

Measuring Residential Segregation

Trevon D. Logan^{*} and John M. Parman[†]

December 12, 2013

Abstract

We develop a new measure of residential segregation based on individual-level data. We exploit complete census manuscript files to derive a measure of segregation based upon the racial similarity of next door neighbors. Our measure overcomes several of the shortcomings of traditional segregation indices and allows for a much richer view of the variation in segregation patterns across time and space. With our new measure, we can distinguish between the effects of increasing the racial homogeneity of a location and of increasing the tendency to segregate within a location given a particular racial composition. We provide estimates of how our new measure relates to traditional segregation measures and historical factors. We also show how the segregation measure is related to the health outcomes of African Americans through late nineteenth and twentieth centuries. We conclude with a discussion of how this measure can be used in a variety of ways to improve and extend the analysis of segregation and its effects.

JEL classifications: I1, J1, N3

PRELIMINARY DRAFT

Do Not Cite, Quote, or Circulate Without Permission

^{*}Department of Economics, The Ohio State University and NBER, 1945 N. High Street, 410 Arps Hall, Columbus, OH 43210 e-mail: logan.155@osu.edu

[†]Department of Economics, College of William and Mary and NBER, 130 Morton Hall, Williamsburg, VA 23187 e-mail: jmparman@wm.edu

[‡]We thank David Blau, Joe Ferrie, Daeho Kim and Richard Steckel for suggestions on this project. William D. Biscarri, Nicholas J. Deis, Jackson L. Frazier, Adaeze Okoli, Terry L. Pack and Stephen Prifti provided excellent research assistance. The usual disclaimer applies.

"We make our friends; we make our enemies; but God makes our next door neighbor."

- Gilbert K. Chesterton, *Heretics* (1905)

1 Introduction

This paper introduces a new measure of residential segregation. Our measure uses the availability of the complete manuscript pages for the federal census to identify the races of next-door neighbors. We measure segregation by comparing the number of household heads in an area living next to neighbors of a different race to the expected number under complete segregation and under no segregation (random assignment). The resulting statistic provides a measure of how much residents tend to segregate themselves given a particular racial composition for an area. The measure allows us to distinguish between the effects of differences in racial composition and the tendency to segregate given a particular racial composition. A particular advantage is that we can aggregate it to any boundary without losing the underlying properties since it is defined at the individual level. Furthermore, the measure is equally applicable to both urban and rural areas.

To our knowledge, our measure of segregation is the first to exploit actual residential living patterns and the first to be equally applicable to rural and urban areas. Previous advances in the measurement of segregation have attempted to use smaller geographic units (Echenique & Fryer, 2007; Reardon et al., 2008), but none have exploited the actual pattern of household location that we do here. Similarly, our measure of segregation applies to the entire United States, not only urban areas. While analysis of segregation has primarily been focused on cities, there are few theoretical reasons to believe that rural segregation is unimportant or unrelated to socioeconomic outcomes for rural residents (Lichter et al., 2007).

The importance of residential segregation in explaining modern racial differences in socioeconomic outcomes is well known. There are a variety of studies linking segregation in the United States to schooling and labor market outcomes for blacks (Kain, 1968; Cutler et al., 1999; Cutler & Glaeser, 1997; Collins & Margo, 2000). Segregation has also been shown to impact the health of the black

community through a lack of access to health care (Almond et al., 2006; Chay et al., 2009). Additionally, there is a growing literature on the importance of neighborhood effects and social networks suggesting that segregated neighborhoods could contribute to racial gaps in a variety of socioeconomic outcomes (Case & Katz, 1991; Brooks-Gunn et al., 1993; Borjas, 1995; Cutler et al., 2008; Ananat, 2011; Ananat & Washington, 2009; Echenique & Fryer, 2007).

It is clear that any explanation of modern racial differences in socioeconomic outcomes must account for the effects of residential segregation on a host of factors. The literature suggests that the effects of segregation on socioeconomic outcomes are potentially strong but also complicated. The effects depend on the precise pattern of segregation, the extent of social interactions within and between groups, and the extent to which residential segregation leads to differential access to schools, health care and labor markets.

Despite the extensive documentation of the importance of segregation in the modern economy, we have little quantitative evidence on the evolution of segregation patterns over time. Segregation measures are inherently static. Traditional segregation measures are ill-suited to describe the evolution of segregation over time. Segregation can change in significant ways over time, and the effects of segregation may change over time as well. For example, between Reconstruction and World War II there was dramatic change in the black population's urban location. In 1870 roughly 90 percent of blacks lived outside of cities and by 1940 more than half lived in urban areas. Given the significant impacts of segregation in the modern economy, it is important to understand how changes in segregation patterns influenced outcomes for those in cities and rural areas. A broad, long-run view of segregation will help us understand its function and change over time.

In what follows we derive our measure of segregation and perform a simulation exercise to verify the properties of the measure. The simulation establishes that our measure captures residential housing patterns and performs as predicted as the dispersion of households by race and the underlying racial composition of the area vary. We then apply our measure to the full, 100% census, exploiting the census takers' sequenced alignment of households to identify the race of household heads and their neighbors.

The results uncover a substantial amount of heterogeneity in segregation within and across regions

in both cities and rural areas. We also show how our measure of segregation differs from the existing segregation measures in important ways. Our measure is correlated with the percent black in a county, but shows that the percentage black hides a considerable amount of racial segregation and integration. Even more, our measure is weakly correlated with standard segregation measures such as dissimilarity and isolation. A key result in our comparison is that we show that traditional measures are very sensitive to geographic boundaries while our measure is not.

Finally, we show that our segregation measure has a strong effect on a range of individual and aggregate outcomes not only historically but at present. Using a variety of data sources, we show that our segregation measure is well correlated with health. This relationship holds even when controlling for the underlying racial composition and traditional measures of segregation. We conclude by noting how our new measure of segregation can be used in a variety of studies for both the historical and contemporary effects of segregation.

2 Traditional Measures of Segregation

A wide range of measures have been introduced to measure segregation. Massey & Denton (1988) provide an overview of twenty different measures in use in the segregation literature. These various measures all capture different dimensions of segregation. Massey & Denton broadly categorize these dimensions as centralization, concentration, exposure, evenness and clustering. The majority of the segregation literature has focused on the dimensions of evenness and exposure. In particular, these are the dimensions of segregation that economists have held as most important in determining how segregation influences socioeconomic outcomes.

Evenness is the differential distribution of social groups across geographic subunits. As evenness decreases it becomes more likely that minorities have significantly different access to schooling and health resources as well as labor markets due to their concentration in specific subunits. Exposure measures the degree of potential contact between social groups. To the extent that social networks and peer effects are important for outcomes, differences in the levels of exposure will have potentially significant consequences for groups excluded from such networks due to limited contact. The eco-

nomics literature on segregation has typically relied on two specific measures of these dimensions: the index of dissimilarity as a measure of evenness and the index of isolation as a measure of exposure.

The index of dissimilarity is essentially a measure of how similar the distribution of minority residents among geographical units is to the distribution of non-minority residents among those same units. The measure is typically calculated at the city-level and is based on the distribution of minorities across census tracts or other sub-units within the city. Formally, if i is an index for the N census tracts within a city, B_i is the number of black residents in tract i , B_{total} is the total number of black residents in the city, W_i is the number of white residents in tract i , and W_{total} is the total number of white residents, the index of dissimilarity for the city is:¹

$$\text{Dissimilarity} = \frac{1}{2} \sum_{i=1}^N \left| \frac{B_i}{B_{\text{total}}} - \frac{W_i}{W_{\text{total}}} \right| \quad (1)$$

One way to interpret this index is as measure of how evenly black residents are distributed across tracts within a city. If black residents are distributed identically to white residents (a tract with 10 percent of the black residents also has 10 percent of the white residents), the index of dissimilarity will be zero. As black residents become less evenly dispersed across census tracts, the index of dissimilarity takes on a larger value.

The index of isolation provides a measure of the exposure of minority residents to other individuals outside of their group. Using the same notation as above, the index of isolation for a city is given by:

$$\text{Isolation} = \sum_{i=1}^N \left(\frac{B_i}{B_{\text{total}}} \cdot \frac{B_i}{B_i + W_i} \right) \quad (2)$$

This is essentially a measure of the racial composition of the census tract for the average black resident, where racial composition is measured as the percentage of the residents in the tract who are black. If there is little segregation, this measure will approach the percent black for the city as a whole. If there is extensive segregation (blacks are highly isolated), this measure will approach one as the tracts

¹Note that the index of dissimilarity typically compares the minority group of interest, black residents in this case, to all other residents. Throughout this paper, we restrict our attention to only black residents and white residents. Consequently, any term in a segregation measure that is a function of non-black residents can instead be thought of more simply as a function of white residents.

containing black residents become more and more homogeneous.

Cutler et al. (1999) and Collins & Margo (2000) use these measures to consider the changes in urban residential patterns over the twentieth century and find that levels of segregation rose dramatically over the twentieth century as blacks migrated to cities and then became more concentrated in city centers as white residents gradually moved to suburbs. They find that the high levels of segregation in American cities in the late twentieth century were largely absent in the early twentieth century. However, while all cities followed the general trend of rising segregation levels over time, at any single point in time there was substantial heterogeneity across cities in the level of segregation. These patterns across cities tended to persist over time, with the most segregated cities at the turn of the century also being the most segregated cities at the end of the century.² It is this variation in segregation across cities that Troesken (2002) exploits when looking at the health improvements resulting from the provision of water and sewerage service during the Jim Crow era. Cities that were more segregated as measured by the index of isolation saw smaller health improvements for black residents relative to white residents.

There is a general critique of using isolation and dissimilarity to measure segregation that has special importance when considering the history and evolution of segregation. Echenique & Fryer (2007) note that these measures are highly dependent on the way the boundaries of the geographical subunits are drawn. Indeed, it is an issue that Cutler et al. must deal with when the available data switches from ward-level data to census tract-level data in 1950. In the cases where Cutler et al. have data at both the ward and census tract levels, the correlation between the index of dissimilarity using wards and the same index using census tracts is only 0.59—0.35 with one outlier removed.

What makes this particularly problematic for historical segregation is that political motivations when drawing ward boundaries can have dramatic effects on segregation measures. A city in which wards are drawn to minimize the voting power of black residents by dispersing their votes across

²This is an important point to consider for our purposes. As discussed in the following sections, our segregation measure is dependent on the unique availability of the original manuscript pages of the 1880 federal census in digital form. Consequently, we can estimate only a single cross-section of our segregation measure. However, the persistence of the relative levels of segregation across cities noted by Cutler et al. (1999) suggests that patterns we identify for 1880 may provide information on patterns of segregation in subsequent time periods as well. The future availability of complete census manuscript files will allow us to construct estimates of our segregation measure from 1900 to 1940 in the near future.

wards may appear to be highly integrated. If the same city had wards drawn to make it easier to discriminate in the provision of public services by placing all black residents in the same ward, it would appear completely segregated according to the segregation measures.

The sensitivity of segregation measures to the way in which boundaries are drawn, which itself could be the product of or cause of segregation, makes it difficult to interpret any observed variation in segregation across cities without accounting for the historical context in which boundaries were drawn. For example, consider the case of Richmond, in the which the ward lines were drawn to include over one third of the city’s black population within the Jackson Ward, making that ward 80 percent black in 1880 (Rabinowitz, 1996, page 98). The efforts to minimize black voting power through gerrymandering were publicly discussed in cases such as Raleigh, where the Republican leaders advised black residents to move to the Fifth Ward which had not been gerrymandered. The local newspaper noted that blacks attempting this would find it “difficult to get houses when it is known they move only to carry the election and keep control of a much plundered city” (from the *Daily Sun* as quoted in Rabinowitz (1996, page 105)).

While Cutler et al., Collins & Margo and Troesken demonstrate that these traditional measures of segregation can be applied to historical data, the studies also highlight the limitations of these measures. Estimation of either index requires observing variation in the racial composition among geographical subunits making up a larger geographical unit of interest. Even with the problems noted above, cities have a somewhat natural subunit of wards or, in more recent decades, census tracts. Rural counties do not have a comparable subunit.³

The index of dissimilarity and the index of isolation therefore allow us to understand historical levels of segregation in cities but not in the areas surrounding those cities or in rural counties. This presents a severe limitation to our understanding of segregation and how it has evolved over time. As Figure 1 shows, the overwhelming majority of the population lived in rural areas in the late 1800s and early 1900s. If we want to understand how segregation influenced outcomes prior to World War II, it

³Lichter et al. calculate dissimilarity for rural areas using 1990 and 2000 census data using census blocks. They find that the pattern of segregation in rural communities is similar to the pattern in urban communities. African Americans are the most segregated racial group in both rural and urban areas. They note that the highly aggregated nature of the census block in rural communities limits their ability to speak to the forces shaping the segregation patterns they observe.

is essential to understand how segregation operated in rural areas. This is the experience relevant to roughly half of the white population until 1910 and over half of the black population until 1940. As much of the analysis of segregation patterns concerns black migratory patterns, the segregation levels in the sending communities of individuals who migrated to urban areas during the Great Migration is important. The index of isolation and index dissimilarity can tell us how segregation in cities changed with the influx of these individuals but they cannot tell us how segregation in rural areas contributed to that migration and changed as a result of it.

3 A New Measure of Segregation

Our measure is an intuitive approach to residential segregation. We assert that the location of households in adjacent units can be used to measure the degree of integration or segregation in a community, similar to Schelling’s classic model of household alignment. Indeed, we take to heart the first Schelling model, where households are aligned on a line with neighbors. Areas that are well integrated will have a greater likelihood of opposite race neighbors that corresponds to the underlying racial proportion of households in the area. The opposite is also true— segregated areas will have a lower likelihood of opposite race neighbors than the racial proportions would predict.

This measure does not suffer from the limitations of using political boundaries for geographical subunits and in fact does not require geographical subunits at all, making it possible to look at segregation in *any* geographical area, a key innovation of our approach. Our measure relies on the individual-level data available in federal census records. With the 100% sample of the federal census available through the Minnesota Population Center’s Integrated Public Use Microdata Series (IPUMS), it is possible to identify the races of next door neighbors. Rather than asking whether an individual lives in a ward or tract with many black residents, a question that hinges on how wards or tracts are defined, we can ask whether an individual lives next to a black or white neighbor, a question that can be consistently and universally applied to all households.

While the most celebrated aspect of Schelling’s model is that very small preferences for same-race neighbors could lead to complete segregation, his model also implies the measure that we use here.

At its heart, the Schelling concept of segregation was based on next-door neighbors.⁴ The popular discussions of segregation and preferences for racial integration, particularly in survey data, use neighbors as the criteria. While studies such as Clark (1991) ask respondents about racial proportions, a more standard approach is Farley et al. (1997), which shows examples of a neighborhood layout and a reference household or Zubrinsky & Bobo (1996b,a), which elicits preferences for same race neighbors. As such, our measure of segregation is the most aligned to the definition of residential segregation.

Our approach has a number of additional advantages. First, we focus on households as opposed to the population. The degree of residential segregation depends on the number of *households* of different types, not the number of individuals. If members of one group have larger household sizes or different household structure (for example, more likely to live in multiple generation households) there will be a difference between the population share and the household share. Household structure and size are known to vary by race historically and at present (Ruggles, 1994). Another advantage is that this measure is also an intuitive proxy for social interactions. Neighbors are quite likely to have some sustained interactions with each other, and an increasing likelihood of opposite race neighbors implies that the average level of interactions across racial lines would be higher. Indeed, social interaction models of segregation are inherently spatial and assume that close proximity is related to social interactions (Echenique & Fryer, 2007; Reardon et al., 2008).⁵

Specifically, the measure compares the observed number of black households in a area living next to a white neighbor to the predicted number given the overall racial composition of the area. We calculate the predicted number of black households with white neighbors given the number of black and white households in the area assuming that households are randomly located by race and assuming that households are completely segregated (only the households on the edge of the all black community have white neighbors). The segregation measure is then simply an estimate of how far the

⁴In the classic formulation, households had preferences over the race of their neighbor and their neighbor’s neighbor.

⁵We are careful to stress that our approach to segregation is focused on a measure of residential segregation. We do not propose a model of optimal household location choice as household location decisions are a function of their own preferences and the location of neighbors of preferred type. The problems of aggregating such measures over a community are further compounded by inherent geographic differences in locations that may be related to household location decisions. Here, our concern is deriving an intuitive measure that captures a key feature of residential living patterns by race. Models which apply a network approach to segregation assume that a person’s social network is closely tied to their residential living pattern, and without direct information on the actual social network such measures may not capture actual exposure to more or less segregated individuals.

actual number of black households with white neighbors is between these two extremes. In essence, our measure is a counterfactual between the observed and hypothetical distribution of households in a given area.

3.1 Next Door Neighbors and Census Enumeration

We exploit a feature of historical census enumeration to derive our segregation measure. Census enumerators went door-to-door to survey households. This implies that the position on the manuscript census form gives us the best possible measure of the actual location and composition of households as one would "walk down the street" from residence to residence. Proximity in the manuscript census form is, by design, a measure of residential proximity. We assume that adjacent appearance on the manuscript census form as evidence of being neighbors.⁶

There are several historical facts which support this assumption (Magnuson & King, 2005; Magnuson, 2009). First, enumerators were expected to be from the districts they were enumerating and to be familiar with the area and its residents. Second, the official training of enumerators specifically required an accurate accounting of dwellings containing persons in order of enumeration. A personal visit to each household was required. As enumerators were allowed to obtain information when household members were not present, it is highly likely that ordering in the census was in alignment to actual living patterns. Third, enumeration was publicly checked—after enumeration, census law required the public posting of each enumeration for public comment and correction. Specifically, enumeration was to be publicly posted for comment for a period of several days. In some large cities local newspapers and other local contingents checked the early returns for accuracy. This was allowed to ensure that complete counting was performed and also to ensure that household assignment of individuals was correct. Fourth, enumeration was often cross-checked with external sources such as voting records and other municipal information that would be recorded in sequenced household order. Fifth, the accuracy of the records had to be ascertained before the enumerator received payment. Census

⁶An obvious concern for rural communities would be the distance between neighbors identified in census manuscript files. Since enumeration districts were quite compact, even for rural areas, these adjacent households were closer than one may assume. Those at quite a distance would be placed in a different enumeration district. A second consideration is that African Americans were largely landless—they were usually not living on independent farms but rather more likely to live in compact tenant farming communities (Ransom & Sutch, 2001).

officials adopted a rough tracking system that allowed them to detect gross over- or under-counting of households.

Moreover, the monitoring and training of census takers became standardized with the tenth decennial census (in 1880). For our purposes, the advances in census enumeration beginning in 1880 are key. Earlier censuses are known to be controversial and demographic historians dispute their accounting of the population. In fact, the misreporting on census forms and the public outcry against the 1870 returns prompted reforms in the appointment, training and monitoring of census takers. For example, census officials did deny the appointment of enumerators who were politically connected or judged to be unqualified. This is very important for the South, as early enumeration (in 1870) was criticized as enumerators did not venture to homes as required by law. (Magnuson, 2009) notes that “[beginning in] 1880, ‘they went from cabin to cabin and did what the census laws require -paid personal visits to every place where it was likely that a person could find shelter.’” In northern and urban areas contemporary commentary concerning census enumeration in 1880 was positive— enumerators were found to be careful and thorough.⁷ In general, the policies and procedures of enumeration since 1880 give us confidence that our approach is the best available proxy for household location by race.

3.2 Deriving the Segregation Measure

Construction of the measure begins by identifying neighbors in the census. The complete set of household heads in the census are sorted by reel number, microfilm sequence number, page number and line number. This orders the household heads by the order in which they appear on the original census manuscript pages, meaning that next-door neighbors appear next to one another. There are two different methods for identifying each household head’s next-door neighbors. The first is to simply define the next-door neighbors as the household head appearing before the individual on the census manuscript page and the household head appearing after the individual on the census manuscript page. An individual that is either the first or last household head on a particular census page will only have one next door neighbor identified using this method.

Naturally, one must be particularly careful to test the proposition that adjacency is a measure of

⁷Some locations rescinded this praise when final counts revealed population levels lower than expected.

neighbor status. To allow for the next door neighbor appearing on either the previous or next census page and to account for the possibility that two different streets are covered on the same census manuscript page an alternative method for identifying neighbors is also used that relies on street name rather than census manuscript page. In this alternative measure next-door neighbors are now identified by looking at the observations directly before and after the household head in question and declaring them next-door neighbors *if and only if* the street name matches the street name of the individual of interest (and the street name must be given, two blank street names are not considered a match). This approach has the advantage of finding the last household head on the previous page if an individual is the first household head on his census manuscript page or the first household head on the next page if the individual was the last household head on a manuscript page. However, the number of observations is reduced substantially relative to the first method because many individuals have no street name given. Few roads had names in historical census records. This is particularly true in rural areas.

Once next door neighbors are identified, an indicator variable is constructed that equals one if the individual has a next door neighbor of a different race and zero if both next-door neighbors are of the same race as the household head.⁸ As described above, two versions of this indicator variable are used, one in which all observations are used, one in which only those observations for which both next-door neighbors are observed are used. This latter version reduces the sample size but, for the remaining individuals, gives a more accurate measure of the percentage of individuals with a neighbor of a different race.

Formally, we begin with the following:

- b_{all} : the total number of black household heads in the area
- $n_{b,B=1}$: the number of black household heads in the area with two observed neighbors
- $n_{b,B=0}$: the number of black household heads in the area with one observed neighbor

⁸Based on the race assigned at enumeration. This is similar to the *Racesing* coding of race constructed by IPUMS. One key feature of *racesing* for our purposes is places people with their race given as ‘mulatto’ in the same category as people with their race given as ‘black’. So a black individuals living next to two neighbors listed on the census as mulatto would be considered to be of the same race as his neighbors. In the current version of the segregation measure, the sample is restricted to only black or white household heads. Consequently saying a black household head has a neighbor of a different race is the equivalent to saying he has a white neighbor.

- x_b : the number of black household heads in the area with a neighbor of a different race

The equivalent for the set of white household heads are similarly defined.

Given these measures, the basic measure of segregation is calculated as the distance the area is between the two extremes of complete segregation and the case where neighbor's race is entirely independent of an individual's own race. There are a total of four versions of the segregation measure. Each of these measures corresponds to one of the two different methods of defining next-door neighbors (whether the specific street of residence is identified on the census manuscript form) and whether all individuals with a neighbor present are included or only those individuals with both neighbors identified are used.

In the case of random neighbors, the number of black residents with at least one white neighbor will be a function of the fraction of black households relative to all households. In particular, the probability that any given neighbor of a black household will be black will be $\frac{b_{all}-1}{(b_{all}-1)+w_{all}}$. The probability that the second neighbor will be black if the first neighbor is black will then be $\frac{b_{all}-2}{b_{all}-2+w_{all}}$. The probability that a black household head will have at least one white neighbor can be written as a function of these probabilities by expressing it as:

$$p(\text{white neighbor}) = 1 - \left(\frac{b_{all} - 1}{b_{all} - 1 + w_{all}} \right) \left(\frac{b_{all} - 2}{b_{all} - 2 + w_{all}} \right) \quad (3)$$

where the second term comes from the assumption that the races of adjacent neighbors are uncorrelated, a reasonable assumption given that we are considering randomly located neighbors. The expected value of x_b under random assignment of neighbors would then be:

$$E(\bar{x}_b) = p(\text{white neighbor}) \cdot n_b \quad (4)$$

$$E(\bar{x}_b) = n_b \left(1 - \left(\frac{b_{all} - 1}{b_{all} - 1 + w_{all}} \right) \left(\frac{b_{all} - 2}{b_{all} - 2 + w_{all}} \right) \right) \quad (5)$$

The calculation of this upper bound on x_b must be modified slightly when including household heads for which only one neighbor is observed. In this case, the expected number of black household heads with a white neighbor under random assignment of neighbors will be composed of two different

terms, the first corresponding to those household heads with both neighbors observed and the second corresponding to those household heads with only one neighbor observed. Letting B be an indicator variable equal to one if both neighbors are observed and equal to zero if only one neighbor is observed, the expected total number of black household heads with a white neighbor is then:

$$E(\overline{x_b}) = p(\text{white neighbor} | B = 1) \cdot n_{b,B=1} + p(\text{white neighbor} | B = 0) \cdot n_{b,B=0} \quad (6)$$

$$E(\overline{x_b}) = n_{b,B=1} \left(1 - \left(\frac{b_{all} - 1}{b_{all} - 1 + w_{all}} \right) \left(\frac{b_{all} - 2}{b_{all} - 2 + w_{all}} \right) \right) + n_{b,B=0} \left(1 - \frac{b_{all} - 1}{(b_{all} - 1) + w_{all}} \right) \quad (7)$$

Under complete segregation, the number of black individuals living next to white neighbors would simply be two, the two individuals on either end of the neighborhood of black residents, giving a lower bound for the value of x_b . However, it is necessary to account for observing only a fraction of the household heads. The expected observed number of black household heads living next to a white neighbor when sampling from an area with only two such residents will be:

$$E(\underline{x_b}) = p(\text{observe one of the two in } n_b \text{ draws}) \cdot 1 + p(\text{observe both in } n_b \text{ draws}) \cdot 2 \quad (8)$$

$$E(\underline{x_b}) = \frac{1}{\frac{1}{2}(n_b + 1)} \left(1 - \prod_{i=0}^{n_b-1} \frac{b_{all} - i - 2}{b_{all} - i} \right) + 2 \left(1 - \frac{1}{\frac{1}{2}(n_b + 1)} \right) \left(1 - \prod_{i=0}^{n_b-1} \frac{b_{all} - i - 2}{b_{all} - i} \right) \quad (9)$$

The product in the expression above gives the probability of selecting neither of the two black household heads with white neighbors in n_b successive draws from the b_{all} black household heads. Thus one minus this product is the probability of drawing either one or both of the two household heads with white neighbors. Note that the product notation is used above because it makes it easier to see how the probability is being derived. In practice, the product reduces to $\frac{(b_{all}-n_b)(b_{all}-n_b-1)}{b_{all}(b_{all}-1)}$. The ratio $\frac{1}{\frac{1}{2}(n_b+1)}$ gives the fraction of these cases that correspond to drawing just one of the two household heads with white neighbors. This comes from noting that with n_b draws, that there are n_b ways to draw one of the two household heads while there are $\sum_{i=1}^{n_b-1} (n_b - i)$ or $n_b(n_b - 1) - \frac{(n_b-1)n_b}{2}$ ways to draw both of the household heads.

Finally, in the case where household heads with only one observed neighbor are included, it is necessary to account for the probability that a black household head with a white neighbor will be

drawn but that white neighbor is not the observed neighbor. The expected value of x_b accounting for the probability that the white neighbor is unobserved for a household head with only one observed neighbor is:

$$E(\underline{x}_b) = \left(\frac{n_{b,B=1}}{n_b} + \frac{n_{b,B=0}}{n_b} \cdot \frac{1}{2} \right) \quad (10)$$

$$\cdot \left[\frac{1}{\frac{1}{2}(n_b + 1)} \left(1 - \prod_{i=0}^{n_b-1} \frac{b_{all} - i - 2}{b_{all} - i} \right) + \right. \quad (11)$$

$$\left. 2 \left(1 - \frac{1}{\frac{1}{2}(n_b + 1)} \right) \left(1 - \prod_{i=0}^{n_b-1} \frac{b_{all} - i - 2}{b_{all} - i} \right) \right] \quad (12)$$

In this equation, the fraction of black household heads with only one observed neighbor, $\frac{n_{b,B=0}}{n_b}$, has its expected value of x_b reduced by an additional factor of $\frac{1}{2}$ to account for the fact that if one of these individuals is one of the two black household heads living next to a white neighbor there is only a 50 percent chance that the white neighbor is the observed neighbor.

The degree of segregation in an area, α , can then be defined as the distance between these two extremes, measured from the case of no segregation:

$$\alpha = \frac{E(\overline{x}_b) - x_b}{E(\overline{x}_b) - E(\underline{x}_b)} \quad (13)$$

This segregation measure increases as black residents become more segregated within an area, equaling zero in the case of random assignment of neighbors (no segregation) and equalling one in the case of complete segregation.⁹

4 Simulations of the Segregation Measure

To confirm that our measure is accurately reflecting segregation as we have defined it, the distance an area is between the extremes of randomly assigned neighbors and completely segregated neighbors, we

⁹Note that it is possible for this measure to be less than zero if the particular sample of household heads is actually more integrated than random assignment of neighbors. For example, suppose every other household head on the manuscript pages were black in an area that is 50 percent black. With random assignment of neighbors we would expect to observe at least some black household heads having black neighbors. In this case, x_b would be larger than $E(\overline{x}_b)$ making α negative. The measure can also exceed one in the rare cases where only zero or one black household heads with a white neighbor are observed. In these cases x_b may actually be smaller than $E(\underline{x}_b)$.

have run a series of simulations to check that these two benchmarks are being properly calculated as the number of black households, the racial composition of the area and the overall population of the area vary. Simulated areas are generated containing between 20 and 1000 white households in increments of 20 households. For each particular number of white households, areas are simulated containing between one and 1000 black households in increments of one household. For each combination of white and black households, we calculate the number of observed black household heads living next to a white neighbor under complete segregation and under no segregation given a particular level of missing households.

For the case of no segregation, we generate a random number for each household and then sort the households on the basis of this number. Neighbors are defined as households next to each other on this sorted list. This gives us neighbor locations that are completely independent of race. We then randomly draw the appropriate number of households given the chosen percent missing and count the number of black households with white neighbors in this sample. For the case of complete segregation, we randomly choose two black households as the two households with white neighbors (the households on either side of the black neighborhood). We then randomly draw the appropriate percentage of households based on the fraction of households that are missing and see whether one or both of the households with a white neighbor is in the randomly drawn sample.

Both of these calculations are repeated for 1000 different draws of random numbers generating 1000 simulated areas for each particular combination of black and white households. The result is 1000 observations of the number of observed black households with white neighbors and the number of observed black households with no white neighbors under complete segregation and under no segregation for each combination of the total number of black households and total number of white households in an area. These values let us calculate our segregation measure and check whether the value is equal to one on average for the completely segregated simulated area and equal to zero on average for the areas with no segregation.

Figure 2 shows the mean, 5th percentile and 95th percentile for these simulated values of the segregation measure by number of black households and by percent black. We include graphs for both simulations in which five percent of the observations are missing and simulations in which

twenty percent of the observations are missing. For all of the graphs, we use observations with either one or both neighbors observed (the results look quite similar when restricting the sample to only those observations with both neighbors observed). From the graphs it is clear that our measure is well behaved, equalling one on average when an area is completely segregated and zero on average when an area has no segregation.

One feature worth noting is that the measure is less well behaved when the number of black households is very small. At very small numbers of black households (typically at fewer than five black households) it becomes difficult to distinguish between randomly located households and segregated households since the majority of black households will have a white neighbors in either case. This is a natural product of the fact that segregation is difficult to define without critical masses of both groups. Once there are over five black households, however, we get a clear distinction between the segregated and unsegregated cases.

The values of α vary a fair amount for completely integrated counties, particularly for counties with a higher percentage of black residents. This is a product of the number of black households located next to white households varying a fair amount when households are randomly assigned, causing deviations from the expected number of black households with white neighbors. For the completely segregated counties, the only variation comes from whether the two black households with white neighbors are observed, leading to far fewer (and far smaller) deviations from the expected number. Even with this inherent variation, the measure captures the degree of residential segregation as intended. Overall, the simulations show that our measure of segregation accurately reflects the racial residential patterns in the underlying community.

5 Comparison of Segregation Measures

Our measure begins from a fundamentally different unit of analysis than existing measures, making it difficult to perform a direct comparison of the methods. Analytically, aggregating our measure to the level of the census tract and block reveals different information about segregation in the subunit than the population shares used in traditional measures. The traditional measures, at their base,

require only population shares by race, while our measure uses alignment and is not hierarchical. Subunit differences in segregation measure would reflect subtle, but potentially important, differences in spatial distribution.¹⁰

A useful example of this problem would be schools. Suppose that a school district had a large number of schools and students were of only two races, white and black, each of whom was fifty percent of the student population. If each school was fifty percent white both isolation and dissimilarity, defined at the school level (the subunit of the school district) would imply that the school district was integrated. If every *classroom* were segregated, however, no student *within* a school would have a classmate of a different race. Our segregation measure, defined from the likelihood that the student next to you is of a different race, would capture this segregation. Since our measure, calculated both within schools and for the entire school districts, reveals different information the analytical comparison is not as informative as we would like. Indeed, under this extreme example our segregation measure would predict complete segregation while the traditional measures would imply complete integration.

To assess how our segregation measure compares to traditional measures, we calculate our segregation measure, the index of dissimilarity and the index of isolation using the federal census. We compare our measure to traditional measures to see if our approach reveals new insight into the pattern of segregation. Our unit of analysis throughout is the county. We choose counties as the unit because it allows us to analyze the differences in segregation between urban and rural areas, counties are well-defined civil jurisdictions and additional information is available at the county level which allows us to analyze the correlates of segregation using our measure in addition to traditional measures.

As noted earlier, dissimilarity and isolation are typically only calculated at the city level using wards as the geographic subunit for the calculation. Given our interest in applying our measure to both urban and rural counties, we cannot take this approach. Rural areas do not have such subdivisions. We need a geographic subunit that will be available for both urban and rural areas

¹⁰One critique of traditional segregation measures is that they may fail to capture the emergence of separate neighborhoods housing members of the same race but of different socioeconomic status. For example, Bayer et al. argue that increasing wealth among African Americans led to the development of middle-class black communities which increased measures of segregation.

for comparison. One of the few candidates for such a unit is the census enumeration district. The enumeration district is on average a smaller unit in terms of population than a ward but still contains several hundred households, on average.¹¹ The typical rural enumeration district in the 1880 census contains 350 households while the typical urban enumeration district contains 450 households.¹² The mean number of enumeration districts in a rural county is 10 while the mean for urban counties is 39. Given that ward-level data is not available for rural counties and that the values of the traditional segregation measures vary with the fineness of the geographical subunit, we calculate the traditional segregation measures using enumeration district as the subunit for both urban and rural counties in order to make meaningful comparisons across counties. A key advantage of enumeration districts is that they were designed to maintain the boundaries of civil divisions (towns, election districts, wards, precincts, etc.). The use of enumeration districts guards against finding differences between the measures that are simply the product of higher level aggregation (calculating dissimilarity and isolation over a larger area) as opposed to actual differences in living arrangements by race.¹³

We calculate our segregation measure at the county level and dissimilarity and isolation at the county level using enumeration districts as the subunit.¹⁴ Figure 3 depicts the variation in our segregation measure and the traditional measures for rural counties across regions. The figure depicts ranges of the measures with the end points of the range being one standard deviation above and below the mean. For simplicity, the results presented in this section focus only on the samples using the manuscript page definition of neighbors and requiring that at least one neighbor’s race be observed. This gives us a larger number of households producing less noisy data, particularly for counties with very small numbers of black households overall.¹⁵ When calculating the means and standard deviations, counties are weighted by the number of black household heads to provide a more accurate

¹¹An enumeration district is actually more comparable in size to a census block, the geographical subunit used by Echenique & Fryer (2007), than a ward or census tract.

¹²Enumeration districts averaged roughly 1,500 persons.

¹³Note that on average, since the enumeration districts are smaller units than wards, our estimates of dissimilarity and isolation in urban counties will tend to be higher than those of Cutler et al. (1999) and Troesken (2002).

¹⁴To make the sample used for calculating the index of isolation and the index of dissimilarity comparable to the sample used in the calculation of our segregation measure, we drop all household heads for which race is not observed and neighbor’s race is not observed. As with the calculation of our segregation statistic, this leads to four different samples: all household heads for which at least one neighbor’s race is observed using the manuscript page definition of neighbor, only those household heads for which both neighbors’ races are observed using the manuscript page definition of neighbor, and both of these samples using the street name definition of neighbor.

¹⁵Comparisons of the segregation measures when using the other sample restrictions are available upon request.

picture of the experience of the typical black household and to minimize the effects of outlier counties with only one or two black households. The figure focuses on only those regions in which over one percent of the population is black. Means and standard deviations for the measures across all regions are in Table 1 giving unweighted values and Table 2 giving values weighted by the number of black households. These tables also include means and standard deviations for the urban counties.

For a more detailed view of how the geographical distribution of segregation varies by measure, maps of the United States and maps of the regions where blacks constitute more than one percent of the population are given in Figure 4 and Figure 5, respectively. The most striking feature is that the index of dissimilarity shows the North and, more generally, areas with a low percentage of black residents as more segregated on average while our measure identifies the South as more segregated (but not necessarily the areas of the South with dense populations). That is, the percent black does not reveal the same spatial pattern of segregation as our neighbor-measure does. Also worth noting is that there is a distinct, discontinuous change in the index of dissimilarity when moving from the South to the North; the southern borders of Pennsylvania, Ohio and Indiana in particular stand out. These patterns are likely due to differences across states in the way enumeration districts are drawn. The index of dissimilarity is highly sensitive to the way these subunits are defined. Our measure, based on individual-level data, does not depend on these definitions of enumeration districts and shows a much more gradual transition in levels of segregation across space.

These figures reveal a substantial amount of heterogeneity in segregation within regions, across regions and between urban and rural areas. However, the data also reveal that the patterns of segregation depend heavily on the chosen measure of segregation. The rankings of regions in terms of how segregated they are and the differences in segregation between rural and urban counties differ significantly depending on the measure. To get a better sense of how the measures relate to one another, correlations between the measures are provided in Table 3. Our measure is positively correlated with the percentage of households who are black and with the index of isolation. Surprisingly, our measure is negatively correlated with the index of dissimilarity for both rural and urban counties. However, after weighting by the number of black households in each county this correlation turns positive.

In general, the correlations in Table 3 show that our measure is weakly correlated with traditional

measures of segregation. This is likely due to the fineness of our measure as opposed to the groupings required of traditional measures. For example, Echenique & Fryer propose a spectral index of segregation that is well correlated with the percent black (.90) and isolation (.93), but less well correlated with dissimilarity (.42). Our measure is substantially less well correlated with any of these measures (.43, .70 and .29 for percent black, isolation and dissimilarity respectively) but does share the same general pattern of correlations. Given the relatively low correlation of our measure of segregation with other indices of segregation and the simulation results which show that our measure does reflect underlying segregation patterns, it appears that our measure captures a dimension of segregation not reflected in traditional segregation indices.

6 Historical Predictors of Segregation

The previous section established that our measure of segregation was distinct from other segregation measures. We conjecture that one key reason for the distinction is the fact that other measures are inherently dependent on geographic boundaries which can give very different inference for segregation. Since our segregation measure captures a feature of segregation that is missing from other measures, below we consider the proximate determinants of our segregation measure. We stress that these results are preliminary and should be interpreted with caution—they are an attempt to see how this measure of segregation varies with antebellum demographic and economic characteristics of counties.

We begin by considering the historical demographic predictors of the segregation measure for 1880. Living arrangements by race were doubtlessly influenced by the antebellum living arrangements and changes from the plantation to tenant farming systems (Kaye, 2007). At the same time, significant flows of African Americans from rural areas in the south to southern cities also would have influenced living patterns (Boles, 1984). Immediately after the end of the Civil War there was a substantial movement of African Americans as they sought to unite families and create communal bonds denied during chattel slavery Sterling (1976). Even so, the majority of African Americans lived in the rural south at the time of our census measure. With our unique measure of segregation, we can now investigate how features of the antebellum era influenced living patterns fifteen years later.

We present preliminary results of historical demographic predictions of the segregation measure in Table 4, which shows the demographic and economic predictors of the segregation measure. The results show that the percent of the county that was slave in 1860 is positively related to the segregation measure. A one standard deviation increase in the percent of the county population that is slave moves the segregation measure by .58 standard deviations, a sizeable effect. The relationship turns negative when the black percentage in 1880 is included, however. This most likely reflects the effects of the strong correlation in the percentage of black households over time in southern counties. The most interesting results show that segregation is positively related to the percent free black in a county in 1860, but negatively related to the percent of the black population that was free in 1860. A one standard deviation increase in the percent of the black population that was free in 1860 decreases the segregation measure by .50 standard deviations while a one standard deviation increase in the percent of the total population that is free black increases the segregation measure by .19 standard deviations. This likely reflects the fact that our measure can detect the development of mutually exclusive homogeneous communities, a feature only recently uncovered in the historical literature (Kaye, 2007). It is well known, however, that free blacks lived in separate communities in the antebellum era (Wade, 1967). After emancipation, ex-slaves would develop their own communities which may not be contiguous to free blacks. This would give rise to two different areas which, by design, decreases the segregation measure as it increases the chances that blacks would have white neighbors.

The relationship between economic variables and the segregation measure are presented in Table 5. The results show that cotton production, measured in bales of ginned cotton, is well correlated with the segregation measure. Both cotton production before and after the Civil War are correlated with more segregated areas. While the cotton production measures are well correlated with segregation, the relationship is not substantively large. A one standard deviation in cotton production in 1869 increases the segregation measure by less than .20 standard deviations. When looking at changes in cotton production the relationship is largely the same quantitatively. The growth of cotton production from 1850 to 1860 is positively related to segregation, but growth of cotton production between 1860 and 1870 is negatively related to segregation. Similarly, growth in the acres of improved farmland between

1850-1860 and 1860-1870 are negatively correlated with the segregation measure. This implies that counties where more land was placed into agricultural production were correlated with more integrated environments. The growth of the value of farms, however, is positively related to segregation.¹⁶

In short, our segregation measure in 1880 is related to population characteristics such as the county fraction that is slave, free and the fraction of the black population that is free. Economic and agricultural measures are not strongly related to the segregation measure. While they are related statistically, their explanatory power pales in comparison to the demographic measures. In some respects, it is difficult to say whether these results confirm what we would expect about living arrangements by race, as with few exceptions, notably Kaye (2007), we know very little about the formation of slave neighborhoods in the antebellum era and the degree to which individual living patterns reflected local economic conditions. Given the destructive nature of the Civil War and the upheaval in the society in general, it is unclear how much of the 1880 segregation measure would reflect antebellum conditions. From the preliminary results in Table 4 and Table 5 it appears that demographic antebellum measures are more strongly related to segregation than economic measures. We take this as suggestive evidence that our measure captures the underlying demographic patterns of community and race—a measure based on neighbors should be strongly correlated with racial demographic characteristics and less correlated with economic factors, *per se*.

7 Segregation and Mortality

As our measure captures a feature of individual living patterns, we turn our attention to individual level results. Contemporary demographic scholarship has established that segregation is related to health outcomes for African Americans (Williams & Collins, 2001, 1999; Jackson et al., 2000). Several potential pathways have been suggested: African Americans in more segregated environments may

¹⁶Additional results (unreported) showed that other features that could be arguably linked to African American residential patterns were not shown to be significantly related to our segregation measure. For example, Fogel & Engerman assert that large plantations, particularly those with more than 15 slaves, were the key to southern agricultural efficiency. Ransom & Sutch suggest that living patterns after the war were quite related to slave quarter arrangements. While no longer living as close to the plantation home, ex-slaves and their descendants lived in racially distinct communities that mirrored that structure. We find that the relationship between our segregation measure and the number of large slave-owning farms is insignificant. When the segregation measure is regressed on the number of farms with more than 15 slaves per county the regression yields a coefficient of -2.26e-06 (s.d. 2.88e-06). We obtain similar results for farm size and for the effects of very large plantations (those with more than 50 slaves).

be more stressed, have less access to medical services, and may be subjected to environmental insults such as poorer air quality. Given the limited measures of residential segregation available for rural residents, there has been little analysis of the role of segregation in explaining mortality patterns in the past. Below, we consider several historical data sources to uncover the relationship between segregation and mortality in the past.

7.1 Mortality and Segregation in Antebellum Cohorts

We continue by looking at the relationship between segregation and mortality. We exploit data on the Colored Troops of the Civil War. These men were born in 1840, on average, so the segregation measure in 1880 is not related to their birth outcomes (and determining birth location for many of these troops is difficult— see Logan (2009)). The location of a veteran after the war is related to two outcomes— their longevity and their pension receipt. We show the results in Table 6.

In the first column of the table we see that the segregation index is positively related to age at death. In the first three columns we add additional controls that would be related to the health and mortality of colored veterans of the Civil War. When segregation is added to the specification the result is highly significant. A one standard deviation increase in the segregation index for these veterans increases their length of life by 0.054 standard deviations.

In the next four columns we consider determinants of the pension amount awarded to colored veterans. Unlike the white Civil War veterans, political variables related to pension receipt will not hold for the colored troops since they lived in the South and were largely disenfranchised by the time pension legislation was adopted. An additional issue for African American veterans was how to find a doctor who would certify their illnesses and injuries to qualify them for a pension. Racial discrimination in medical services was acute at this time. Segregation may be related to the probability of obtaining a pension and the amount of the pension that a veteran received if it was related to a stronger social network that would give them better information about how to apply for a pension and which doctors would be most favorable to their conditions. In this way, it is likely that segregation had a positive effect. It is important to note that these veterans applied for pensions after 1880, so our segregation measure refers to their location at the time of pension application. Even so,

we find that segregation has a large and positive effect on the amount of a pension that a veteran received. These preliminary results suggest that the racial residential pattern of the location of the veteran after the war, and not the racial composition itself, was well correlated with the pension amount received.

7.2 Mortality and Segregation in Postbellum Cohorts

To consider the effects of segregation on mortality for cohorts born after the Civil War, we use a new data source: the death certificates from the state of Missouri. We match the counties of birth in the death certificates to the segregation measure. Since the majority of those in the death records are born after the date of the measure, we posit that the segregation measure captures the degree of segregation in the community into which one was born. As this is the complete record of deaths recorded in Missouri during the time period (1880 to 1909), we use the simple regression where the dependent variable is the length of life.¹⁷ A particular advantage of the Missouri data is that the African American population in the state is relatively large (more than ten percent of households in the state are African American) and geographically dispersed.

In Table 7 we analyze lifespan and lifespan conditional on survival to age 10. The results for lifespan show that those born in more segregated counties lived four fewer years, on average. From the third column we find that a one standard deviation increase in the segregation measure (for Missouri only), lessens the lifespan by 0.07 standard deviations. When we interact race and segregation, to see if the effects of segregation differ by race, we find that segregation had a large and negative correlation with mortality. A one standard deviation increase in the segregation measure lessens the lifespan for African Americans by more than .10 standard deviations, a large effect. While dissimilarity and isolation are correlated with length of life, their relationship is not as strong substantively or statistically as the effect of segregation. The mortality results are not driven by infant mortality. In Table 7 the effect remains when we exclude deaths to those under age 10. Indeed, the effect is largely the same— a one standard deviation increase in the segregation measure reduces the lifespan by .10

¹⁷There is some literature that the ages of death for African Americans are reported with errors on death records. However, the previous analysis has shown that the errors lead to African Americans having longer lifespans as ages of the elderly at death are upward biased.

standard deviations. In general, these results for one state show that segregation was strongly related to mortality outcomes for African Americans in the early twentieth century.

7.3 Infant Mortality and Segregation

We use 1900 and 1910 census data to construct measures of child mortality.¹⁸ We restrict our attention to women in southern states who have had at least one birth. From this we derive the number of children who have died for each mother and whether or not a woman has lost a child. Given the differences in disease environments and health at this time, we divide the sample and estimate the relationship for rural and urban counties. In Table 8 and Table 9 we estimate a regression where the dependent variable is the number of children who have died for rural and urban mothers, respectively. The results show that segregated environments are positively correlated with the number of infant deaths. Interactions of segregation with race show that more segregated environments are not well correlated with racial differences in infant mortality. The result is less prominent in urban areas, however, suggesting that the effects of segregation are different when one lives in an urban versus rural location. We consider the extensive margin of child and infant mortality in Table 10 and Table 11. While the results for rural mothers shows that there is a strong relationship, the results for urban areas show that the extensive margin is not well correlated with segregation in urban areas. Given the critiques of estimates of census-based methods of inferring child and infant mortality, it is unclear whether the results could be driven by measurement error. Even with this caveat, it appears that segregation is related to the mortality of African Americans in the past.

8 Lynching and Segregation

Segregation has long been viewed as a factor in explaining racial violence in the past. DeFina & Hannon argue that city level measures of segregation are linked to contemporary segregation measures today. Bailey et al. argue that lynching victims were socially isolated, and Tolnay & Beck argue that racial violence was related to black migration patterns in the early twentieth century. The number

¹⁸See Keyfitz (1985) for various methods. We use the number of children born minus the number of children present to estimate child mortality.

of lynchings in the United States reached their peak in 1892 (Cook, 2012). As such, it is useful to investigate whether the number of lynchings varied with the measure of segregation based on neighbors. Recent historical scholarship by Loewen (2005), Jaspin (2007), and Kantrowitz (2012) has shown that lynching was only one small piece of a larger movement of racial violence in the United States in the late nineteenth century. This movement included the ethnic cleansing of entire counties and the prohibition of African American residence in certain towns (known as "sundown towns" as blacks found to be present after dark would be subject to death.) The full quantitative history of these events is still unknown. What is also still not known is the degree to which living patterns were correlated with racial violence.

In Table 12 we regress the number of lynchings per county on the segregation measure and the percent of households that were black in 1880. As lynching was highly differential by region we control for state fixed effects in all specifications. We use the existing national lynching database from the Historical American Lynching (HAL) project for the number of lynchings by county from 1882-1930. A key advantage here is that lynchings in the HAL data come from years after the 1880 census used to measure segregation. The results show that increases in segregation were strongly related to the number of lynchings per county. For the OLS results, a one standard deviation change in the segregation measure increases the number of lynchings per county by 0.10 standard deviations, a sizeable effect. The result is robust to non-linear models which account for the count nature of the segregation measure. The probit estimates suggest that the probability of a lynching was much higher in a county that was highly segregated. For example, the tobit estimates show that increases in the segregation measure were strongly related to lynchings per county. Indeed, a one standard deviation increase in the segregation measure results in an additional lynching in a county, on average. In general, the results of Table 12 suggests that our segregation measure was strongly related to lynching at both the extensive and intensive margins in the late nineteenth and early twentieth centuries.

One interesting feature of the results in Table 12 is that the traditional measures of segregation are not well correlated with the number or presence of lynchings in a county. In some respects, the results add quantitative support for the histories of racial cleansing offered by Jaspin (2007) and Loewen (2005). In particular, Jaspin (2007) notes that general county characteristics, such as racial makeup,

did not predict racial cleansing in a county, and Loewen (2005) found that towns which forbade African American inhabitants did not share observable characteristics. While boundary-based measures imply that segregation had no impact on lynching, the neighbor-based measure of segregation shows that an increasing likelihood of opposite race neighbors is correlated with less lynching activity.

9 Conclusion

In this paper we have derived a new measure of segregation from the complete 1880 census and used the simple criteria of the race of a neighbor. If neighbors are less much likely to be of a different race than random assignment would predict then that location is more segregated than another that is closer to random assignment. We first showed how this measure has distinct advantages over the most commonly used measures of segregation in the literature. Indeed, we showed that our measure is weakly correlated with the most commonly used measures of segregation. Next, we showed how segregation varied in significant ways in the United States in the late 19th century. There was substantial regional heterogeneity in segregation and between rural and urban areas.

We then turned to the effects of segregation on health, mortality and racial violence. At the individual level we found that segregation was a strong predictor of the pension receipt of African American Civil War Veterans, mortality, and racial violence. The surprisingly strong relationship holds despite our controls for the state, the other measures of segregation, and contemporaneous controls that should mute the effect of historical segregation *per se*.

At a minimum, this project shows that segregation in rural communities is an important part of the segregation story in the United States that should be investigated. As we noted earlier, more than three quarters of the population lived in rural areas in 1880. Understanding the relationship between segregation and urbanization would help us understand the dynamics of segregation in cities and rural communities in the twentieth century. These links have important implications for the skill mix of cities, public finance, education, health and other measures of social well-being. Better knowledge of segregation's past will give us the tools to outline its full impact in the past and, most important, the present.

References

- Almond, D., Chay, K., & Greenstone, M. (2006). Civil rights, the war on poverty, and black-white convergence in infant mortality in the rural south and mississippi. *MIT Department of Economics working paper*.
- Ananat, E. O. (2011). The wong sides(s) of the tracks: The causal effects of racial segregation on urban poverty and inequality. *American Economic Journal: Applied Economics*, 3(2), 34–66.
- Ananat, E. O., & Washington, E. (2009). Segregation and black political efficacy. *Journal of Public Economics*, 93(5-6), 807–822.
- Bailey, A. K., Tolnay, S. E., Beck, E., & Laird, J. D. (2011). Targeting lynch victims: Social marginality or status transgressions? *American Sociological Review*, 76(3), 412–436.
- Bayer, P., Fang, H., & McMillan, R. (2005). Separate when equal? racial inequality and residential segregation. Tech. rep., National Bureau of Economic Research.
- Boles, J. B. (1984). *Black southerners, 1619-1869*. University of Kentucky Press.
- Borjas, G. J. (1995). Ethnicity, neighborhoods, and human-capital externalities. *The American Economic Review*, 85(3), 365–390.
- Brooks-Gunn, J., Duncan, G. J., Klebanov, P. K., & Sealand, N. (1993). Do neighborhoods influence child and adolescent development? *American journal of sociology*, (pp. 353–395).
- Case, A. C., & Katz, L. F. (1991). The company you keep: The effects of family and neighborhood on disadvantaged youths. *NBER Working Papers*.
- Chay, K. Y., Guryan, J., & Mazumder, B. (2009). Birth cohort and the black-white achievement gap: The roles of access and health soon after birth. Tech. rep., National Bureau of Economic Research.
- Clark, W. (1991). Residential preferences and neighborhood racial segregation; a test of the schelling segregation model. *Demography*, 28(1), 1–19.
- Collins, W. J., & Margo, R. A. (2000). Residential segregation and socioeconomic outcomes: When did ghettos go bad? *Economics Letters*, 69(2), 239–243.
- Cook, L. D. (2012). Converging to a national lynching database: Recent developments and the way forward. *Historical Methods*, 45(2), 55–63.
- Cutler, D. M., & Glaeser, E. L. (1997). Are ghettos good or bad? *Quarterly Journal of Economics*, 112(3), 827–872.
- Cutler, D. M., Glaeser, E. L., & Vigdor, J. L. (1999). The rise and decline of the american ghetto. *Journal of Political Economy*, 107(3), 455–506.
- Cutler, D. M., Glaeser, E. L., & Vigdor, J. L. (2008). When are ghettos bad? lessons from immigrant segregation in the united states. *Journal of Urban Economics*, 63(3), 759–774.

- DeFina, R., & Hannon, L. (2011). The legacy of black lynching and contemporary segregation in the south. *Review of Black Political Economy*, 38(2), 165–181.
- Echenique, F., & Fryer, R. G. (2007). A measure of segregation based on social interactions. *The Quarterly Journal of Economics*, 122(2), 441–485.
- Farley, R., Fielding, E. L., & Krysan, M. (1997). The residential preferences of blacks and whites: a four-metropolis analysis. *Housing Policy Debate*, 8(4), 763–800.
- Fogel, R. W., & Engerman, S. L. (1974). *Time on the cross: The economics of American negro slavery*. Norton.
- Jackson, S. A., Anderson, R. T., Johnson, N. J., & Sorlie, P. D. (2000). The relation of residential segregation to all-cause mortality: a study in black and white. *American Journal of Public Health*, 90(4), 615–617.
- Jaspin, E. (2007). *Buried in the bitter waters: The hidden history of racial cleansing in America*. Basic Books.
- Kain, J. F. (1968). Housing segregation, negro employment, and metropolitan decentralization. *The Quarterly Journal of Economics*, 82(2), 175–197.
- Kantrowitz, S. D. (2012). *More than freedom: Fighting for black citizenship in a white republic, 1829-1889*. Penguin Books.
- Kaye, A. E. (2007). *Joining places: Slave neighborhoods in the old south*. University of North Carolina Press.
- Keyfitz, N. (1985). *Applied Mathematical Demography*. Springer-Verlag.
- Lichter, D. T., Parisi, D., Grice, S. M., & Taquino, M. C. (2007). National estimates of racial segregation in rural and small-town america. *Demography*, 44(3), 563–581.
- Loewen, J. W. (2005). *Sundown towns: A hidden dimension of American racism*. Touchstone Books.
- Logan, T. D. (2009). Health, human capital and african american migration before 1910. *Explorations in Economic History*, 46(2), 169–185.
- Magnuson, D. L. (2009). History of enumeration procedures, 1790-1940. Accessed through <https://usa.ipums.org/usa/voliii/enumproc1.shtml>.
- Magnuson, D. L., & King, M. L. (2005). Comparability of the public use microdata samples: Enumeration procedures. *Historical Methods*, 28(1), 27–32.
- Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social forces*, 67(2), 281–315.
- Rabinowitz, H. N. (1996). *Race relations in the urban South, 1865-1890*. University of Georgia Press Athens, GA.

- Ransom, R. L., & Sutch, R. (2001). *One kind of freedom: The economic consequences of emancipation*. Cambridge University Press.
- Reardon, S. F., Matthews, S. A., O'Sullivan, D., Lee, B. A., Firebaugh, G., Farrell, C. R., & Bischoff, K. (2008). The geographic scale of metropolitan racial segregation. *Demography*, 45(3), 489–514.
- Ruggles, S. (1994). The origins of african-american family structure. *American Sociological Review*, 59, 136–151.
- Ruggles, S., Sobek, M., Alexander, T., Fitch, C., Goeken, R., Hall, P., King, M., & Ronnander, C. (2009). Integrated public use microdata series sample of the 1880 federal census. Accessed through usa.ipums.org/usa/.
- Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, 1(1), 143–186.
- Sterling, D. (1976). *The Trouble they seen: The story of Reconstruction in the words of African Americans*. Da Capo Press.
- Tolnay, S. E., & Beck, E. (1992). Racial violence and black migration in the american south, 1910 to 1930. *American Sociological Review*, 57(1), 103–116.
- Troesken, W. (2002). The limits of jim crow: Race and the provision of water and sewerage services in american cities, 1880-1925. *The Journal of Economic History*, 62(3), 734–772.
- Wade, R. C. (1967). *Slavery in the cities: The south 1820-1860*. Oxford University Press.
- Williams, D. R., & Collins, C. A. (1999). Segregation and mortality: the deadly effects of racism? *Sociological Forum*, 14(3), 495–523.
- Williams, D. R., & Collins, C. A. (2001). Racial residential segregation: a fundamental cause of racial disparities in health. *Public Health Reports*, 116(5), 404–416.
- Zubrinisky, C. L., & Bobo, L. (1996a). Attitudes toward residential integration: perceived status differences, mere in-group preference, or racial prejudice? *Social Forces*, 74(3), 883–909.
- Zubrinisky, C. L., & Bobo, L. (1996b). Prismatic metropolis: race and residential segregation in the city of angels. *Social Science Research*, 25(3), 335–374.

Table 1: Means of county-level segregation measures by region, 1880.

	Rural counties				Urban counties			
	Alpha	Percent black	Dissimilarity	Isolation	Alpha	Percent black	Dissimilarity	Isolation
New England	0.101 (0.127)	0.007 (0.010)	0.581 (0.219)	0.075 (0.115)	0.191 (0.172)	0.014 (0.014)	0.582 (0.136)	0.061 (0.054)
Middle Atlantic	0.145 (0.145)	0.015 (0.020)	0.541 (0.146)	0.076 (0.105)	0.201 (0.135)	0.017 (0.014)	0.527 (0.118)	0.048 (0.048)
East North Central	0.120 (0.144)	0.012 (0.026)	0.523 (0.257)	0.126 (0.168)	0.143 (0.133)	0.025 (0.047)	0.506 (0.186)	0.061 (0.058)
West North Central	0.139 (0.156)	0.015 (0.034)	0.366 (0.323)	0.113 (0.173)	0.184 (0.164)	0.047 (0.057)	0.479 (0.138)	0.065 (0.043)
South Atlantic	0.275 (0.132)	0.338 (0.231)	0.201 (0.127)	0.056 (0.065)	0.459 (0.114)	0.484 (0.196)	0.28 (0.123)	0.107 (0.077)
East South Central	0.276 (0.129)	0.266 (0.234)	0.289 (0.119)	0.091 (0.093)	0.378 (0.12)	0.306 (0.234)	0.314 (0.079)	0.083 (0.043)
West South Central	0.336 (0.161)	0.233 (0.248)	0.241 (0.197)	0.09 (0.127)	0.391 (0.147)	0.398 (0.26)	0.218 (0.137)	0.073 (0.046)
Mountain	0.079 (0.146)	0.010 (0.024)	0.303 (0.267)	0.134 (0.218)	0.07 (0)	0.005 (0)	0.536 (0)	0.018 (0)
Pacific	0.048 (0.119)	0.005 (0.006)	0.357 (0.258)	0.119 (0.17)	0.066 (0.071)	0.01 (0.007)	0.484 (0.147)	0.041 (0.022)

Notes: Standard deviations are given in parentheses. The rural/urban distinction is based on the 1880 census definition of an urban area being a place with more than 8,000 inhabitants.

Table 2: Means of county-level segregation measures by region with counties weighted by number of black households, 1880.

	Rural counties				Urban counties			
	Alpha	Percent black	Dissimilarity	Isolation	Alpha	Percent black	Dissimilarity	Isolation
New England	0.241 (0.135)	0.015 (0.009)	0.512 (0.115)	0.053 (0.057)	0.373 (0.127)	0.02 (0.008)	0.611 (0.098)	0.12 (0.052)
Middle Atlantic	0.282 (0.158)	0.041 (0.03)	0.496 (0.112)	0.08 (0.069)	0.419 (0.165)	0.027 (0.013)	0.609 (0.131)	0.126 (0.087)
East North Central	0.233 (0.129)	0.051 (0.064)	0.484 (0.147)	0.083 (0.067)	0.299 (0.115)	0.066 (0.08)	0.467 (0.141)	0.076 (0.037)
West North Central	0.259 (0.116)	0.101 (0.071)	0.386 (0.145)	0.078 (0.066)	0.386 (0.127)	0.101 (0.046)	0.405 (0.101)	0.089 (0.035)
South Atlantic	0.311 (0.107)	0.521 (0.179)	0.199 (0.092)	0.057 (0.048)	0.518 (0.116)	0.509 (0.189)	0.323 (0.089)	0.127 (0.051)
East South Central	0.374 (0.093)	0.536 (0.234)	0.286 (0.099)	0.099 (0.063)	0.448 (0.04)	0.468 (0.211)	0.314 (0.066)	0.112 (0.026)
West South Central	0.433 (0.092)	0.527 (0.218)	0.261 (0.118)	0.096 (0.076)	0.412 (0.079)	0.446 (0.189)	0.254 (0.096)	0.09 (0.037)
Mountain	0.274 (0.241)	0.057 (0.074)	0.481 (0.198)	0.192 (0.173)	0.07 (0)	0.005 (0)	0.536 (0)	0.018 (0)
Pacific	0.059 (0.121)	0.011 (0.007)	0.395 (0.138)	0.082 (0.099)	0.064 (0.047)	0.013 (0.005)	0.494 (0.107)	0.029 (0.012)

Notes: Standard deviations are given in parentheses. The rural/urban distinction is based on the 1880 census definition of an urban area being a place with more than 8,000 inhabitants.

Table 3: Correlations in segregation measures for rural and urban counties, 1880.

Rural counties									
<u>Unweighted</u>				<u>Weighted by number of black households</u>					
Percent				Percent					
Alpha	black	Dissimilarity	Isolation	Alpha	black	Dissimilarity	Isolation	Alpha	black
1			Alpha	1			Alpha	1	
Percent black	0.6065	1	Percent black	0.4303	1		Percent black	0.4303	1
Dissimilarity	-0.36	-0.5199	1	Dissimilarity	-0.2055	1	Dissimilarity	0.2853	-0.2055
Isolation	-0.0201	-0.1643	0.5303	Isolation	0.0763	0.7624	Isolation	0.5476	0.0763
Urban counties									
<u>Unweighted</u>				<u>Weighted by number of black households</u>					
Percent				Percent					
Alpha	black	Dissimilarity	Isolation	Alpha	black	Dissimilarity	Isolation	Alpha	black
1			Alpha	1			Alpha	1	
Percent black	0.6819	1	Percent black	0.2769	1		Percent black	0.2769	1
Dissimilarity	-0.4438	-0.5503	1	Dissimilarity	-0.5316	1	Dissimilarity	0.1411	-0.5316
Isolation	0.4999	0.3471	0.18	Isolation	0.0627	0.5046	Isolation	0.6931	0.0627

Table 4: Historical Demographic Predictors of the 1880 Segregation Measure

Segregation and Historical County Level Measures: Historical Predictors of 1880 Segregation

	Neighbor-Based Segregation Measure					
Slave Percentage of Population, 1860	0.323*** (0.0178)		0.423*** (0.0134)	0.293*** (0.0180)		
Free Percentage of Population, 1860		0.857*** (0.137)	0.892*** (0.129)	1.072*** (0.137)		
Free Black Percent of Black Population, 1860		-0.0992*** (0.0199)		-0.0809*** (0.0202)		
Δ Slave Pct. of Population, 1850-1860					0.118 (0.0928)	0.119 (0.0928)
Δ Free Pct. of Population, 1850-1860					0.210 (0.390)	0.277 (0.416)
Δ Free Black Pct. of Black Population, 1850-1860					-0.0405 (0.111)	-0.0405 (0.111)
Constant	0.116*** (0.0178)	0.109*** (0.0174)	0.217*** (0.0268)	0.190*** (0.0265)	0.108*** (0.0174)	0.109*** (0.0175)
Region Fixed Effects	X	X	X	X	X	X
Observations	1,864	1,864	1,755	1,755	1,550	1,489
R-squared	0.394	0.310	0.302	0.412	0.323	0.312

Standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Table 5: Historical Economic Predictors of the 1880 Segregation Measure

Segregation and Historical County Level Measures: Historical Predictors of 1880 Segregation		Neighbor-Based Segregation Measure	
Cotton (Bales Ginned), 1860	3.48e-06*** (5.10e-07)		
Cotton (Bales Ginned), 1870	8.45e-06*** (6.93e-07)		
Δ Cotton (Bales Ginned), 1850-1860	3.11e-06*** (8.15e-07)		
Δ Cotton (Bales Ginned), 1860-1870	-3.78e-06*** (7.95e-07)		
Δ Acres of improved farmland, 1850-1860	-2.44e-07** (9.68e-08)		
Δ Acres of improved farmland, 1860-1870	-3.77e-07*** (8.79e-08)		
Δ Average value of farms per acre, 1850-1860	0.000126*** (2.11e-05)		
Δ Average value of farms per acre, 1860-1870	0.000222** (9.59e-05)		
Constant	0.116*** (0.0178)	0.116*** (0.0177)	0.116*** (0.0174)
Region Fixed Effects	X	X	X
Observations	1,844	1,984	1,841
R-squared	0.325	0.332	0.339
		0.324	0.313
		1,539	1,841
		0.324	0.302
		1,539	1,461
		0.314	0.306

Standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

Table 6: The Correlation of Segregation with Pensions and Mortality, Colored Troops

Segregation, Pensions, and Mortality: Colored Civil War Veterans

Dependent Variable	Age at Death	Age at Death	Age at Death	Age at Death	Pension Amount	Pension Amount	Pension Amount	Pension Amount
War Illness	-0.345 (0.865)	-0.251 (0.761)	-0.349 (0.749)	-0.851 (0.971)	-0.571 (0.991)	-0.286 (0.830)	-0.352 (0.797)	-0.654 (1.093)
War injury	2.000*** (0.731)	1.427** (0.708)	1.419** (0.704)	0.744 (0.853)	0.672 (0.866)	2.431** (1.147)	2.418** (1.131)	1.001 (1.062)
POW	2.486* (1.379)	2.471** (1.193)	1.935 (1.215)	1.564 (3.273)	2.311 (3.282)	0.221 (1.253)	-0.134 (1.214)	-0.207 (2.232)
Percent Black	1.931 (2.096)	1.733 (1.842)	2.660 (1.887)	1.476 (3.061)	2.295 (5.043)	0.576 (2.345)	1.183 (2.300)	-0.246 (5.792)
Pension Amount		0.245*** (0.0339)	0.241*** (0.0319)	0.238*** (0.0495)	0.245*** (0.0512)			
No. Black Physicians			-0.0138** (0.00634)	-0.0181** (0.00869)			-0.00892 (0.00624)	-0.0117* (0.00592)
Segregation			4.930** (2.315)	4.580** (2.256)			6.369*** (2.389)	5.929*** (2.235)
Pct. Black Physicians				-4.746 (5.289)				1.503 (7.552)
Constant	1,001*** (102.3)	1,047*** (92.90)	1,061*** (92.05)	1,033*** (127.5)	1,017*** (129.0)	12.92*** (0.902)	13.20*** (0.824)	11.78*** (0.879)
Observations	6,993	6,987	6,987	4,302	4,302	6,987	6,987	4,302
R-squared	0.177	0.268	0.277	0.252	0.239	0.009	0.011	0.017

Robust standard errors in parentheses clustered at county level *** p<0.01, ** p<0.05, * p<0.1

Table 7: The Correlation of Segregation Mortality, Death Records from Missouri

Segregation and Mortality: Missouri Death Records

Method	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Dependent Variable	Lifespan	Lifespan	Lifespan	Lifespan	Lifespan	Lifespan	Lifespan	Lifespan
Female	1.576*** (0.133)	1.555*** (0.135)	1.551*** (0.133)	1.554*** (0.133)	1.000 (1.025)	0.881 (1.011)	0.881 (1.012)	0.890 (1.016)
Black	-2.556 (1.800)	-2.392 (1.791)	8.009*** (2.473)	7.851*** (2.535)	-5.468*** (0.559)	-5.233*** (0.547)	-1.361 (0.955)	-1.588 (0.971)
Percent Black	2.387 (4.492)	5.083 (5.439)	3.787 (5.482)	13.92 (10.62)	0.919 (7.127)	5.622 (5.580)	5.039 (5.568)	16.54 (11.39)
Segregation		-4.512** (1.723)	-2.350 (2.148)	-3.314 (2.076)		-7.902*** (1.831)	-7.021*** (1.987)	-8.322*** (1.743)
Black*Segregation			-33.09*** (6.362)	-32.59*** (6.509)			-12.59*** (2.974)	-11.82*** (2.905)
Dissimilarity				7.011* (4.050)				9.716** (4.829)
Isolation				-14.27* (8.220)				-24.37** (10.89)
Constant	25.62*** (0.501)	26.56*** (0.678)	26.13*** (0.687)	23.93*** (2.116)	43.49*** (0.755)	45.19*** (0.954)	45.02*** (0.943)	42.51*** (2.369)
Observations	134,029	132,236	132,236	132,236	79,187	78,139	78,139	78,139
R-squared	0.002	0.002	0.003	0.003	0.007	0.009	0.009	0.010

Robust standard errors in parentheses clustered at the county level *** p<0.01, ** p<0.05, * p<0.1

Table 8: Effects of segregation and race on health in rural counties, number of children lost as dependent variable.

	Number of children lost					
Age	0.063 (0.001)***	0.063 (0.001)***	0.063 (0.001)***	0.063 (0.001)***	0.063 (0.001)***	0.063 (0.001)***
Age squared	-0.000 (0.000)***	-0.000 (0.000)***	-0.000 (0.000)***	-0.000 (0.000)***	-0.000 (0.000)***	-0.000 (0.000)***
Income (occscore)	-0.003 (0.001)***	-0.003 (0.001)***	-0.003 (0.001)***	-0.003 (0.001)***	-0.003 (0.001)***	-0.003 (0.001)***
Literate	-0.310 (0.009)***	-0.305 (0.009)***	-0.300 (0.009)***	-0.302 (0.009)***	-0.298 (0.009)***	-0.298 (0.009)***
Black	0.513 (0.009)***	0.494 (0.010)***	0.411 (0.025)***	0.495 (0.010)***	0.444 (0.037)***	0.415 (0.061)***
Segregation index		0.171 (0.036)***	0.226 (0.042)***	0.172 (0.036)***	0.226 (0.042)***	0.226 (0.042)***
Percent black		0.028 (0.023)	-0.072 (0.029)**	0.017 (0.024)	-0.073 (0.029)**	-0.073 (0.029)**
Black x Seg. index			-0.051 (0.071)		-0.058 (0.072)	-0.020 (0.158)
Black x Pct. black			0.242 (0.041)***		0.214 (0.046)***	0.247 (0.086)***
Doctors per 100 households				-0.046 (0.014)***	-0.033 (0.015)**	-0.033 (0.015)**
Black x Doctors					-0.023 (0.028)	0.019 (0.077)
Black x Doctors x Seg.						-0.059 (0.214)
Black x Doctors x Pct. black						-0.052 (0.117)
R^2	0.18	0.18	0.18	0.18	0.18	0.18
N	250,890	233,618	233,618	233,168	233,168	233,168

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Linear regressions with state-year fixed effects. Sample includes all women in Southern states reporting a positive number for children ever born in the 1900 or 1910 census.

Table 9: Effects of segregation and race on health in urban counties, number of children lost as dependent variable.

	Number of children lost					
Age	0.059 (0.002)***	0.059 (0.002)***	0.059 (0.002)***	0.059 (0.002)***	0.059 (0.002)***	0.059 (0.002)***
Age squared	-0.000 (0.000)***	-0.000 (0.000)***	-0.000 (0.000)***	-0.000 (0.000)***	-0.000 (0.000)***	-0.000 (0.000)***
Income (occscore)	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***
Literate	-0.521 (0.020)***	-0.514 (0.022)***	-0.507 (0.022)***	-0.508 (0.022)***	-0.503 (0.022)***	-0.503 (0.022)***
Black	0.457 (0.017)***	0.459 (0.018)***	0.230 (0.057)***	0.462 (0.018)***	0.098 (0.076)	-0.198 (0.146)
Segregation index		0.116 (0.079)	0.083 (0.084)	0.201 (0.081)**	0.209 (0.087)**	0.179 (0.088)**
Percent black		-0.046 (0.058)	-0.195 (0.067)***	-0.142 (0.061)**	-0.295 (0.069)***	-0.290 (0.069)***
Black x Seg. index			0.238 (0.113)**		0.109 (0.117)	0.780 (0.388)**
Black x Pct. black			0.343 (0.077)***		0.433 (0.087)***	0.455 (0.214)**
Doctors per 100 households				-0.156 (0.033)***	-0.187 (0.037)***	-0.188 (0.037)***
Black x Doctors					0.176 (0.053)***	0.529 (0.156)***
Black x Doctors x Seg.						-0.753 (0.403)*
Black x Doctors x Pct. black						-0.071 (0.256)
R^2	0.17	0.17	0.17	0.17	0.17	0.17
N	71,125	64,200	64,200	64,200	64,200	64,200

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Linear regressions with state-year fixed effects. Sample includes all women in Southern states reporting a positive number for children ever born in the 1900 or 1910 census.

Table 10: Effects of segregation and race on health in rural counties, indicator variable for having lost a child as dependent variable.

	Lost a child					
Age	0.090 (0.001)***	0.090 (0.001)***	0.090 (0.001)***	0.090 (0.001)***	0.090 (0.001)***	0.090 (0.001)***
Age squared	-0.001 (0.000)***	-0.001 (0.000)***	-0.001 (0.000)***	-0.001 (0.000)***	-0.001 (0.000)***	-0.001 (0.000)***
Income (occscore)	-0.004 (0.001)***	-0.003 (0.001)***	-0.003 (0.001)***	-0.003 (0.001)***	-0.003 (0.001)***	-0.003 (0.001)***
Literate	-0.122 (0.007)***	-0.120 (0.007)***	-0.117 (0.007)***	-0.117 (0.007)***	-0.115 (0.007)***	-0.115 (0.007)***
Black	0.213 (0.007)***	0.219 (0.008)***	0.201 (0.020)***	0.220 (0.008)***	0.214 (0.029)***	0.278 (0.048)***
1900	-0.084 (0.007)***	-0.083 (0.007)***	-0.083 (0.007)***	-0.083 (0.007)***	-0.082 (0.007)***	-0.082 (0.007)***
Segregation index		0.044 (0.028)	0.102 (0.033)***	0.044 (0.028)	0.102 (0.033)***	0.101 (0.033)***
Percent black		-0.071 (0.018)***	-0.133 (0.022)***	-0.082 (0.018)***	-0.138 (0.023)***	-0.137 (0.023)***
Black x Seg. index			-0.128 (0.056)**		-0.140 (0.056)**	-0.187 (0.124)
Black x Pct. black			0.149 (0.032)***		0.136 (0.036)***	0.030 (0.067)
Doctors per 100 households				-0.043 (0.011)***	-0.039 (0.012)***	-0.039 (0.012)***
Black x Doctors					-0.004 (0.022)	-0.099 (0.060)
Black x Doctors x Seg.						0.074 (0.167)
Black x Doctors x Pct. black						0.170 (0.091)*
N	250,853	233,583	233,583	233,133	233,133	233,133

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Probit regressions with state dummies included. The dependent variable is equal to one if the number of children ever born exceeds the number of children surviving. Sample includes all women in Southern states reporting a positive number for children ever born in the 1900 or 1910 census.

Table 11: Effects of segregation and race on health in urban counties, indicator variable for having lost a child as dependent variable.

	Lost a child					
Age	0.079 (0.002)***	0.078 (0.002)***	0.078 (0.002)***	0.078 (0.002)***	0.078 (0.002)***	0.079 (0.002)***
Age squared	-0.001 (0.000)***	-0.001 (0.000)***	-0.001 (0.000)***	-0.001 (0.000)***	-0.001 (0.000)***	-0.001 (0.000)***
Income (occscore)	-0.004 (0.001)***	-0.003 (0.001)***	-0.003 (0.001)***	-0.003 (0.001)***	-0.003 (0.001)***	-0.003 (0.001)***
Literate	-0.174 (0.015)***	-0.175 (0.016)***	-0.169 (0.016)***	-0.172 (0.016)***	-0.167 (0.016)***	-0.167 (0.016)***
Black	0.225 (0.012)***	0.228 (0.013)***	0.117 (0.042)***	0.230 (0.013)***	0.049 (0.057)	-0.132 (0.108)
1900	-0.005 (0.012)	-0.003 (0.012)	-0.003 (0.012)	-0.003 (0.012)	-0.003 (0.012)	-0.003 (0.012)
Segregation index		-0.006 (0.058)	0.012 (0.062)	0.041 (0.060)	0.079 (0.065)	0.062 (0.066)
Percent black		-0.081 (0.043)*	-0.209 (0.050)***	-0.135 (0.045)***	-0.262 (0.052)***	-0.260 (0.052)***
Black x Seg. index			0.012 (0.084)		-0.056 (0.086)	0.329 (0.287)
Black x Pct. black			0.292 (0.057)***		0.339 (0.065)***	0.377 (0.159)**
Doctors per 100 households				-0.087 (0.024)***	-0.100 (0.028)***	-0.101 (0.028)***
Black x Doctors					0.092 (0.039)**	0.310 (0.116)***
Black x Doctors x Seg.						-0.434 (0.299)
Black x Doctors x Pct. black						-0.076 (0.190)
N	71,108	64,189	64,189	64,189	64,189	64,189

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Probit regressions with state dummies included. The dependent variable is equal to one if the number of children ever born exceeds the number of children surviving. Sample includes all women in Southern states reporting a positive number for children ever born in the 1900 or 1910 census.

Table 12: The Correlation of Segregation with Number of Lynchings per County

Segregation and Lynching

Method	Negative Binomial	Poisson	Probit	OLS	Tobit
Dependent Variable	Number of Lynchings	Number of Lynchings	Lynching in a County	Lynchings (Lynchings>0)	Number of Lynchings
Segregation Index	1.917*** [0.398]	1.464*** [0.208]	0.544*** [0.154]	3.188* [1.698]	5.965*** [1.660]
Percent Black	1.348*** [0.216]	1.252*** [0.105]	0.220** [0.102]	5.264*** [0.906]	5.801*** [0.961]
Isolation Index	-0.0455 [0.820]	0.405 [0.455]	-0.257 [0.243]	0.333 [3.721]	-1.980 [3.123]
Dissimilarity Index	-1.511*** [0.518]	-1.362*** [0.289]	-0.206 [0.176]	-2.396 [2.238]	-3.067 [2.065]
Constant	0.650*** [0.202]	0.765*** [0.108]		2.328*** [0.861]	0.961 [0.874]
State Fixed Effects	X	X	X	X	X
Observations	2,100	2,100	783	597	2,100
R-squared				0.243	

Standard errors in brackets are clustered at the county level *** p<0.01, ** p<0.05, * p<0.1

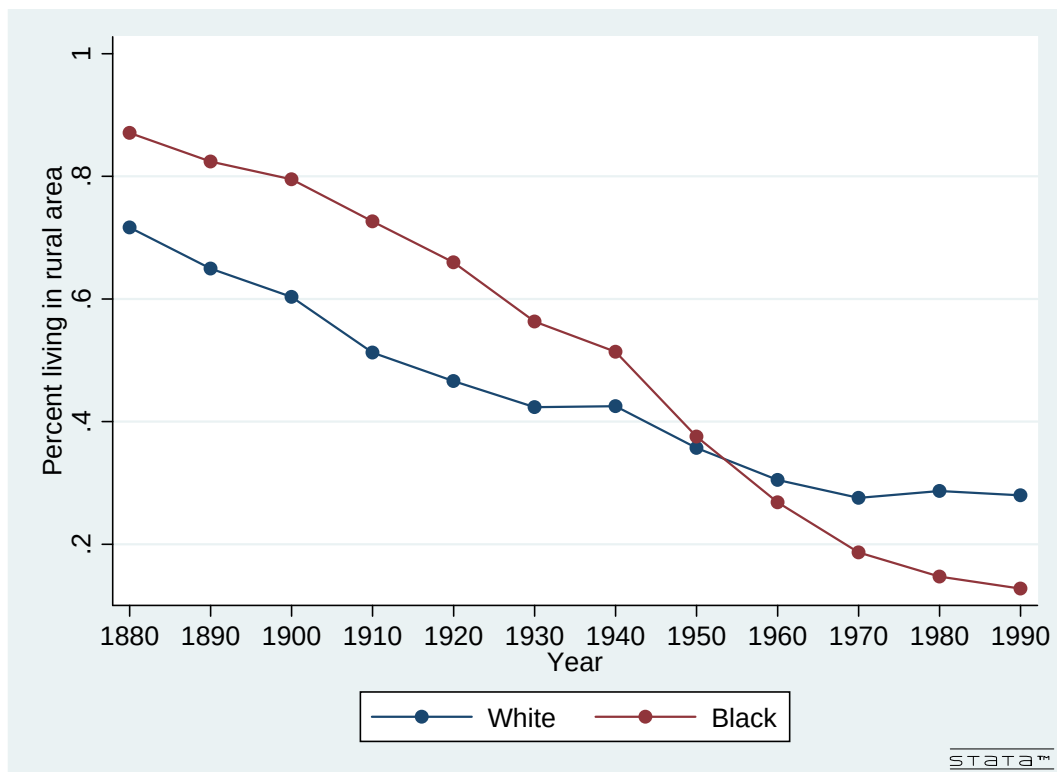


Figure 1: Percentage of population living in rural areas, 1880-1990. Source: Historical Statistics of the United States, Table Aa716-775.

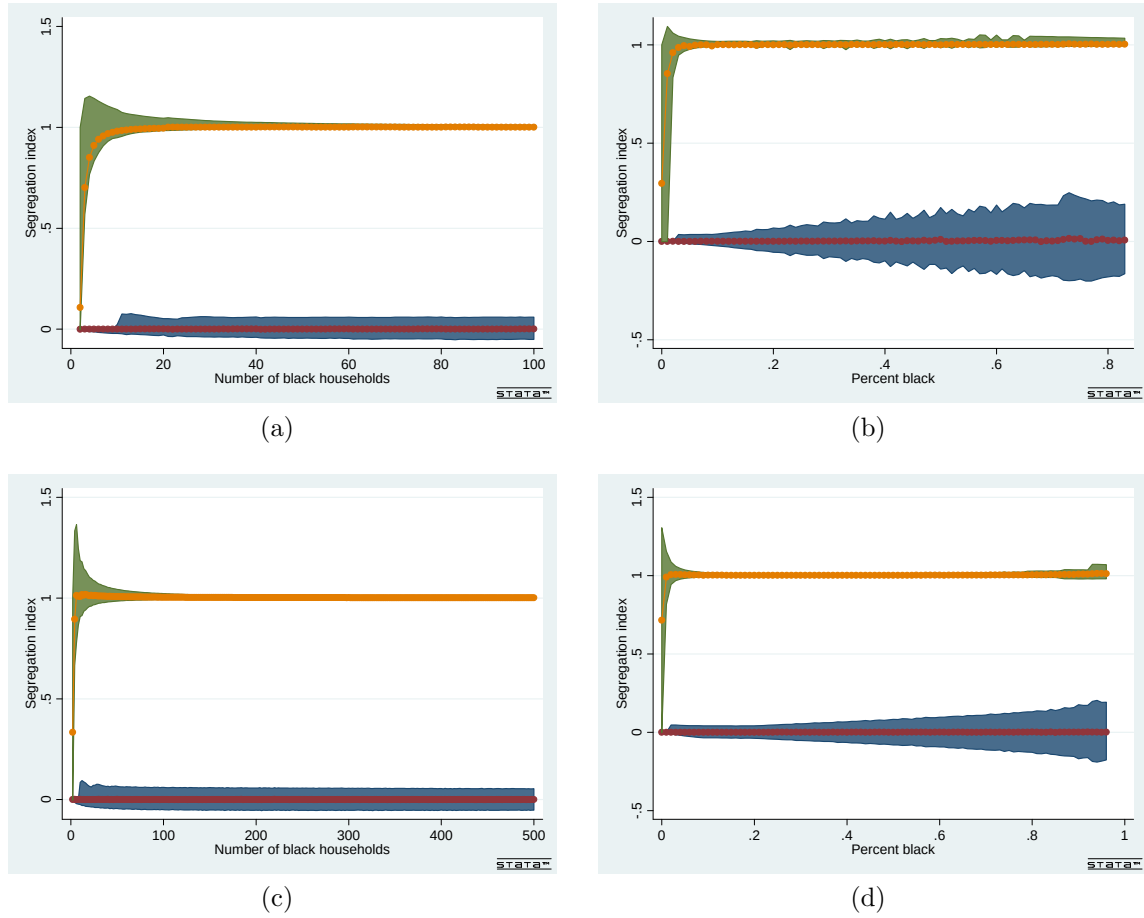
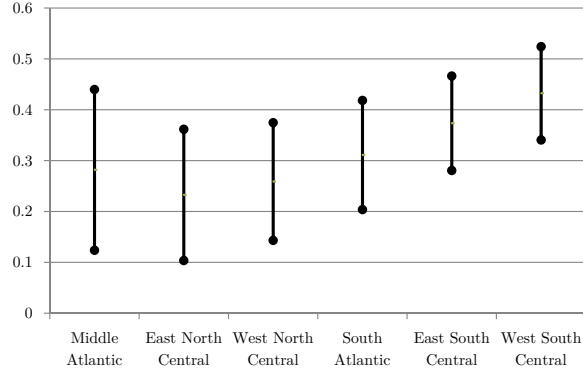
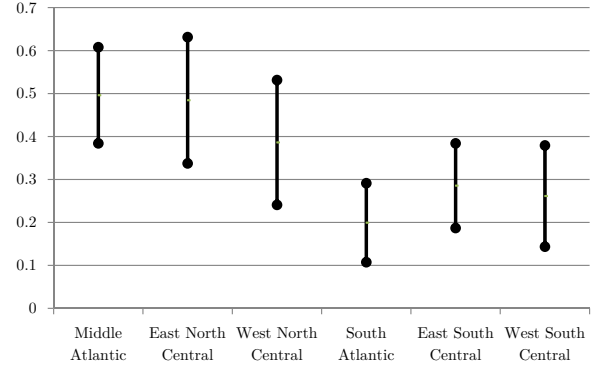


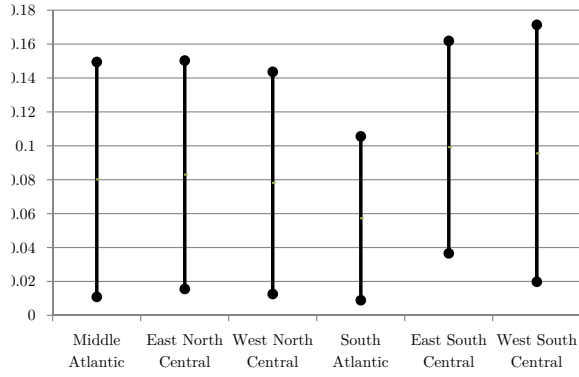
Figure 2: Simulated values for α under no segregation and complete segregation (points give the mean of α while the bands represent the 5th and 95th percentiles): (a) by number of black households with 5% of households missing; (b) by percent black with 5% of households missing; (c) by number of black households with 20% of households missing; and, (d) by percent black with 20% of households missing.



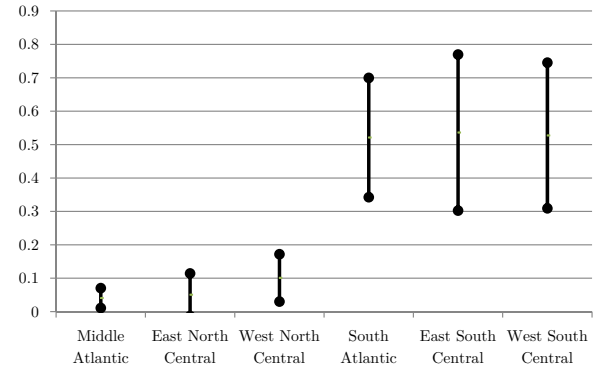
(a)



(b)



(c)



(d)

Figure 3: Black household weighted measures of segregation for rural counties by region. Upper and lower endpoints of bars correspond to one standard deviation above and below the mean, respectively. The measures in each panel are: (a) our α measure, (b) the index of dissimilarity, (c) the index of isolation, and (d) the percentage of household heads who are black.

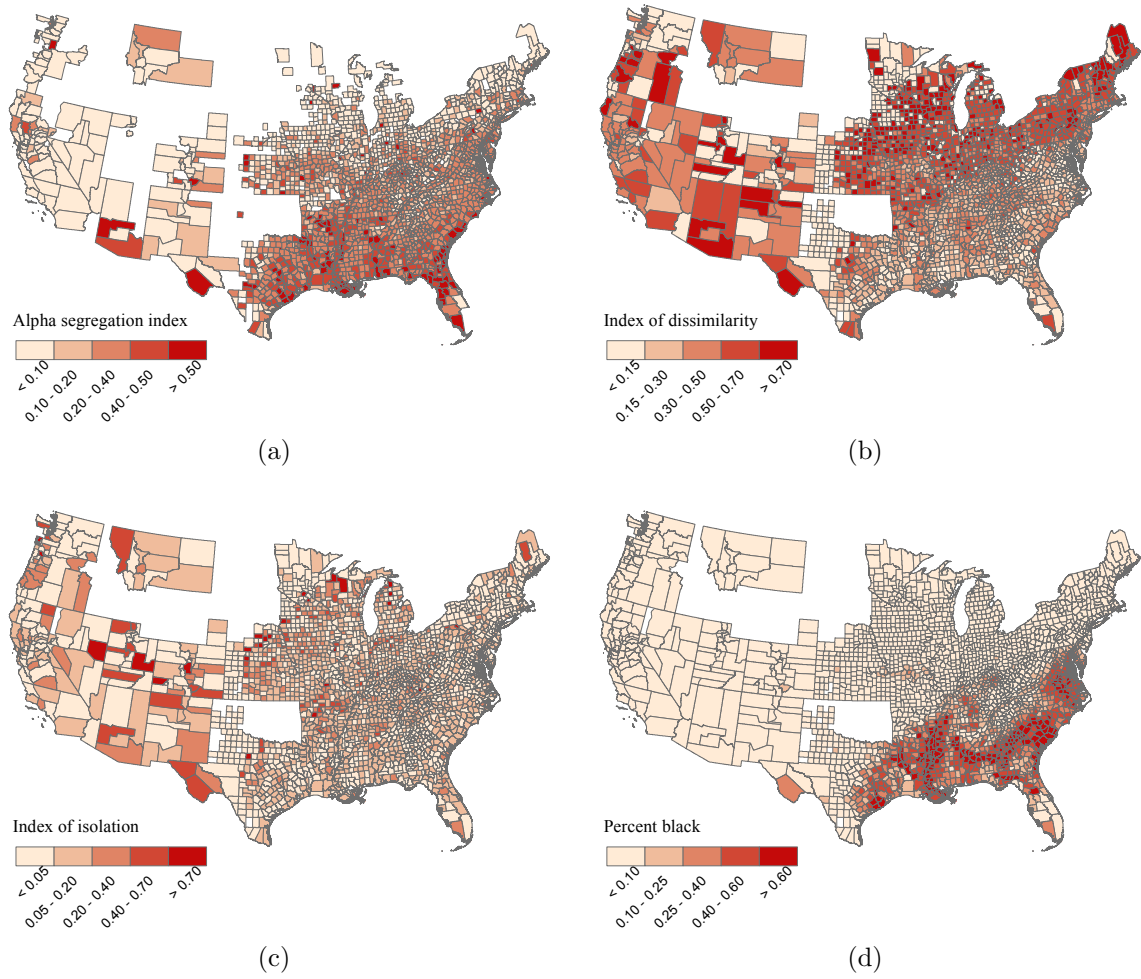


Figure 4: Segregation measures by county, 1880: (a) our α measure; (b) index of dissimilarity; (c) index of isolation; and, (d) percent black.

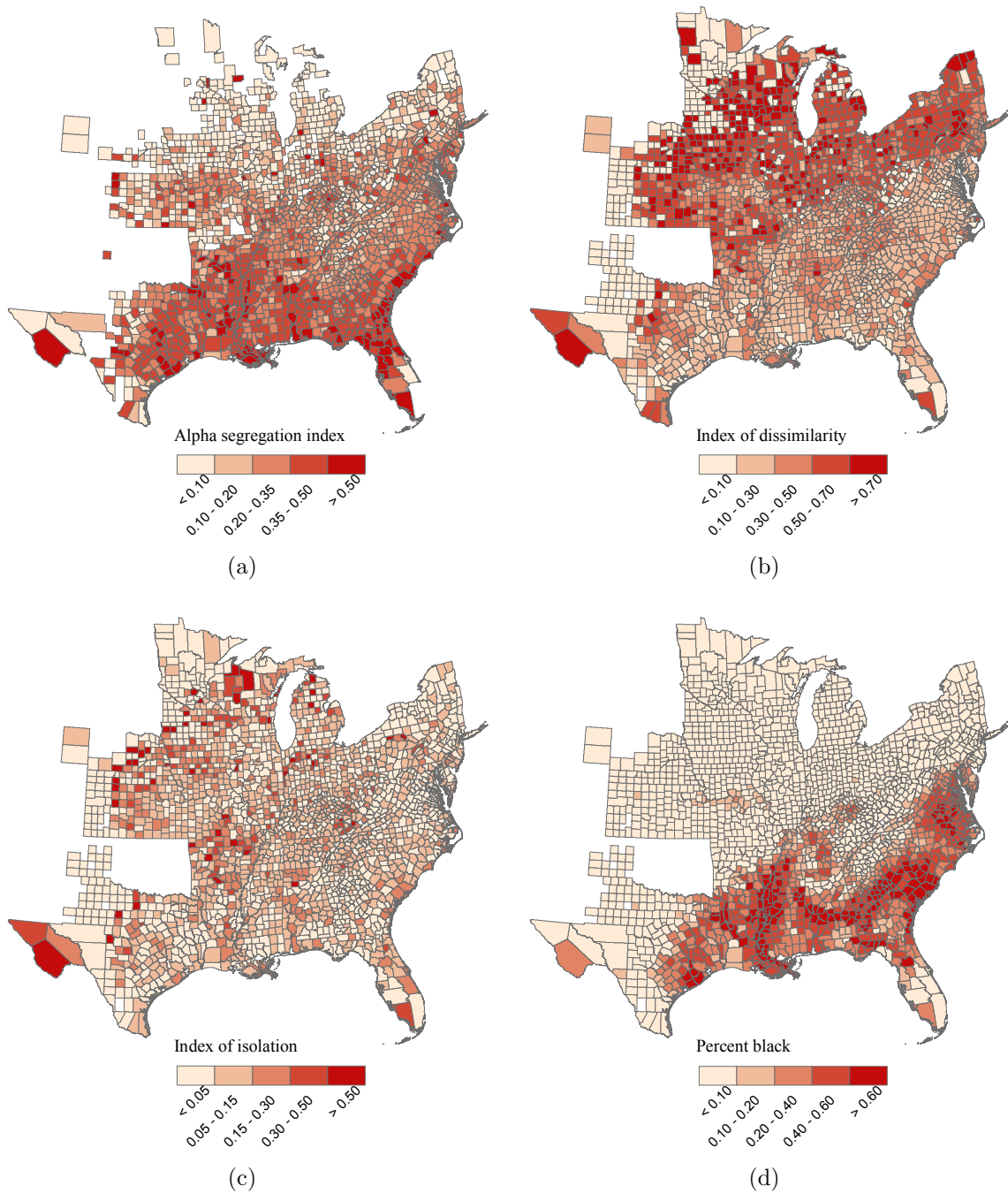


Figure 5: Segregation measures by county for regions with greater than 1% black population, 1880: (a) our α measure; (b) index of dissimilarity; (c) index of isolation; and, (d) percent black.