1989.

[4] C. Winship, "A revaluation of indexes of residential segregation," *Social Forces*, vol. 55, no. 4, pp. 1058–1066, 1977.

[5] L. H. Larsen, *The Rise of the Urban South*. The University Press of Kentucky, Lexington, Kentucky, 1985.

[6] G. Myrdal, *An American Dilemma*. Harper and Row, New York, 1962.

[7] C. S. Johnson, *Patterns of Negro Segregation*. Harper and Brothers Publishers, New York, 1943.

[8] R. L. Morrill, "The Negro ghetto: Problems and alternatives," *The Geographical Review*, vol. 55, pp. 339–361, July 1965.

[9] E. W. Burgess, "The growth of the city," in *The City*, R. E. Park, E. W. Burgess, and R. D. Mackenzie, Eds. University of Chicago Press, 1925.

[10] A. B. Schnare, *The Persistence of Racial Segregation in Housing*. The Urban Institute, Washington, D.C., 1978.

[11] J. F. Jakubs, "Recent racial segregation in U.S. SMSAs," *Urban Geography*, vol. 7, no. 2, pp. 146–163, 1986.

[12] "The dynamics of change: An analysis of growth in metropolitan Atlanta over the past two decades," Policy Research Center, Georgia State University, Atlanta, GA, Tech. Rep., 1993.

[13] E. J. Fournier, "Racial residential pattern of the Atlanta metro area : Historic evolution and development," November 1990, paper given at 45th Annual Meeting of the Association of American Geographers, Southeast Division, Columbia, South Carolina.

[14] U. S. Senate, *Desegregation and the cities - the trends and policy choices*. U.S. Government Printing Office, Washington, D.C., 1977.

[15] D. S. Massey and M. L. Eggers, "The ecology of inequality: Minorities and the concentration of poverty, 1970-1980," *American Journal of Sociology*, vol. 95, no. 5, pp. 1153–1188, 1990.

[16] D. L. Hamilton, S. Carpenter, and G. D. Bishop, "Desegregation of suburban neighborhoods," in *Groups in Contact: The Psychology of Desegregation*, N. Miller and M. Brewer, Eds. Academic Press, NY, 1984.

[17] R. D. Peterson and L. J. Krivo, "Racial segregation and Black urban homicide," *Social Forces*, vol. 4, pp. 1001–1026, June 71.

[18] J. Yinger, "Prejudice and discrimination in the urban housing market," in *Current Issues in Urban Economics*, P. Mieszkowski and M. Straszheim, Eds. Johns Hopkins Univerity Press, Baltimore, MD, 1979, pp. 430–468.

[19] R. L. Morrill and O. F. Donaldson, "Geographical perspectives on the history of Black America," in *Readings in Social Geography*, E. Jones, Ed. Oxford University Press, Oxford, England, 1975, pp. 15–38.

[20] D. McEntire, "Government and racial discrimination in housing," *Journal of Social Issues*, vol. 13, no. 4, pp. 60–67, 1957.

[21] ——, *Residence and Race*. University of California Press, Berkeley, CA, 1960.

[22] W. A. Leigh, "Civil rights legislation and the housing status of Black Americans: An overview," *The Review of Black Political Economy*, vol. 19, no. 3-4, pp. 5–28, 1991.

[23] D. S. Massey and N. A. Denton, *American Apartheid: Segregation and the Making of the Underclass*. Harvard University Press, Cambridge, MA, 1993.

[24] M. L. Streitwieser and J. L. Goodman, "A survey of recent research on race and residential location," *Population Research and Policy Review*, vol. 2, no. 3, pp. 253–283, 1983.

[25] E. P. Wolf, "The invasion-succession sequence as a self-fulfilling prophecy," *Journal of Social Issues*, vol. 13, no. 4, pp. 7–20, 1957.

[26] E. F. Schietinger, "Racial succession and value of small residential properties," *American Sociological Review*, vol. 16, no. 6, pp. 832–835, 1951.

[27] M. J. Bailey, "Effects of race and other demographic factors on the value of single-family homes," *Land Economics*, vol. 42, no. 2, pp. 215–220, May 1966.

[28] K. Commission, *Report of the National Advisory Commission on Civil Disorders*. E. P. Dutton and Co, Inc., New York, 1968.

[29] H. Barker and R. E. Anderson, *Equal Housing Opportunities in the South*. Southern Regional Council, Atlanta, GA, 1971.

[30] A. H. Pascal, "The economics of housing segregation," RAND, Santa Monica, CA, Tech. Rep., 1967, memorandum RM-5510-RC.

[31] L. M. Laurenti, "Property values and race," in *Readings in Urban Economics*, M. Edel and J. Rothenberg, Eds. Macmillan, New York, NY, 1972, pp. 267–288.

[32] R. C. Weaver, *The Negro Ghetto*. Harcourt Brace, Publishers, New York, NY, 1948.

[33] T. J. Sugrue, "A dream still deferred," New York Times, March 27 2011.

[34] K. T. Jackson, *Crabgrass Frontier: The Suburbanization of the United States*. Oxford University Press, New York, 1985.

[35] R. J. Johnston, *The American Urban System: A Geographical Perspective*. St. Martin's Press, New York, 1982.

[36] J. Heilbrun, *Urban Economics and Public Policy*, 3rd ed. St. Martin's Press, New York, 1987.

[37] G. Grier and E. Grier, "Obstacles to desegregation in America's urban areas," *Race*, vol. 6, no. 1, pp. 3–17, 1964.

[38] R. Farley, "Residential segregation in urbanized areas of the United States in 1970: An analysis of social class and racial differences," *Demography*, vol. 14, no. 4, pp. 497–518, November 1977.

[39] M. Coulibaly, "Racial and income segregation in low-income housing: 1934-1992," *Review of Radical Political Economics*, vol. 25, no. 3, pp. 104–112, 1993.

[40] R. L. Morrill, "Racial segregation and class in a liberal metropolis," *Geographical Analysis*, vol. 27, no. 1, pp. 22–41, 1995.

[41] D. R. Fusfeld, "The economy of the urban ghetto," in *Financing the Metropolis*, J. P. Crecine, Ed. Sage Publications, Beverly Hills, CA, 1970, vol. 4, pp. 369–399, urban Affairs Annual Reviews.

[42] R. Stanfield, "Apartheid, American-style," *National Journal*, vol. 25, no. 28, p. 1782, July 1993.

[43] The Economist, "The other America," *The Economist*, vol. 328, no. 7819, pp. 17–18, 1993.

[44] ——, "Whither shall it wander?" *The Economist*, vol. 328, no. 7819, pp. 18–19, 1993.

[45] ——, "Back to the ghetto?" *The Economist*, vol. 328, no. 7819, pp. 13–14, 1993.

[46] R. Cook and M. Christensen, "Georgia's 11th district thrown out," The Atlanta Journal, June 29 1995, a1, A10.

[47] R. W. Lake and S. C. Cutter, "A typology of black suburbanization in New Jersey since 1970," *Geographical Review*, vol. 70, no. 2, pp. 167–181, 1980.

[48] N. Fainstein, "Race, class and segregation: Discourses about African Americans," *International Journal of Urband and Regional Research*, vol. 17, no. 3, pp. 384–403, 1993.

[49] K. E. Taeuber and A. F. Taeuber, "The Negro as an immigrant group: Recent trends in racial and ethnic segregation in Chicago," *American Journal of Sociology*, vol. 69, no. 4, pp. 369–382, 1964.

[50] J. F. Kain, "Housing segregation, Negro employment, and metropolitan decentralization," *The Quarterly Journal of Economics*, vol. 82, no. 2, pp. 175–197, 1968.

[51] P. M. Huutchinson, "The effects of accessibility and segregation on the employment of the urban poor," in *Patterns of Racial Discrimination, Volume 1: Housing*, G. von Furstenberg *et al.*, Ed. Lexington Books, Lexington, Mass., 1974.

[52] R. D. Green and J. P. Reidy, "Accumulation, urban segregation and the Black role in the U.S. economy: A stylized history," *Review of Radical Political Economics*, vol. 24, no. 2, pp. 83–90, 1992.

[53] R. D. Green and D. M. James, "Is job accessibility a serious problem for Black workers?" *Review of Radical Political Economics*, vol. 25, no. 2, pp. 85–92, 1993.

[54] W. W. Goldsmith and E. J. Blakely, *Separate Societies: Poverty and Inequality in U.S. Cities*. Temple University Press, Philadelphia, PA, 1992.

[55] J. F. Jakubs and S. H. Pollard, "Suburbanization and South Carolina cities," *Urban and Regional Notes, Bureau of Urban and Regional Affairs, University of South Carolina*, June 1973.

[56] E. M. Hoover, "The measurement of industrial location," *Review of Economic Statistics*, vol. 18, no. 1, pp. 162–171, February 1936.

[57] ——, "Interstate redistribution of population, 1850-1940," *Journal of Economic History*, vol. 1, no. 1, pp. 199–205, November 1941.

[58] M. O. Lorenz, "Methods of measuring the concentration of wealth," *Quarterly Publications of the American Statistical Association*, vol. 9, pp. 209–219, June 1905.

[59] J. Morgan, "The anatomy of income distribution," *Review of Economics and Statistics*, vol. 44, pp. 270–283, August 1962.

[60] J. L. Gastwirth, "The estimation of the Lorenz curve and Gini index," *Review of Economics and Statistics*, vol. 54, pp. 306–316, August 1972.

[61] M. J. Bowman, "A graphical analysis of personal income distribution in the United States," *American Economic Review*, vol. 35, pp. 608–628, September 1945.

[62] M. W. Smelker, "Shifts in the concentration of income," *Review of Economics and Statistics*, vol. 30, pp. 215–222, August 1948.

[63] R. Zhang, "Measurement and evaluation of changes in income distributions: A welfare approach (China, Taiwan, Hong Kong)," [CD-ROM], 1992, abstract from: ProQuest File: Dissertation Abstracts Item: 9238372.

[64] W. F. Lubbe, "The red meat marketing scheme: An evaluation in a dynamic environment," [CD-ROM], 1992, abstract from: ProQuest File: Dissertation Abstracts.

[65] A. S. Memon, "Impacts of land reform on farm production and income distribution in the agricultural sector of Sindh province of Pakistan," [CD-ROM], 1993, abstract from: ProQuest File: Dissertation Abstracts Item: 9325745.

[66] C. Lee, "The economic impact of international inbound tourism on the South Korean economy and its distributional effects on income classes," [CD-ROM], 1992, abstract from: ProQuest File: Dissertation Abstracts Item: 9315100.

[67] J. A. Jahn, C. F. Schmid, and C. Schrag, "The measurement of ecological segregation," *American Sociological Review*, vol. 12, pp. 293–304, June 1947.

[68] R. A. Hornseth, "A note on 'The Measurement of Ecological Segregation' by Julius Jahn, Calvin F. Schmid, and Clarence Schrag," *American Sociological Review*, vol. 12, no. 5, pp. 603–604, 1947.

[69] J. J. Williams, "Another commentary on so-called segregation indices," *American Journal of Sociology*, vol. 13, no. 3, pp. 298–303, June 1948.

[70] J. A. Jahn, "The measurement of ecological segregation: Derivation of an index based upon the criterion of reproducibility," *American Sociological Review*, vol. 15, no. 1, pp. 100–104, 1950.

[71] D. O. Cowgill and M. S. Cowgill, "An index of segregation based on block statistics," *American Sociological Review*, vol. 16, no. 6, pp. 825–831, December 1951.

[72] M. J. White, "Segregation and diversity measures in population distribution," *Population Index*, vol. 52, no. 2, pp. 198–221, 1986.

[73] C. J. Dawkins, "Measuring the spatial pattern of residential segregation," *Urban Studies*, vol. 41, no. 4, pp. 833–851, April 2004.

[74] M. J. White, "The measurement of spatial segregation," *American Journal of Sociology*, vol. 88, no. 5, pp. 1008–1018, 1983.

[75] K. E. Taeuber and A. F. Taeuber, *Negroes in Cities*. Aldine Publishing Company, Chicago, IL, 1965.

[76] B. S. Morgan, "The properties of a distance-based segregation index," *Journal of Socio-Economic Planning Sciences*, vol. 16, pp. 167–171, 1982.

[77] C. F. Cortese, R. F. Falk, and J. K. Cohen, "Further considerations on the methodological analysis of segregation indices," *American Sociological Review*, vol. 41, no. 4, pp. 630–637, 1976.

[78] J. F. Jakubs, "Residential segregation: The Taeuber index reconsidered," *Journal of Regional Science*, vol. 17, no. 2, pp. 281–283, 1977.

[79] ——, "A consistent conceptual definition of the index of dissimilarity," *Geographical Analysis*, vol. 11, no. 3, pp. 315–321, July 1979.

[80] ——, "A distance-based segregation index," *Journal of Socio-Economic Planning Sciences*, vol. 15, no. 6, pp. 129–131, 1981.

[81] B. S. Morgan, "An alternative approach to the development of a distance-based segregation index," *American Journal of Sociology*, vol. 88, no. 6, pp. 1237–1249, 1983.

[82] R. L. Morrill, "On the measure of geographic segregation," *Geography Research Forum*, vol. 11, pp. 25–36, 1991.

[83] D. W. S. Wong, "Spatial indices of segregation," *Urban Studies*, vol. 30, no. 3, pp. 559–572, 1993.

[84] R. F. Falk, C. F. Cortese, and J. Cohen, "Utilizing standardized indices of residential segregation: Comment on Winship," *Social Forces*, vol. 57, no. 2, pp. 713–716, December 1978.

[85] D. S. Massey, "On the measurement of segregation as a random variable," *American Sociological Review*, vol. 43, pp. 587–590, 1978.

[86] H. F. Inman and E. L. Bradley, "Approximations to the mean and variance of the index of dissimilarity in 2xC tables under a random allocation model," *Sociological Methods and Research*, vol. 20, no. 2, pp. 242–255, 1991.

[87] W. P. Frisbie, "Urban sociology in the United States: the past 20 years," *American Behavioral Scientist*, vol. 24, no. 2, pp. 177–214, November/December 1980.

[88] R. A. Elgie, "The segregation of socioeconomic groups in urban areas: a comment," *Urban Studies*, vol. 16, pp. 191–195, 1979.

[89] J. F. Jakubs, *Measuring Segregation in Point Patterns: A Distance-based Approach*. Instrument Society of America, 1984, pp. 111–115.

[90] W. Bell, "A probability model for the measurement of ecological segregation," *Social Forces*, vol. 32, no. 4, pp. 357–364, 1954.

[91] O. D. Duncan and B. Duncan, "Residential distribution and occupational stratification," *American Journal of Sociology*, vol. 60, no. 5, pp. 493–503, 1955.

[92] J. Kim and C. W. Mueller, *Introduction to Factor Analysis: What It Is and How To Do It*, ser. Sage University Paper Series on Quantitative Applications in the Social Sciences. Sage Publications, Beverly Hills and London, 1978, vol. 07-013.

[93] ——, *Factor Analysis: Statistical Methods and Practical Issues*, ser. Sage University Paper Series on Quantitative Applications in the Social Sciences. Sage Publications, Beverly Hills and London, 1978, vol. 07-014.

[94] R. J. Rummel, *Applied Factor Analysis*, 2nd ed. Northwestern University Press, Evanston, IL, 1970.

[95] G. M. Herek, "Attitudes toward lesbians and gay men: A factor-analytic study," *Journal of Homosexuality*, vol. 10, no. 1, pp. 39–51, 1984, fall.

[96] J. Schmid and J. M. Leiman, "The development of hierarchical factor solutions," *Psychometrika*, vol. 22, no. 1, pp. 53–61, 1957.

[97] A. L. Comrey, "Common methodological problems in factor analytic studies," *Journal of Counsulting and Clinical Psychology*, vol. 46, no. 4, pp. 648–659, 1978.

[98] J. C. Loehlin, *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis*, 2nd ed. Lawrence Erlbaum Associates, Hillsdale, NJ, 1992.

[99] G. Strang, *Linear Algebra and Its Applications*, 2nd ed.   Academic Press, New York, 1980.

[100] L. L. Thurstone, *Multiple Factor Analysis*.   University of Chicago Press, Chicago, 1947.

<div align="center">**Appendix A**</div>

<div align="center">**COMPUTATIONAL FORMULAE FOR INDICES**</div>

A.1   Introduction

In their 1988 paper, Massey and Denton used twenty common indices of segregation for their analysis. Each index was chosen for its measurement of a particular dimension of the problem. The indices for each dimension are presented here, each with its computational formula and relevant citations. All formulae are taken from Massey and Denton.

A.2   Dimension : Evenness

A.2.1   Index of Dissimilarity (D)

Computational Formula

$$D = \sum_{i=1}^{n} \frac{t_i \mid p_i - P \mid}{2 \times T \times P \times (1 - P)}$$

Parameters

t_i total population in tract i;

p_i total minority population in tract i;

T total population of entire metropolitan area; and

P total minority population of entire metropolitan area.

Citations

Taeuber and Taeuber, 1965

A.2.2 Gini Coefficient

Computational Formula

$$G = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{t_i \times t_j \times \mid p_i - p_j \mid}{2 \times T^2 \times P \times (1 - P)}$$

Parameters

t_i total population in tract i;

t_j total population in tract j;

p_i total minority population in tract i;

p_j total minority population in tract j;

T total population of entire metropolitan area; and

P total minority population of entire metropolitan area.

Citations

Gastwirth, 1972; Allison, 1978

A.2.3 Entropy Index (a.k.a. Information Index)

Computational Formula

$$
\begin{aligned}
H &= \sum_{i=1}^{n} \frac{t_i \times (E - E_i)}{E \times T} \\
E &= P \times \log\left(\frac{1}{P}\right) + (1 - P) \times \log\left(\frac{1}{(1 - P)}\right) \\
E_i &= p_i \times \log\left(\frac{1}{p_i}\right) + (1 - p_i) \times \log\left(\frac{1}{(1 - p_i)}\right)
\end{aligned}
$$

E total city entropy;

E_i tract's entropy;

t_i total population in tract i;

p_i total minority population in tract i;

T total population of entire metropolitan area; and

59

P total minority population of entire metropolitan area.

Citations

Theil and Finizza, 1971; Theil, 1972

A.2.4 Atkinson Index

Computational Formula

$$A = 1 - \frac{P}{(1 - P)} \left| \sum_{i=1}^{n} \frac{(1 - p_i)^{1-b} \times p_i^b \times t_i}{P \times T} \right|^{\frac{1}{(1-b)}}$$

Parameters

t_i total population in tract i;

p_i total minority population in tract i;

T total population of entire metropolitan area;

P total minority population of entire metropolitan area; and

b shape parameter that determines how to weight the increments to segregation contributed by different portions of the Lorenz curve.

$0 < b < 0.5$	units where $p_i < P$ contribute more
$b = 0.5$	all units contribute equally
$0.5 < b < 1.0$	units where $p_i > P$ contribute more

Citations

James and Taeuber, 1985; White, 1986

A.3   Dimension : Exposure

A.3.1   Bell's Isolation/Interaction Index

Computational Formula

Isolation Index

$$\Pi^*_{xx} = \sum_{i=1}^{n} \left(\frac{x_i}{X} \times \frac{x_i}{t_i} \right)$$

Interaction Index

$$\Pi^*_{xy} = \sum_{i=1}^{n} \left(\frac{x_i}{X} \times \frac{y_i}{t_i} \right)$$

Parameters

x_i number of X members;

y_i number of Y members;

t_i total population of unit i; and

X number of X members city-wide.

Citations

Bell, 1954; Lieberson, 1981

A.3.2   Eta (correlation ratio)

Computational Formula

$$V = \eta^2 = \frac{\Pi^*_{xx} - \frac{X}{T}}{1 - \frac{X}{T}}$$

Parameters

Π^*_{xx} Bell's Isolation Index;

T total population of entire metropolitan area;

X number of X members city-wide.

Citations

White, 1986

A.4 Dimension : Concentration

A.4.1 Delta

Computational Formula

$$DEL = \frac{1}{2} \sum_{i=1}^{n} \left| \frac{x_i}{X} - \frac{a_i}{A} \right|$$

Parameters

x_i number of X members;

X number of X members city-wide;

a_i land area of unit i; and

A total land area in city.

Citations

Hoover, 1941; Duncan, Cuzzort, and Duncan, 1961

A.4.2 Absolute Concentration Index

Computational Formula

$$ACO = 1 - \frac{\sum_{i=1}^{n} \left(\frac{x_i a_i}{X} \right) - \sum_{i=1}^{n_1} \left(\frac{t_i a_i}{T} \right)}{\sum_{i=n_2}^{n} \left(\frac{t_i a_i}{T_2} \right) - \sum_{i=1}^{n_1} \left(\frac{t_i a_i}{T} \right)}$$

A.4.3 Relative Concentration Index

Computational Formula

$$RCO = \frac{\dfrac{\sum_{i=1}^{n} \dfrac{x_i a_i}{X}}{\sum_{i=1}^{n} \dfrac{y_i a_i}{Y}} - 1}{\dfrac{\sum_{i=1}^{n_1} t_i \dfrac{a_i}{T_1}}{\sum_{i=n_2}^{n} t_i \dfrac{a_i}{T_2}} - 1}$$

Parameters

a_i land area of unit i, ranked from smallest to largest;

n_1 rank of tract where the cumulative total population of areal units equals the total minority population of the city, summing from smallest unit up;

n_2 rank of tract where the cumulative total population of areal units equals the total minority population of the city, summing from largest unit down;

T_1 total population of tracts from 1 to n_1;

T_2 total population of tracts from n_1 to n;

x_i number of X members;

y_i number of Y members;

t_i total population of unit i;

Y number of Y members city-wide; and

X number of X members city-wide.

Citations

Massey and Denton, 1988

A.5 Dimension : Centralization

A.5.1 Proportion in Central City

Computational Formula

$$PCC = \frac{X_{cc}}{X}$$

Parameters

X_{cc} number of X members living within the central city; and

X number of X members city-wide.

Citations

Grebler, Moore, and Guzman, 1970; Massey, 1979

A.5.2 Relative Centralization Index

Computational Formula

$$RCE = \left(\sum_{i=1}^{n} [X_{i-1} \times Y_i] \right) - \left(\sum_{i=1}^{n} [X_i \times Y_{i-1}] \right)$$

Parameters

n number of areal units, ordered by increasing distance from CBD;

X_i proportion of X population in tract i; and

Y_i proportion of Y population in tract i.

Citations

Duncan and Duncan, 1955b

A.5.3   Absolute Centralization Index

Computational Formula

$$ACE = \left(\sum_{i=1}^{n} [X_{i-1} \times A_i] \right) - \left(\sum_{i=1}^{n} [X_i \times A_{i-1}] \right)$$

Parameters

n number of areal units;

X_i proportion of X population in tract i; and

A_i cumulative proportion of land area through unit i.

Citations

Massey and Denton, 1988

A.6 Dimension : Clustering

A.6.1 Absolute Clustering Index

Computational Formula

$$ACL = \frac{\sum_{i=1}^{n} \frac{x_i}{X} \sum_{j=1}^{n} c_{ij} \times x_j - \frac{X}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}}{\sum_{i=1}^{n} \frac{x_i}{X} \sum_{j=1}^{n} c_{ij} \times t_j - \frac{X}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}}$$

Parameters

x_i number of X members in tract i;

X number of X members city-wide;

n number of areal units;

t_j total population of tract j;

c_{ij} distance function between areas ($\exp(-d_{ij})$);

d_{ij} distance between areal unit centroids; and

d_{ii} $(0.6a - i)^{0.5}$ (after White, 1983).

Citations

Massey and Denton, 1988

A.6.2 White's Index of Spatial Proximity

Computational Formula

$$P = \frac{N_1 P_{11} + N_2 P_{22}}{(N_1 + N_2) \times P_{00}}$$

$$P_{00} = \frac{1}{(N_1 + N_2)^2} \sum_i \sum_j (N_{1i} + N_{2i}) \times (N_{1j} + N_{2j}) \times f(d_{ij})$$

$$P_{11} = \frac{1}{(N_1)^2} \sum_i \sum_j N_{1i} \times N_{1j} \times f(d_{ij})$$

$$P_{22} = \frac{1}{(N_2)^2} \sum_i \sum_j N_{2i} \times N_{2j} \times f(d_{ij})$$

Parameters

N_1 total population of group 1 in city;

N_2 total population of group 2 in city;

N_{1i} population of group 1 in ith tract; and

N_{2i} population of group 2 in ith tract.

Citations

White, 1983; White, 1986

A.6.3 Relative Clustering Index

Computational Formula

$$RCL = \frac{P_{xx}}{P_{yy}} - 1$$

Parameters

P_{xx} $\frac{\sum_{i=1}^{n} \sum_{i=1}^{n} x_i x_j c_{ij}}{X^2}$

P_{yy} $\frac{\sum_{i=1}^{n} \sum_{i=1}^{n} x_i y_j c_{ij}}{XY}$

c_{ij} $\exp(-d_{ij})$

Citations

Massey and Denton, 1988

Appendix B

CORRELATION MATRIX

B.1 Complete Correlation Matrix for Indices

The first step in a factor analysis is to examine the correlations between the variables. From the correlation matrix it can be determined if the data are well suited to the technique. Possible problems can also be identified such as the perfect inverses discovered in the 1980 Census data. The full 19 x 19 correlation matrix follows.

Index	Atkin1	Atkin5	Atkin9	ACE	ACL	ACO	Eta	D	DPxx	DPxy	DEL	Entropy	Gini	Pxx	Pxy	RCE	RCO	RCL	SP
Atkin1 Atkinson Index at 0.01	1	0.81985 <.0001	0.69017 <.0001	0.182 0.0009	0.4905 <.0001	0.26982 <.0001	0.5395 <.0001	0.77622 <.0001	0.34469 <.0001	-0.34545 <.0001	0.43532 <.0001	0.67489 <.0001	0.75452 <.0001	0.43069 <.0001	-0.43069 <.0001	0.33541 <.0001	0.27613 <.0001	0.32498 <.0001	0.50925 <.0001
Atkin5 Atkinson Index at 0.05	0.81985 <.0001	1	0.96619 <.0001	0.1717 0.0017	0.70317 <.0001	0.22963 <.0001	0.85187 <.0001	0.98369 <.0001	0.63473 <.0001	-0.63551 <.0001	0.41662 <.0001	0.95226 <.0001	0.98027 <.0001	0.75911 <.0001	-0.75911 <.0001	0.34315 <.0001	0.28262 <.0001	0.34341 <.0001	0.70001 <.0001
Atkin9 Atkinson Index at 0.09	0.69017 <.0001	0.96619 <.0001	1	0.16142 0.0033	0.66759 <.0001	0.17975 0.001	0.85898 <.0001	0.95169 <.0001	0.66476 <.0001	-0.66602 <.0001	0.36799 <.0001	0.9394 <.0001	0.97501 <.0001	0.7897 <.0001	-0.7897 <.0001	0.29686 <.0001	0.25665 <.0001	0.29455 <.0001	0.65524 <.0001
ACE Absolute Centralization Index	0.182 0.0009	0.1717 0.0017	0.16142 0.0033	1	0.00925 0.8671	0.53469 <.0001	0.06992 0.2052	0.16735 0.0023	-0.06531 0.2367	0.06374 0.2482	0.62472 <.0001	0.12409 0.0242	0.17524 0.0014	-0.00376 0.9457	0.00376 0.9457	0.57989 <.0001	0.45704 <.0001	-0.04182 0.449	0.043 0.4363
ACL Absolute Clustering Index	0.4905 <.0001	0.70317 <.0001	0.66759 <.0001	0.00925 0.8671	1	-0.18065 0.001	0.87479 <.0001	0.68033 <.0001	0.89443 <.0001	-0.89463 <.0001	0.01834 0.7399	0.82412 <.0001	0.65437 <.0001	0.87413 <.0001	-0.87413 <.0001	0.14255 0.0095	-0.08489 0.1238	0.31473 <.0001	0.95492 <.0001
ACO Absolute Concentration Index	0.26982 <.0001	0.22963 <.0001	0.17975 0.001	0.53469 <.0001	-0.18065 0.001	1	-0.01476 0.7894	0.24698 <.0001	-0.30952 <.0001	0.30858 <.0001	0.83928 <.0001	0.11462 0.0374	0.24902 <.0001	-0.16941 0.002	0.16941 0.002	0.43755 <.0001	0.79106 <.0001	-0.11685 0.0338	-0.11175 0.0425
Eta Correlation Index	0.5395 <.0001	0.85187 <.0001	0.85898 <.0001	0.06992 0.2052	0.87479 <.0001	-0.01476 0.7894	1	0.82329 <.0001	0.89921 <.0001	-0.8996 <.0001	0.13764 0.0123	0.96486 <.0001	0.81406 <.0001	0.9749 <.0001	-0.9749 <.0001	0.22675 <.0001	0.13257 0.016	0.18699 0.0006	0.84432 <.0001
D Dissimilarity Index	0.77622 <.0001	0.98369 <.0001	0.95169 <.0001	0.16735 0.0023	0.68033 <.0001	0.24698 <.0001	0.82329 <.0001	1	0.60602 <.0001	-0.60629 <.0001	0.43359 <.0001	0.9346 <.0001	0.98582 <.0001	0.73061 <.0001	-0.73061 <.0001	0.33686 <.0001	0.27904 <.0001	0.36473 <.0001	0.68284 <.0001
DPxx Distance Decay Interaction	0.34469 <.0001	0.63473 <.0001	0.66476 <.0001	-0.06531 0.2367	0.89443 <.0001	-0.30952 <.0001	0.89921 <.0001	0.60602 <.0001	1	-0.99986 <.0001	-0.12524 0.0229	0.7944 <.0001	0.60636 <.0001	0.95959 <.0001	-0.95959 <.0001	0.06369 0.2486	-0.07551 0.1712	0.0893 0.1054	0.83682 <.0001
DPxy Distance Decay Isolation	-0.34545 <.0001	-0.63551 <.0001	-0.66602 <.0001	0.06374 0.2482	-0.89463 <.0001	0.30858 <.0001	-0.8996 <.0001	-0.60629 <.0001	-0.99986 <.0001	1	0.12485 0.0233	-0.79503 <.0001	-0.60698 <.0001	-0.95975 <.0001	0.95975 <.0001	-0.06426 0.2443	0.07503 0.1739	-0.08906 0.1063	-0.83683 <.0001
DEL Delta	0.43532 <.0001	0.41662 <.0001	0.36799 <.0001	0.62472 <.0001	0.01834 0.7399	0.83928 <.0001	0.13764 0.0123	0.43359 <.0001	-0.12524 0.0229	0.12485 0.0233	1	0.2836 <.0001	0.43924 <.0001	-0.00622 0.9104	0.00622 0.9104	0.40586 <.0001	0.64778 <.0001	0.08948 0.1047	0.07018 0.2035
Entropy Entropy Index	0.67489 <.0001	0.95226 <.0001	0.9394 <.0001	0.12409 0.0242	0.82412 <.0001	0.11462 0.0374	0.96486 <.0001	0.9346 <.0001	0.7944 <.0001	-0.79503 <.0001	0.2836 <.0001	1	0.92239 <.0001	0.90172 <.0001	-0.90172 <.0001	0.29524 <.0001	0.21005 0.0001	0.29343 <.0001	0.80879 <.0001
Gini Gini Index	0.75452 <.0001	0.98027 <.0001	0.97501 <.0001	0.17524 0.0014	0.65437 <.0001	0.24902 <.0001	0.81406 <.0001	0.98582 <.0001	0.60636 <.0001	-0.60698 <.0001	0.43924 <.0001	0.92239 <.0001	1	0.7314 <.0001	-0.7314 <.0001	0.3333 <.0001	0.28674 <.0001	0.3236 <.0001	0.64964 <.0001
Pxx Isolation Index	0.43069 <.0001	0.75911 <.0001	0.7897 <.0001	-0.00376 0.9457	0.87413 <.0001	-0.16941 0.002	0.9749 <.0001	0.73061 <.0001	0.95959 <.0001	-0.95975 <.0001	-0.00622 0.9104	0.90172 <.0001	0.7314 <.0001	1	-1 <.0001	0.15398 0.0051	0.03956 0.4739	0.1245 0.0237	0.82398 <.0001
Pxy Interaction Index	-0.43069 <.0001	-0.75911 <.0001	-0.7897 <.0001	0.00376 0.9457	-0.87413 <.0001	0.16941 0.002	-0.9749 <.0001	-0.73061 <.0001	-0.95959 <.0001	0.95975 <.0001	0.00622 0.9104	-0.90172 <.0001	-0.7314 <.0001	-1 <.0001	1	-0.15398 0.0051	-0.03956 0.4739	-0.1245 0.0237	-0.82398 <.0001
RCE Relative Centralization Index	0.33541 <.0001	0.34315 <.0001	0.29686 <.0001	0.57989 <.0001	0.14255 0.0095	0.43755 <.0001	0.22675 <.0001	0.33686 <.0001	0.06369 0.2486	-0.06426 0.2443	0.40586 <.0001	0.29524 <.0001	0.3333 <.0001	0.15398 0.0051	-0.15398 0.0051	1	0.5242 <.0001	-0.00918 0.868	0.16442 0.0027
RCO Relative Concentration Index	0.27613 <.0001	0.28262 <.0001	0.25665 <.0001	0.45704 <.0001	-0.08489 0.1238	0.79106 <.0001	0.13257 0.016	0.27904 <.0001	-0.07551 0.1712	0.07503 0.1739	0.64778 <.0001	0.21005 0.0001	0.28674 <.0001	0.03956 0.4739	-0.03956 0.4739	0.5242 <.0001	1	-0.38627 <.0001	-0.08786 0.1111
RCL Relative Clustering Index	0.32498 <.0001	0.34341 <.0001	0.29455 <.0001	-0.04182 0.449	0.31473 <.0001	-0.11685 0.0338	0.18699 0.0006	0.36473 <.0001	0.0893 0.1054	-0.08906 0.1063	0.08948 0.1047	0.29343 <.0001	0.3236 <.0001	0.1245 0.0237	-0.1245 0.0237	-0.00918 0.868	-0.38627 <.0001	1	0.4175 <.0001
SP Spatial Proximity	0.50925 <.0001	0.70001 <.0001	0.65524 <.0001	0.043 0.4363	0.95492 <.0001	-0.11175 0.0425	0.84432 <.0001	0.68284 <.0001	0.83682 <.0001	-0.83683 <.0001	0.07018 0.2035	0.80879 <.0001	0.64964 <.0001	0.82398 <.0001	-0.82398 <.0001	0.16442 0.0027	-0.08786 0.1111	0.4175 <.0001	1

Figure B.1: Correlation Matrix for 1980 Segregation Indices

69

Appendix C

SAS CODE LISTINGS

C.1 Dataset Creation

```
filename tract80 "./raw_data/tract1980black.csv";
libname store "./store";
;

data group1;
    infile tract80 DSD DLM="," firstobs=2;
    input MSA_id total pctmin
          Atkin1 Atkin5 Atkin9 ACE ACL ACO Eta D DPxx DPxy DEL Entropy Gini
          Pxy Pxx RCE RCL RCO SP;
    label pctmin = 'Percent Black'
          D       = 'Dissimilarity Index'
          Gini    = 'Gini Index'
          Entropy = 'Entropy Index'
          Atkin1  = 'Atkinson Index at 0.01'
          Atkin5  = 'Atkinson Index at 0.05'
          Atkin9  = 'Atkinson Index at 0.09'
          Pxx     = 'Isolation Index'
          Pxy     = 'Interaction Index'
          Eta     = 'Correlation Index'
          DEL     = 'Delta'
          RCO     = 'Relative Concentration Index'
          ACO     = 'Absolute Concentration Index'
          ACE     = 'Absolute Centralization Index'
          RCE     = 'Relative Centralization Index'
          ACL     = 'Absolute Clustering Index'
          RCL     = 'Relative Clustering Index'
          SP      = 'Spatial Proximity'
          DPxx    = 'Distance Decay Interaction'
          DPxy    = 'Distance Decay Isolation';
;
run;
;
```

C.2 Code for Initial Analysis Using Massey and Denton's Approach

```
libname study "./store";
;
OPTIONS nocenter;

data workset;
   set study.black80;
;
proc factor data=workset
            scree
            reorder
            rotate=varimax
            score
            outstat=pass1;
   var Atkin1--SP;
   title 'Varimax Factor Analysis of 1980 Census Data - Black';
;
run;
;
```

C.3 Code for Factor Analysis

C.3.1 Initial Set-Up

```
libname study "./store";
;
OPTIONS nocenter;

data workset;
   set study.black80;
;
proc factor data=workset
            method=p
            priors=smc
            reorder
            mineigen=1
            rotate=varimax
            score
            outstat=pass1;

*  first attempt with all indices;
   var Atkin1--SP;
   title 'Varimax Factor Analysis of 1980 Census Data - Black';
;
run;
;
```

C.3.2 Final Set-Up

```
libname study "./store";
;
OPTIONS nocenter;

data workset;
   set study.black80;
;
proc factor data=workset
            method=p
            priors=smc
            reorder
            mineigen=1
            rotate=varimax
            score
            outstat=pass1;

*  first attempt with all indices;
*  second attempt with DPxy Pxy removed;
*  third attempt with RCL removed;
   var Atkin1 Atkin5 Atkin9 ACE ACL ACO Eta D DPxx DEL Entropy Gini
       Pxx RCE RCO SP;
   title 'Varimax Factor Analysis of 1980 Census Data - Black';
;
run;
;
```