Poisson Indices of Segregation

Angelo Mele

Johns Hopkins University

Abstract

Existing indices of residential segregation are based on a partition of the city in neighborhoods: given a spatial distribution of racial groups, the index measures different segregation levels for different partitions. I propose a spatial approach, which estimates segregation at the individual level and produces the entire spatial distribution of segregation. This method provides different rankings of cities in terms of segregation and new insights on the effect of segregation on socioeconomic outcomes. Using Census data and controlling for endogeneity using instrumental variables, I show that reduced form estimates of the impact of segregation on socioeconomic outcomes are not robust to the spatial approach.

Keywords: spatial segregation, spatial processes, nonparametric estimation

1. Introduction

This paper studies the relationship between racial residential segregation and socioeconomic outcomes, using a new family of indices derived from spatial statistics. The spatial separation of racial groups in US metropolitan areas is well documented by a large body of research in sociology and economics.[1] Most of the studies find a negative correlation between residential segregation and socioeconomic outcomes of minorities. The empirical strategy in this literature consists of regressing a measure of socioeconomic performance on several controls and an index that proxies for the level of

[1]See for example Massey and Denton (1988), Massey and Denton (1993), Cutler and Glaeser (1997), Cutler et al. (1999),Ananat (2011), Echenique and Fryer (2007), Frankel and Volij (2011), Card and Rothstein (2007), Collins and Margo (2000), Ferrara and Mele (2011),Ananat and Washington (2009).

Preprint submitted to Regional Science and Urban Economics *April 12, 2012*

segregation in the metropolitan area.

However, the majority of existing indices are based on a partition of the city in neighborhoods, that directly affects the estimated segregation.[2] Given a spatial distribution of the racial groups in the city, the index measures different segregation levels if researchers adopt different partitions.

In this paper, I develop a method to measure segregation that considers individuals and their locations as primitives, avoiding the problem of arbitrary partitions. The method consists of estimating a continuous spatial density for the location probability of each racial group. When there is no segregation the spatial density of each group should be flat and equal to the proportion of the group in the metropolitan area. The index measures the deviation of the estimated spatial density from the flat density at each location, providing the entire distribution of segregation among individuals and over space. The segregation level of the city is measured as the average of the estimated individual segregation.

The intuition behind this formulation is simple. Suppose to select a random coordinate in the metropolitan area and draw a circle of 1km radius around the point. Compute the share of blacks living in the circle: this is the probability of black location in that small area. Now let's shrink the radius until the area around the point becomes infinitesimal. The limit of the black share is the probability that the individual at *that location* is African American. Now suppose to repeat this procedure for all the points in the metropolitan area: the result will be a continuous spatial density, that describes the probability of blacks location in the city. If there is no segregation the spatial distribution of blacks does not vary over the metropolitan area, it is flat. Therefore the metropolitan area segregation will be higher the greater the difference between the *actual* spatial distribution of racial groups and the flat spatial density.

This approach has several advantages with respect to the traditional *neighborhood-based* approach. First, the index does not depend on arbitrary partitions of the city. Second, the methodology provides the entire distribution of segregation over space and among individuals. As a result, the researcher is able to identify which regions and individuals account for the

[2]The Spectral Segregation Index of Echenique and Fryer (2007) is an exception. Their index uses individual locations as primitive of the index and therefore does not depend on an arbitrary partition of the city in neighborhoods.

aggregate city-level segregation. Third, the computational burden for the measurement of segregation is minimal, since estimation is performed using simple non-parametric techniques available in standard statistical software.

I provide the theoretical background for the methodology and apply the technique to the measurement of segregation in US metropolitan areas, using Census data. Results show that the spatial index provides a different ranking of the US metropolitan areas in terms of segregation levels, than the one implied by traditional measures. The differences among the spatial and traditional approach are more pronounced in cities with lower income levels, smaller population, lower geographic density and higher fraction of African Americans.

Furthermore, using the index to study the effect of segregation on individual outcomes provides new insights on the relationships between spatial separation of racial groups and socioeconomic performance. The least squares results show that the relationship between segregation and outcomes estimated using the traditional approach is different from when I use the spatial approach. I provide evidence that the spatial index has additional explanatory power and conveys information that is not already contained in the traditional index.

If differences among the approaches are due to pure measurement error in the segregation index, they should disappear when using instrumental variables. I use the inter- and intracounty rivers in a metropolitan area as instruments for segregation, and find that the differences among the approaches persist. Using alternative indices of segregation and alternative samples does not change the result.

An interesting result is that for most cities the individual segregation distribution is not bell-shaped: few highly segregated individuals drive the high average segregation levels. I consider using a more robust proxy of segregation at the city level, i.e. the segregation of the median individual in the metropolitan area. In such IV estimates, the effect of segregation on outcomes is never significant.

The empirical evidence points out that that reduced-form estimates of the effect of segregation on outcomes are not robust to the spatial approach. A micro-founded economic model, in which segregation and outcomes are determined as equilibrium quantities, could help to shed light on the economic reasons behind these results.

The paper is related to several strands of literature. The literature on segregation indices is certainly heavily influenced by the work of Massey and Denton (1988). They review the indices of segregation and group them in five categories: evenness, exposure, concentration, centralization and clustering. They show that the dissimilarity index can explain almost the entire variability of segregation in US cities. Reardon and O'Sullivan (2004) extend the traditional theory of segregation indices to spatial measures. They adapt the properties often required to neighborhood-based indices to a framework based on the location of individuals on a city map. They extend the existing indices in this new framework and check if they satisfy the properties required. Segregation is measured as a function of the agents' *local environment*, where the latter is defined by a proximity function. There are two main differences between their framework and mine: 1) the local environment in this paper is infinitesimal, since I consider a continuous spatial density; 2) I assume that locations are the realization of a stochastic process, while in their paper individual coordinates are assumed as given.

Most of the contributions in economics are based on axiomatic approaches, but consider the neighborhood partitions as given (See Frankel and Volij (2011) and Hutchens (2004) for examples). I do not rely on an axiomatization, but I impose assumptions on the stochastic process that generates locations and marks. In this sense, part of this paper's contribution is to operationalize the estimation of the spatial density using a simple spatial process.

The spatial approach can be considered as a complement to the *spectral* approach of Echenique and Fryer (2007): they develop a segregation index based on individuals' social networks, satisfying three axioms. The Spectral Segregation Index (SSI) measures segregation based on social interactions with same race neighbors and it can disaggregate at the individual level. In this sense the SSI shares most of the advantages of the spatial approach, since it is indipendent of neighborhood partitions. The Spectral Segregation Index is more apt to measure segregation in non-spatial contexts (school segregation, employment segregation) where the spatial approach cannot be implemented. On the other hand, the spatial approach has a comparative advantage in dealing with segregation in geographical contexts, where the SSI uses geographical distance only as an approximation for social interactions. In this sense we can distinguish the two approaches: the spectral approach is *individual-specific* while the spatial approach is *location-specific*.

Another important difference between the Spectral Segregation Index and the Spatial Approach is that the former considers an isolated individual as perfectly integrated, while the latter does not. An isolated individual has no interactions with other individuals of same race and therefore his SSI is zero. In the spatial approach an isolated individual implies that the probability of location in that point is positive: since there are no other individuals located around that point, the probability of location for that particular racial group is very close to one, therefore the individual will be segregated. Furthermore, the spatial approach is easily extended to continuous variables while the Spectral Segregation Index is designed for categorical variables only.

I borrow several concepts and results from the literature on point processes.[3,4] In particular, this paper is related to Diggle et al. (2005), which study the clustering of bovine tuberculosis in Cornwall. They assume that the cases of different types of tuberculosis follow a multivariate inhomogeneous poisson process and compute conditional probabilities of a specific type of disease at a specific location. Their definition of segregation is similar to the one contained in this paper, but the conditional probabilities are computed taking into account the control cases, i.e. bovines which did not developed any form of tuberculosis.[5,6]

The use of spatial techniques in economics is very recent. Arbia et al. (2008) apply techniques from spatial statistics to the analysis of firms' location. Quah and Simpson (2003) empirically test an economic model of location of economic activity using spatial processes that exhibit clustering. While the statistical techniques used in these papers are similar to the ones I propose, they do not rely on synthetic indices to analyze the clustering of the spatial process.

---

[3]See Diggle (2003), Moller and Waagepetersen (2004),Stoyan et al. (1987) and Stoyan and Stoyan (1994) for excellent introductions to the theory and some applications.

[4]Statistical models of point patterns are used in spatial epidemiology (Diggle et al. (2005), Kelsall and Diggle (1998)), Neuroscience (Diggle et al. (2006)), Astrophysics, Ecology, Geology (Zhuang et al. (2006)) and Image Recognition.

[5]In their model there are four types of tubercolosis and there is also a control group, i.e. locations in which there is an animal not infected by the disease. We don't have to model the control group in our application.

[6]They provide a test for *detection* of segregation based on Monte Carlo simulation. However, their test is not particulary useful in the present context. indeed, in a segregation study the researcher is interested in comparing segregation levels among cities, therefore testing if, say, New York is more segregated than Chicago.

2. Motivation and Practical Implementation

Residential separation by race is a common feature observed in the majority of US metropolitan areas. Several studies show that racial segregation undermines the socioeconomic performance of African Americans in education, unemployment, earnings and single motherhood, while the remaining racial groups are not affected significantly (Massey and Denton (1993); Cutler and Glaeser (1997)). Similar results hold when the endogeneity of segregation is accounted for using instrumental variables (Ananat (2011)), when segregation is measured using alternative segregation indices (Echenique and Fryer (2007)), and when performance is measured as the black-white test score gap (Card and Rothstein (2007)).[7,8]

The majority of the literature measures the level of segregation of (say) blacks using a synthetic index. The **traditional approach** consists of the following steps:

1. Partition the city in K neighborhoods
2. For each neighborhood k, compute the share of blacks B_k/P_k, where P_k is the number of individuals and B_k the number of blacks in neighborhood k.
3. Choose a distance function to measure the difference among the actual spatial distribution $(B_1/P_1, ..., B_K/P_K)$ and the distribution arising when there is no segregation $(B/P, ..., B/P)$. The most popular segregation measure, the *dissimilarity index*, is based on the absolute deviation, $|B_k/P_k - B/P|$.
4. Compute the neighborhood-level segregation, using an appropriate normalization that insures the aggregate index assumes values between 0 and 1. For the dissimilarity index each neighborhood has segregation

$$\phi_k = \frac{|B_k/P_k - B/P|}{2\,(B/P)\,(1 - B/P)}$$

5. Compute the average segregation of the city

[7] Collins and Margo (2000) suggest that the negative impact of residential segregation on African Americans outcomes is relatively recent, starting from 1980.

[8] Recently Alesina and Zhuravskaya (2011) constructed measures of segregation at the country level. Their results show that countries with high ethnic and linguistic segregation have a lower quality of government.

$$D = \frac{1}{P}\sum_{k=1}^{K} P_k \phi_k \qquad (1)$$

The index measures the proportion of blacks that should change neighborhood in order to achieve a perfectly integrated city. An alternative interpretation is the mean deviation from uniform spatial distribution, where each neighborhood's segregation is weighted according to the population proportions P_k/P.

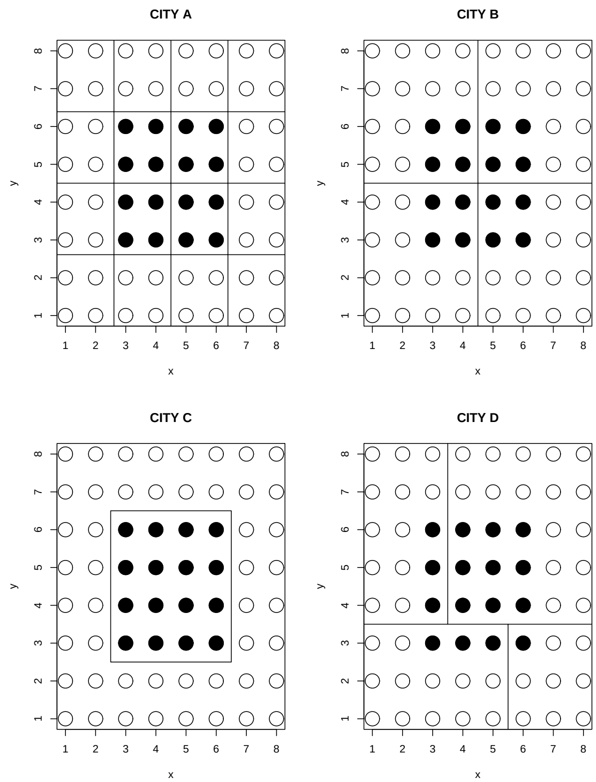
While this approach is suitable for the measurement of segregation in other contexts (employment segregation, school segregation), it has some drawbacks when applied to residential segregation, where the spatial element should be considered explicitly. First, the index is based on an arbitrary partition of the metropolitan area in neighborhoods (as argued by Echenique and Fryer (2007)), making the measurement directly dependent on the specific partition adopted. Figure 1 presents four stylized cities with the same spatial distribution of racial groups, but a different partition in neighborhoods. If segregation is measured using the standard dissimilarity, city A and C are perfectly segregated, city B is perfectly integrated and city D has an intermediate level of segregation. However, the spatial distribution of the racial groups in the four cities is equivalent: the difference in the measured segregation is the result of different partitions.

Second, if we compute the index of segregation using different levels of aggregation of the data (tracts, block groups or blocks) we will observe different values and different ranking of the cities, a problem known in spatial analysis as Modifiable Area Unit Problem (MAUP). In Figure 1, the neighborhoods in city A are obtained by partitioning each of the neighborhoods in city B in four sub-areas of the same size. This results in a dissimilarity of 1 in city A and 0 in city B.

Third, many traditional segregation indices do not take into account the spatial location of individuals over the urban area, thus completely ignoring the within-neighborhood variation in segregation. The dissimilarity index assigns the same segregation level ϕ_k to all individuals living in the same neighborhood. However, the black individual located at (4,5) is surrounded by 8 blacks, while the black individual in (3,3) has 5 white neighbors and 3 black neighbors: an index of segregation should consider the former more segregated than the latter.

If segregation is defined as a function of individual locations, without re-

Figure 1: Different partitions imply different segregation levels



Four stylized cities. Black dots represent the locations of blacks, white dots the locations of whites. The four cities have the same spatial distribution of racial groups. However, when segregation is measured using the neighborhood-based approach, the different partitions in neighborhoods deliver different segregation levels as measured by the dissimilarity index. City A has a dissimilarity $D_A = 1$, while City B has no segregation $D_B = 0$, since each neighborhood contains the same proportion of blacks and whites. Segregation is complete in City C, $D_C = 1$, and intermediate in City D, $D_D = .2291$.

lying on an arbitrary partition in neighborhoods, all these critiques do not apply. This is the main motivation of the present work.

The approach proposed in this paper consists of estimating a spatial density for each racial group, which is compared to the spatial distribution under no segregation. When racial groups do not segregated the probability of location of each racial group is constant over the metropolitan area and the spatial density is flat.

Assume that the researcher has location data x_i for each individual i

in the metropolitan area. The **spatial approach** consists of the following steps:

1. Choose a kernel estimator function $\mathcal{K}_h(u) = \mathcal{K}(u/h)/h^2$, where $\mathcal{K}(\cdot)$ is a given density function. Compute the optimal bandwidth h for the kernel estimator as the minimizer of the Mean Squared Error in formula (17). In this paper, I use a multiplicative quartic kernel, $\mathcal{K}(u) = k(u_1)k(u_2)$, for $u = (u_1, u_2)$.
2. For each individual location x_i, compute the intensity of the spatial pattern for blacks $\lambda_b(x_i)$ and for the entire population $\lambda_0(x_i)$

$$\widehat{\lambda}_b(x_i) = \sum_{j=1}^{n} \frac{\mathcal{K}_h(x_i - x_j)\,\mathbf{1}_{\{m(x_j)=b\}}}{\int_{\mathcal{S}} \mathcal{K}_h(\xi - x_i)\,d\xi} \tag{2}$$

$$\widehat{\lambda}_0(x_i) = \sum_{j=1}^{n} \frac{\mathcal{K}_h(x_i - x_j)}{\int_{\mathcal{S}} \mathcal{K}_h(\xi - x_i)\,d\xi} \tag{3}$$

   where $\mathbf{1}_{m(x_i)=b}$ is an indicator variable which assumes value 1 if the individual at location x_i is black. The estimated intensity $\widehat{\lambda}_b(x_i)$ can be interpreted as the expected number of blacks living in x_i, while $\widehat{\lambda}_0(x_i)$ is the expected number of residents in x_i.
3. Compute the probability of black location at x_i as

$$\widehat{\rho}_b(x_i) = \frac{\widehat{\lambda}_b(x_i)}{\widehat{\lambda}_0(x_i)} \tag{4}$$

   and the probability of black location when there is no segregation, as the proportion of blacks in the population $\rho_b = B/P$
4. Choose a distance function to measure the difference among the actual spatial distribution $\widehat{\rho}_b(\cdot)$ and the spatial distribution under no segregation ρ_b. The distance function associated with the *spatial dissimilarity* is the absolute deviation, $|\widehat{\rho}_b(x_i) - \rho_b|$.
5. Compute the estimated segregation index $\widehat{\phi}(x_i)$ for each observed location x_i. For the spatial dissimilarity we use formula (12)

$$\widehat{\phi}(x_i) = \frac{|\widehat{\rho}_b(x_i) - \rho_b|}{2\rho_b(1-\rho_b)} \tag{5}$$

6. Compute the estimated global index $\widehat{\mathcal{T}}_D(X)$ as the average individual index. The global spatial dissimilarity is thus

$$\widehat{\mathcal{T}}_D(X) = \frac{1}{n}\sum_{i=1}^{n}\widehat{\phi}(x_i)$$

where n is the number of individuals in the city.

The traditional approach imposes a restriction on the individual level segregation, i.e.

$$\phi_i = \phi_k \text{ for all } i \text{ living in neighborhood } k$$

Therefore the traditional dissimilarity assumes no intra-neighborhood variation of spatial segregation. The approach presented here does not impose such a restriction and explicitly considers the spatial distribution of racial groups within neighborhoods.

The computational burden of the spatial approach is minimal, since the kernel estimation procedure is fully automated in standard statistical software for point pattern analysis. An example with instruction for the installation of packages and code for estimation of the segregation levels is available at `https://jshare.johnshopkins.edu/amele1/research.html`.[9]

3. Spatial Point Processes and Segregation

3.1. Notation, Basic Properties and Definitions

A spatial point process X is a stochastic process that maps points over a set $\mathcal{S} \subseteq \mathbb{R}^2$.[10,11] I denote the random set of locations as $X = \{x_1, ..., x_n\}$, where x_i denotes the generic point of the process. The random variable $N(A)$ indicates the number of points in a bounded set $A \subseteq \mathcal{S}$. I denote realizations of X as x and the realizations of N as n. I write ξ or η to indicate a generic

[9] I use the packages `splancs`, `spatstat` and `spatialkernel` for the statistical software `R`. The kernel estimation is accelerated through fast *C* and *Fortran 90* codes that decrease the computational burden even further. A modified version of `spatialkernel` is contained in the web example.

[10] Alternatively it can be defined as a random counting measure over bounded sets $A \subseteq \mathcal{S}$. See Conley (1999) for a more technical explanation of point processes in the context of spatial GMM.

[11] Diggle (2003), Stoyan et al. (1987), Stoyan and Stoyan (1994), Moller and Waagepetersen (2004) are basic references.

point (coordinate) in $\mathcal{S}$ and x_i for the generic realized point of the process. The area of region A is $|A|$ and $d\xi$ refers to the infinitesimal region containing ξ.

I consider only finite spatial processes, with realizations x in the set $N_{1f} = \{x \subseteq \mathcal{S} : n(x \cap A) < \infty\}$, for any bounded $A \subseteq \mathcal{S}$. A spatial point process is *stationary* if all the probability statements about the process in any bounded set $A \subseteq \mathcal{S}$ are invariant under arbitrary translations. This implies that all the statistics are invariant under translation, e.g. $\mathbb{E}N(A) = \mathbb{E}N_p(A)$, where $N_p(A)$ is the process X translated by the vector p. A point process is *isotropic* if the invariance holds under arbitrary rotations. The process is *simple* (or *orderly)* if there are no coincident points. In this paper I consider *simple nonstationary and anisotropic* processes.

Let X be a spatial point process defined over $\mathcal{S} \subseteq \mathbb{R}^2$. The *intensity function* of the process is a locally integrable function[12] $\lambda : \mathcal{S} \to [0, \infty)$, defined as the limit of the expected number of points per infinitesimal area

$$\lambda(\xi) = \lim_{|d\xi| \to 0} \left\{ \frac{\mathbb{E}[N(d\xi)]}{|d\xi|} \right\} \tag{6}$$

A stationary process has constant intensity $\lambda(\xi) = \lambda$ for all ξ. The *intensity measure* of a point process X is defined for $A \subseteq \mathcal{S}$ as

$$\Lambda(A) = \mathbb{E}N(A) = \int_A \lambda(\xi)\, d\xi \tag{7}$$

and measures the expected number of points of the process in the set A.

3.2. Measuring Segregation

Consider a spatial pattern $X = \{x_i, m(x_i)\}_{i=1}^n$ characterized by the locations x_i's in the city $\mathcal{S}$ and marks $m(x_i)$. The mark attached to a location is a random variable describing the characteristics of an individual living at x_i. The mark could indicate the racial group, income group, income level, education level of the individual located at x_i.

I assume that the locations of individuals X_0 are the realization of an Inhomogeneous Spatial Poisson Point Process over the metropolitan area $\mathcal{S} \subseteq \mathbb{R}^2$ with intensity function $\lambda_0(\xi)$

[12] A function is locally integrable if $\int_A \lambda(\xi)\, d\xi < \infty$ for all bounded $A \subseteq \mathcal{S}$

ASSUMPTION 1. *The individual locations X_0 follow an Inhomogeneous Poisson Process with intensity $\lambda_0(\xi)$ over $\mathcal{S}$*

$$X_0 \sim Poi\left(\mathcal{S}, \lambda_0(\xi)\right)$$

This assumption provides a simple but flexible model for the spatial distribution of households in the urban area.[13] The spatial process generates *exogenous* clustering of residential locations, depending on the functional form of the intensity function. Assumption 1 does not impose any behavioral or equilibrium restriction on how people choose their residential locations.

The second assumption concerns the interaction among marks: I assume that conditional on the realized locations, the marks are independent.

ASSUMPTION 2. *Conditional on X_0, the marks are mutually independent*

As a consequence the presence of a specific attribute at a specific location does not influence the presence of same attributes at other locations. It is important to notice that this assumption *does not* rule out spatial clustering of marks.

Let $\rho\left(\xi, m, X_0 \diagdown \xi\right) \equiv \mathbb{P}\left(m(\xi) = m \mid X_0\right)$ be the probability that an individual located at ξ has mark m, conditional on the locations X_0. The third assumption states that the probability distribution of a mark is location-specific and does not depend on the entire realization x of the process. I assume that this conditional probability depends on the location ξ, but it does not depend on the locations of the other points of the process $X_0 \setminus \xi$.

ASSUMPTION 3. *For all $\xi \in X_0$, for all $m \in \mathcal{M}$*

$$\rho\left(\xi, m, X_0 \setminus \xi\right) = \rho(\xi, m)$$

---

[13]The Inhomogeneous Spatial Poisson process is a nonstationary spatial process such that

1. for any bounded region $A \subseteq \mathcal{S}$

$$\mathbb{P}\left[N_0(A) = n\right] = \left[\Lambda_0(A)\right]^n \frac{\exp\left[-\Lambda_0(A)\right]}{n!}, \qquad n = 0, 1, 2, ...$$

2. for any bounded region $A \subseteq \mathcal{S}$, conditional on $N_0(A) = n$ the locations are i.i.d. with density

$$f(\xi) = \frac{\lambda_0(\xi)}{\int_A \lambda_0(\xi)\, d\xi}$$

Assumptions 2 and 3 imply that the probability that an household has a certain characteristic is not affected by the location or attributes of any other household. Marks are independent *conditioning on the realized locations*, but they are not identically distributed at each point. Each location faces a different mark distribution and clustering can occur exogenously according to the functional form of the intensity function and the mark distribution.

The process X defined according to Assumptions 1 to 3, is called a *Marked Spatial Poisson Process.* I will focus on the case of *discrete marks*, which is the appropriate framework for racial segregation. The extension of definitions and theorems to the continuous or multivariate case are straightforward, and shown in Appendix.

The definition of segregated spatial distribution is operationalized using the conditional mark distributions. There is no segregation when the conditional probability of each attribute/mark does not vary over $\mathcal{S}$: $\rho_m(\xi) = \rho_m$ for all ξ. Such a process is said to exhibit *random labeling.*[14] Therefore the marked poisson process is completely unsegregated if there is random labeling of the locations. The maximum level of segregation is reached when the conditional mark distribution is degenerate: for each point of the process there is a mark occurring with probability one at that location, while the remaining marks occur with probability zero.[15]

DEFINITION 1. *Assume that the process X satisfies Assumptions 1-3. Then:*

1. *The marked point process X is completely unsegregated if and only if the conditional mark distribution follows random labeling, i.e. $\rho_m(\xi) = \rho_m$ for all individuals $\xi \in X_0$ and for all racial groups $m \in \mathcal{M}$.*
2. *The marked point process X is completely segregated if and only if for each individual location $\xi \in X_0$, there is a racial group $m^* \in \mathcal{M}$ such that $\rho_{m^*}(\xi) = 1$ and $\rho_m(\xi) = 0$ for any other racial group $m \neq m^*$.*

An index of segregation measures the level of spatial clustering of the marked point process. I focus on indices measuring the difference between the *actual* spatial distribution of racial groups and the distribution arising under no

[14]See Moller and Waagepetersen (2004) for examples.

[15]See Diggle et al. (2005) for a similar definition. The same idea is proposed in Arbia et al. (2008).

segregation. To have comparability across cities the index is normalized to assume values between 0 and 1, where zero corresponds to the case of no segregation and one to the maximum level of segregation. The index increases with the *difference* between the distributions $\rho_m(\xi)$ and ρ_m: different notions of distances between distribution will result in different indices.

Define N_{1m} to be the set of all the possible realizations of the marked point process.

DEFINITION 2. *A segregation index is a function $\mathcal{T} : N_{1m} \to [0, 1]$ such that*

1. $\mathcal{T}(X) = 1$ *iff X is completely segregated*
2. $\mathcal{T}(X) = 0$ *iff X is completely unsegregated (integrated)*
3. $\mathcal{T}(X)$ *is increasing in the difference between the conditional distributions $\rho_m(\xi)$ and ρ_m.*

If the process X satisfies Assumptions 1-3 it is possible to derive the unconditional moments of any index $\mathcal{T}(X)$ (see Appendix). I specialize the framework and impose another restriction often requested in the literature. I focus on indices that satisfy *additivity*: the segregation level of the city is the sum of individual level segregation. Additivity is very common in the literature on segregation, since it allows the researcher to determine which components provide higher contributions to the global level of segregation. Many of the traditional indices are indeed additive at the neighborhood level.

I focus on indices computed *conditionally* on the realization $N(\mathcal{S}) = n$.

I define an *individual* or *location-dependent* segregation function $\phi(\xi)$, summarizing the difference between $\rho_m(\xi)$ and ρ_m at ξ, and a *global* segregation index that aggregates the individual-level indices at the city level. The global index is computed as *average* of the normalized individual-level segregation indices.

ASSUMPTION 4. *Assume the global index $T(X)$ is the average of the individual indices $\phi(\xi)$, conditional on $N(\mathcal{S}) = n$*

$$\mathcal{T}(X) = \frac{1}{n} \sum_{\xi \in X_0} \phi(\xi) \tag{A4}$$

where $\phi : \mathcal{S} \to \mathbb{R}_+$ is a location-specific segregation index.

The function ϕ maps the location into the segregation level of the individual. I provide examples of possible functional forms for ϕ below. The general distributional results are summarized in Appendix. Most of the existing indices can be adapted to this approach by redefining the neighborhoods as individuals.

The construction of the index requires the choice of a distance function for measuring the difference between spatial distributions. Two popular choices are the absolute deviation

$$d(\xi) = \sum_{m \in \mathcal{M}} |\rho_m(\xi) - \rho_m| \tag{8}$$

and the squared deviation

$$d(\xi) = \sum_{m \in \mathcal{M}} [\rho_m(\xi) - \rho_m]^2 \tag{9}$$

The following proposition provides the value of (E.1) and (E.4) when there is complete segregation. Let ξ^s be a generic point of a completely segregated spatial point process.

PROPOSITION 1. *The value of (8) under complete segregation is*

$$d(\xi^s) = 2 \sum_{m \in \mathcal{M}} \rho_m (1 - \rho_m) \tag{10}$$

The value of (9) under complete segregation is

$$d(\xi^s) = \sum_{m \in \mathcal{M}} \rho_m (1 - \rho_m) \tag{11}$$

Proof. In Appendix C. ■

Incidentally notice that $d(\xi^s)$ is a linear function of the fractionalization of the city, as defined below in (E.10).

Using the absolute deviation we can The location-based dissimilarity index is

$$\phi_D(\xi) = \frac{\sum_{m \in \mathcal{M}} |\rho_m(\xi) - \rho_m|}{2 \sum_{m \in \mathcal{M}} \rho_m (1 - \rho_m)} \tag{12}$$

and the global **Spatial Dissimilarity Index** is

$$\mathcal{T}_D(X) = \frac{1}{n} \sum_{\xi \in X_0} \phi_D(\xi) \tag{13}$$

In the traditional neighborhood-based dissimilarity the conditional probability $\rho_m(\xi)$ is assumed to be the same for all locations in the same neighborhood, while the spatial dissimilarity does not impose such within-neighborhood restriction on the spatial segregation.

Using the same approach we can derive several indices of segregation, summarized in Table 1 and described in detail in Appendix.

3.3. Discussion

In principle, one could estimate the spatial density without the assumptions on the spatial process. The probability of location of an individual can be estimated using a simple kernel estimator. However, when the exact location data of the individuals are unavailable, the Poisson assumption allows estimation using simple kernel regression methods. Without this assumption, the researcher should rely on more computationally intensive simulation methods.

The class of Marked Markov Pairwise Interaction models is a valid alternative to the spatial poisson process.[16] These models assume that the probability of location of an individual with a specific mark, depends on the location of other individuals with the same mark. The degree of spatial correlation among locations is driven by the interaction parameter, which measures the degree of attraction among individuals of the same race. These statistical models incorporate the idea of endogenous clustering, but are much harder to estimate. The literature on spatial statistics has developed approximate methods for estimation, i.e. maximum pseudolikelihood, monte carlo maximum likelihood and markov chain monte carlo simulations. However, these techniques are very expensive from a computational point of view. The assumption of Poisson process is convenient, since it allows clustering of individuals without significant increase in the computational burden for estimation.

Finally, I use the average individual segregation as proxy for the segregation of the metropolitan area. However, the average is not robust to outliers, i.e. individuals that are highly segregated. In the empirical application, I provide evidence that high level of segregation are the results of such out-

[16] See Moller and Waagepetersen (2004) or Diggle (2003) for a simple description of these models.

Table 1: Alternative indices of diversity and segregation

Index	$d(\xi)$	$\phi(\xi)$	$T(X)$
		A. Indices of Spatial Diversity	
Spat. Fractionalization		$I(\xi) = \sum_{m \in \mathcal{M}} \rho_m(\xi)(1 - \rho_m(\xi))$	$\mathcal{T}_I(X) = \frac{1}{n} \sum_{\xi \in X_0} I(\xi)$
Spat. Entropy		$E(\xi) = \sum_{m \in \mathcal{M}} \rho_m(\xi) \ln\left(\frac{1}{\rho_m(\xi)}\right)$	$T_E(X) = \frac{1}{n} \sum_{\xi \in X_0} E(\xi)$
		B. Indices of Spatial Segregation	
Spat. Relative Fract.	$d(\xi) = \lvert I(\xi) - I \rvert$	$\phi_F(\xi) = \frac{\lvert I(\xi) - I \rvert}{I}$	$T_F(X) = \frac{1}{N(\mathcal{S})} \sum_{\xi \in X_0} \phi_F(\xi)$
Spat. Relative Entropy	$d(\xi) = \lvert E(\xi) - E \rvert$	$\phi_H(\xi) = \frac{\lvert E(\xi) - E \rvert}{E}$	$T_H(X) = \frac{1}{N(\mathcal{S})} \sum_{\xi \in X_0} \phi_H(\xi)$
Spat. Exposure	$d(\xi) = \sum_{m \in \mathcal{M}} [\rho_m(\xi) - \rho_m]^2$	$\phi_{Exp}(\xi) = \frac{\sum_{m \in \mathcal{M}} [\rho_m(\xi) - \rho_m]^2}{\sum_{m \in \mathcal{M}} \rho_m(1 - \rho_m)}$	$T_{Exp}(X) = \frac{1}{N(\mathcal{S})} \sum_{\xi \in X_0} \phi_{Exp}(\xi)$
Spat. Relative Exposure	$d(\xi) = [\rho_m(\xi) - \rho_m]^2$	$\phi_{V,m}(\xi) = \frac{[\rho_m(\xi) - \rho_m]^2}{\rho_m(1 - \rho_m)}$	$T_P(X) = \frac{1}{N(\mathcal{S})} \sum_{\xi \in X_0} \sum_{m \in \mathcal{M}} \frac{[\rho_m(\xi) - \rho_m]^2}{(1 - \rho_m)}$

The table collects the formulas for alternative spatial diversity and spatial segregation indices. Details for derivation and description of the measures are contained in Appendix.

liers pulling the segregation towards 1. I also consider using a more robust alternative to the average segregation, such as the median segregation.

4. Estimation Strategy

4.1. Estimation with exact location data

I estimate segregation levels using nonparametric methods borrowed from spatial statistics. If the researcher has access to exact individual location data, the estimation of the intensity function is straightforward.[17] Lemma 2 in Appendix B proves that the a spatial process satisfying Assumptions 1-3 is equivalent to a multivariate Poisson process with independent univariate processes. As a consequence, I can estimate the intensity functions of each racial group independently. This observation leads to a convenient estimate of $\widehat{\rho}_m(\xi)$

$$\widehat{\rho}_m(\xi) = \frac{\widehat{\lambda}_m(\xi)}{\widehat{\lambda}_0(\xi)} \tag{14}$$

where $\widehat{\lambda}_m(\xi)$ is the estimate of the intensity function for the process X_m, corresponding to the spatial process for group m. Diggle (1985) and Berman and Diggle (1989) suggested a nonparametric estimator based on the definition of intensity function, $\widetilde{\lambda}(\xi) = N(\xi, h)/\pi h^2$, where $N(\xi, h)$ is the number of points within distance h from ξ. The estimator counts the points within the disc of radius h and centered in ξ, dividing by the area of the disc πh^2.[18] More generally one can weight the points using a Kernel function, which leads to estimators of the form (see Diggle (2003) p.148 or Moller and Waagepetersen

[17] See Diggle (2003), Diggle et al. (2005).

[18] This can be interpreted as a kernel estimator in which the kernel function is

$$k(u) = \begin{cases} \frac{1}{\pi u^2} & \text{if } 0 \le u \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(2004))[19]

$$\widehat{\lambda}(\xi) = \sum_{i=1}^{n} \frac{\mathcal{K}_h(\xi - x_i)}{\int_{\mathcal{S}} \mathcal{K}_h(\xi - x_i)\, d\xi} \tag{15}$$

where $\mathcal{K}_h(u) = \frac{1}{h^2}\mathcal{K}(u/h)$. In the application, I will use a multiplicative quartic kernel. I also use a gaussian kernel, with virtually no difference in estimated probabilities, but substantial increase in the computational time.

The spatial approach allows direct estimation of the optimal neighborhood size, since the estimation procedure requires the computation of an optimal kernel bandwidth h. In principle h assumes different values in each city, because it takes into account the geographic density. The bandwidth h can be interpreted as defining the *relevant neighborhood* for the individual (the local environment, in the words of Reardon and O'Sullivan (2004)), which is possibly different for each metropolitan area.

I choose h using the Mean Square Error (MSE) minimization procedure suggested in Diggle (1985) and Berman and Diggle (1989). The assumption is that the underlying process is a stationary isotropic Cox point process with rate process $\Lambda(\xi)$.[20] The optimal bandwidth h minimizes

$$MSE(h) = \mathbb{E}_{\Lambda,N}\left\{\left[\widetilde{\lambda}(\xi) - \Lambda(\xi)\right]^2\right\} \tag{16}$$

where the expectation is taken with respect to the process governing Λ and the point process conditioning on Λ's realization. It can be shown that $MSE(h)$ can be rewritten as[21]

$$MSE(h) = \mu\frac{1 - 2\mu K(h)}{\pi h^2} + \left(\pi h^2\right)^{-2} \int\int \mu_2(\|\xi - \eta\|)\, d\eta d\xi \tag{17}$$

[19]There are alternative ways to estimate the conditional mark probability. For example, Diggle et al. (2005) exploit the fact that conditioning on the realized n, the mark distribution is a multinomial distribution and can be estimated through kernel regression. Alternative smoothing techniques can be used. For example, the method of total variation regularization proposed in Koenker and Mizera (2004).

[20]A Cox Process is a point process such that: 1) The rate process $\left\{\Lambda(\xi) : \xi \in \mathbb{R}^2\right\}$ is a non-negative-valued stochastic process; 2) Conditional on the realization $\left\{\Lambda(\xi) = \lambda(\xi) : \xi \in \mathbb{R}^2\right\}$, the point process follows an Inhomogeneous Poisson Point process with intensity $\lambda(\xi)$.

[21]This is a simple method of computing the optimal bandwidth. An alternative method is described in Diggle et al. (2005).

with $\mu = \mathbb{E}[\Lambda(\xi)]$; the term $\mu_2(\|\xi - \eta\|)$ is the *second-order intensity*, defined as $\mu_2(u) = \gamma(u) - \mu^2$, where $\gamma(u) = \mathbb{E}[\Lambda(\xi)\Lambda(\eta)]$ is the covariance of the rate process $\Lambda(\xi)$. The quantity $K(h)$ is

$$K(h) = \lambda^{-1} E[N_o(h)] = 2\pi\lambda^{-2} \int_0^h \mu(\xi)\,\xi d\xi \tag{18}$$

which measures the expected number of *further* points in the circle of radius h and center ξ. I estimate $K(h)$ with the Ripley's estimator: define $w(\xi, u)$ as the proportion of the circumference of the circle with center ξ and radius u, which lies in $\mathcal{S}$, and $w_{ij} = w(x_i, u_{ij})$, where $u_{ij} = \|x_i - x_j\|$.

$$\widehat{K}(h) = \frac{1}{n(n-1)} |\mathcal{S}| \sum_{i=1}^{n} \sum_{j \neq i} w_{ij}^{-1} I_h(u_{ij}) \tag{19}$$

where $I_h(u_{ij}) = I(u_{ij} \leq h)$ is an indicator function. This gives edge-corrected estimates of the $K(h)$ function. The integral in (17) is estimated using the weighted integral suggested by Berman and Diggle (1989). These estimates are plugged in (17) to obtain an estimated $\widehat{MSE}(h)$. The optimal h is the minimizer of $\widehat{MSE}(h)$.

As a practical matter, when estimating the conditional probability, I use the same bandwidth for $\widehat{\lambda}_m(\xi)$ and $\widehat{\lambda}_0(\xi)$, to avoid probabilities greater than one or conditional probabilities not summing up to one. In Appendix D, I apply the estimation procedure to artificial data.

4.2. Estimation with block level data

In many empirical applications, the exact location data are not available. Therefore, I develop an approximated estimation technique to deal with data at the block level. I assume the data contain the number of individuals of each racial group in each block, and the location of the block centroid. This is the case in my empirical application.

The metropolitan area $\mathcal{S}$ is partitioned in K disjoint blocks, $\mathcal{S} = \bigcup_{k=1}^{K} \mathcal{S}_k$ and $\mathcal{S}_k \cap \mathcal{S}_l = \varnothing$, for $k \neq l$. By the *independent scattering property* of the inhomogeneous poisson process the number of points of the process $N_0(\mathcal{S}_k)$ and $N_0(\mathcal{S}_l)$ over disjoint regions $\mathcal{S}_k$ and $\mathcal{S}_l$ are independent (see Appendix B.1 for a proof). The definition of intensity measure implies that $\mathbb{E}N_0(\mathcal{S}_k) =$

$\int_{\mathcal{S}_k} \lambda_0(\xi)\, d\xi$, for any k. One can model the number of points as

$$N_0(\mathcal{S}_k) = \int_{\mathcal{S}_k} \lambda_0(\xi)\, d\xi + u_k$$

where u_k is an error with mean zero, and independent across blocks. For any block k there exists a $\bar{\xi}_k \in \mathcal{S}_k$ such that $\int_{\mathcal{S}_k} \lambda_0(\xi)\, d\xi = \lambda_0(\bar{\xi}_k)\, |\mathcal{S}_k|$ and thus

$$N_0(\mathcal{S}_k) = \lambda_0(\bar{\xi}_k)\, |\mathcal{S}_k| + u_k \tag{20}$$

Notice that $\bar{\xi}_k$ is not necessarily the centroid of the block. An approximation of (20) for any $\xi \in \mathcal{S}_k$ is $N_0(\mathcal{S}_k) \approx \lambda_0(\xi)\, |\mathcal{S}_k| + u_k$.

The expected number of points in $\mathcal{S}_k$ is then approximated as

$$\mathbb{E}\left[N_0(\mathcal{S}_k) \middle| \xi \right] \approx \lambda_0(\xi)\, |\mathcal{S}_k|$$

and thus the function $\lambda_0(\xi)\, |\mathcal{S}_k|$ can be estimated through kernel regression as

$$\widehat{\lambda}_0(\xi)\, |S_k| = \sum_{k=1}^{K} \frac{\mathcal{K}_h(\xi - x_k)}{\sum_{j=1}^{K} \mathcal{K}_h(\xi - x_j)} n_{0k} \tag{21}$$

where x_k's are the centroids of the census blocks and n_{0k} the number of individuals observed in each block. Applying this procedure to each racial group process we can then estimate $\widehat{\lambda}_m(\xi)\, |\mathcal{S}_k|$ for each m.

Taking the ratio $\frac{\widehat{\lambda}_m(\xi)|\mathcal{S}_k|}{\widehat{\lambda}_0(\xi)|\mathcal{S}_k|}$ we get the estimator for $\widehat{\rho}_m(\xi)$

$$\widehat{\rho}_m(\xi) = \frac{\widehat{\lambda}_m(\xi)}{\widehat{\lambda}_0(\xi)} = \frac{\sum_{k=1}^{K} \mathcal{K}_h(\xi - x_k)\, n_{mk}}{\sum_{k=1}^{K} \mathcal{K}_h(\xi - x_k)\, n_{0k}} \tag{22}$$

where n_{0k} is the number of people living in block k and n_{mk} is the number of people belonging to race m living in block k. To estimate the index of segregation, I evaluate the estimated conditional probabilities at the block centroid.

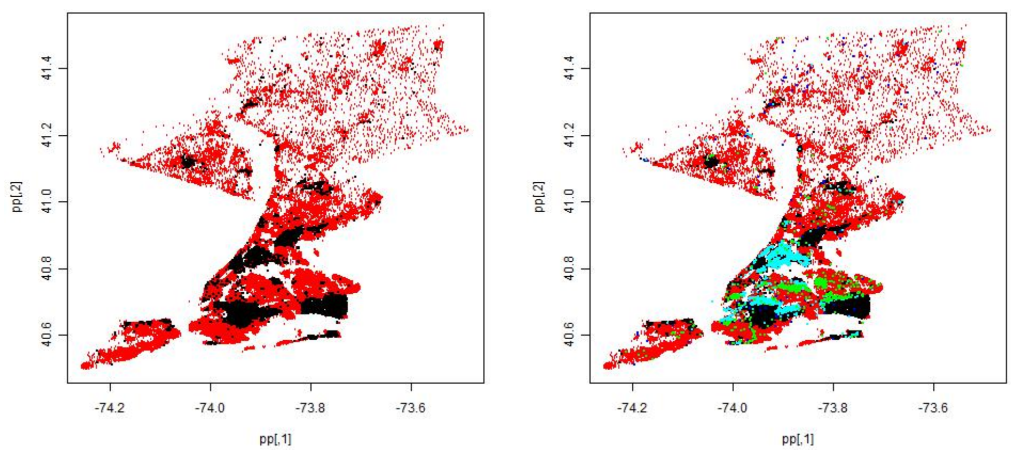
4.3. Data

I apply this approach to census data from the 1990 and 2000 US Census of Population and Housing. The ideal dataset would consist of individual or household level data on location, racial group and socioeconomic characteristics. Unfortunately such data are not publicly available for confidentiality reasons.[22] A possible alternative is the 1% PUMS 1930 Census, where each household's address is reported. However, there are concerns about the spatial randomness of this sample and the geocoding of historical addresses, therefore I prefer to not use these data.

As a necessary compromise between estimation precision and reliability of data, I use the most disaggregated data publicly available: census block data containing the location of the block centroid and the racial composition. In Appendix D I illustrate the methodology using exact locations from artificial datasets.

I have data for all the 331 MSA's (Metropolitan Statistical Areas) and

Figure 2: Segregation in New York City, PMSA



Each dot represents a block centroid. On the left figure, a black dot is a block where the majority of residents is African American, while a red dot indicates a block with a majority of non-blacks. On the right picture, red indicates Whites/Caucasians, black indicates African Americans, green indicates Asians, and light blue indicates Other racial groups (including Hispanics).

PMSA's (Primary Metropolitan Statistical Areas) for years 1990 and 2000. To maintain comparability across census years, I adopt the racial categories in Census 1990: Whites/Caucasians, African Americans, Asian/Pacific Islanders, Native American, Other.

[22] I am in contact with the Census Bureau to gain access to such data.

The left panel of Figure 2 plots all the blocks centroids of New York in 2000: the black dots represent blocks in which the majority is black while red dots are blocks with a majority of non-blacks. The pattern of geographic separation is evident: African Americans are concentrated in Harlem, Bronx and Bedford-Stuyvesant. The right panel of Figure 2 plots all racial groups: black points are African Americans, red points are Whites, green are Asians and light blue correspond to Other racial groups (including Hispanics). [23]

5. Results

5.1. Descriptive results

The estimates for the spatial distribution of segregation in New York are shown in Figure 3 and 4. In Figure 3, the left panel shows the spatial dissimilarity for African Americans, while the right panel displays the spatial exposure. The areas with higher segregation correspond to Bedford-Stuyvesant, Harlem and Bronx.

One of the most striking features of Figure 3 is that most areas show moderate levels of racial segregation. This is the case for most metropolitan areas in the sample: the individual segregation distribution is very skewed and very few areas present extreme levels of segregation. The average new yorker has spatial dissimilarity of 0.6903519, with a median of 0.6423153. The spatial exposure index in the right panel shows a similar pattern. The results for multigroup segregation in Figure 4 are virtually identical, with few areas showing unusual segregation levels.

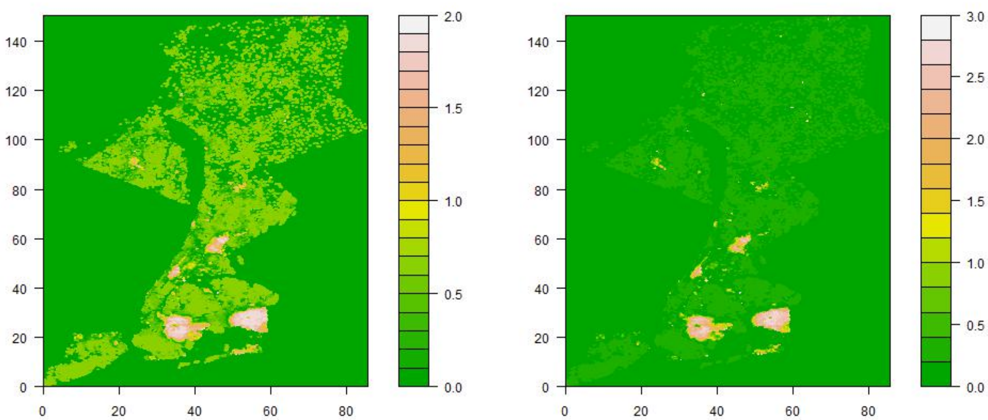
The metropolitan area segregation levels estimated using the spatial approach are different from those estimated with the traditional approach. In Table 2, I present the estimated segregation levels for several metropolitan areas.[24] I compare the spatial dissimilarity with the traditional dissimilarity, the latter estimated using both census tracts and blocks. Panel A and B show the ten most and least segregated MSAs respectively. Panel C shows the results for the most populated MSAs.

The spatial approach provides a different picture of segregation levels in US cities than the traditional approach, both in levels and rankings. For example, Muncie (IN) and Beaumont (TX) have drastically different levels of segregation, when using the traditional approach vis-a-vis the spatial

[23]Other metropolitan areas are available from the author.

[24]The results for all metropolitan areas are available in excel format from the author.

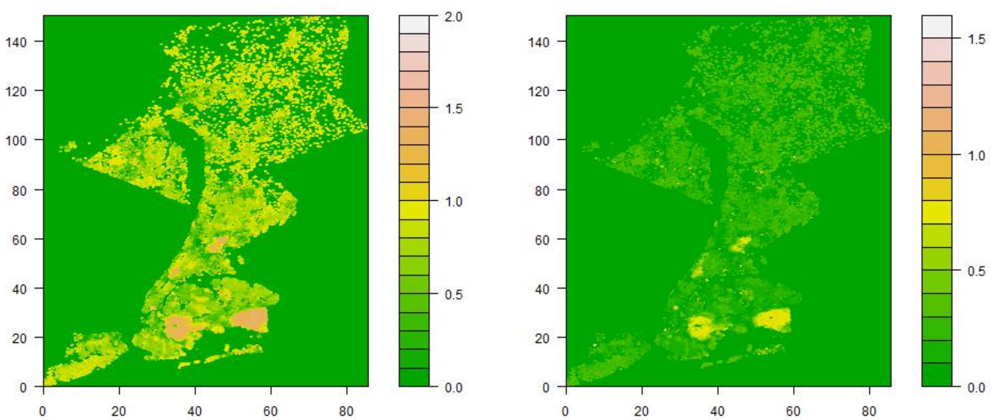
Figure 3: Estimated African American segregation in New York PMSA, 2000



Estimated spatial dissimilarity (left panel) and spatial exposure (right panel) for African Americans in New York.

approach. Figure 5 plots the spatial dissimilarity and the neighborhood-based dissimilarity (computed using census tracts). Each point represents a metropolitan area, indicated with the MSA FIPS code. Spatial dissimilarity is positively associated with the traditional dissimilarity, as expected. However the measured levels of segregation in many metropolitan areas are strikingly different when we compare the two methodologies. For example, the metropolitan area of Muncie (IN), with MSA FIPS code 5280 in the figure, has a dissimilarity of 0.7022 while the spatial dissimilarity is 0.8785. Furthermore, the spatial dissimilarity implies a different ranking of cities in terms of racial segregation: Muncie (IN) is indeed the most segregated metropolitan area according to the spatial approach, while using the traditional approach it was 141st. Table 4 shows evidence that while aggregate (average) segregation could be high in some metropolitan areas, most of the individuals are exposed to moderate levels of spatial separation. For example, the high levels of segregation in Los Angeles depend on very few areas with quite unusually high segregation: 75% of the population is exposed to

Figure 4: Estimated multigroup segregation in New York PMSA, 2000



Estimated spatial dissimilarity (left panel) and spatial exposure (right panel) for all racial groups in New York.

segregation levels below 0.55, while the average of the index is 0.61.[25]

5.2. Segregation and Outcomes

In this section, I explore the economic implications of the spatial approach. I study the effect of racial segregation on individual outcomes, comparing results using the traditional index and the spatial index. I focus on three outcomes, also studied by Cutler and Glaeser (1997): high school graduation, college graduation and idleness.

Table 6 shows some preliminary evidence of the relationship between segregation and outcomes. I present estimates of linear probability models as in Cutler and Glaeser (1997); probit and logit estimates provide the same qualitative results. The estimates focus on the sample of 25-30 years old individuals from the 1% PUMS 1990.[26]

[25]A previous version of the paper contains more details about the individual distribution of segregation.

[26]The sample selection follows the same procedure as in Cutler and Glaeser (1997)

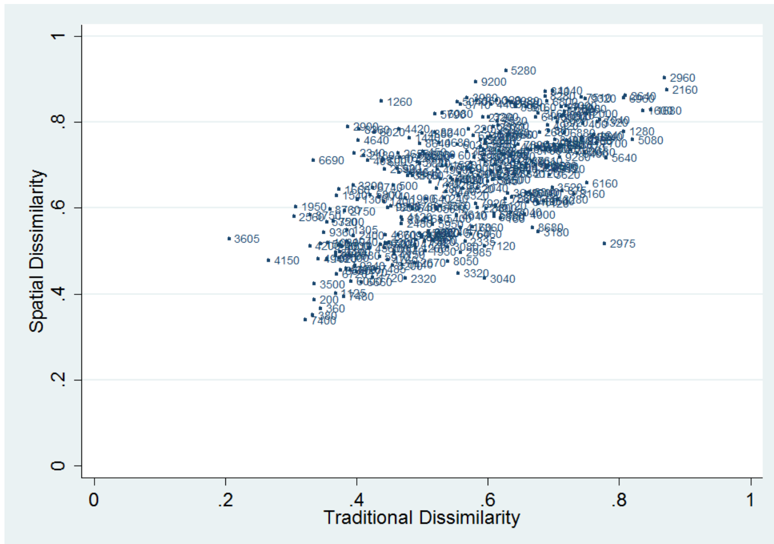
Table 2: Spatial Dissimilarity vs Traditional Dissimilarity (African Americans)

MSA FIPS	Metropolitan Area	Spatial Dissimilarity		Dissimilarity (Blocks)		Dissimilarity (Tracts)	
		Levels	Rank	Levels	Rank	Levels	Rank
	A. Most segregated MSAs in US						
5280	Muncie, IN MSA	0.8785051	1	0.7022	141	0.5282	150
2960	Gary, IN PMSA	0.8747661	2	0.8602	4	0.8093	2
2160	Detroit, MI PMSA	0.8701484	3	0.8655	3	0.8405	1
8080	Steubenville–Weirton, OH–WV MSA	0.848863	4	0.7648	58	0.6256	60
6960	Saginaw–Bay City–Midland, MI MSA	0.8471904	5	0.8123	19	0.7334	12
1320	Canton–Massillon, OH MSA	0.8457054	6	0.738	89	0.5774	99
2640	Flint, MI PMSA	0.8411021	7	0.8268	11	0.7646	6
1000	Birmingham, AL MSA	0.8389853	8	0.8157	17	0.6989	20
840	Beaumont–Port Arthur, TX MSA	0.8273058	9	0.7513	74	0.6481	47
5200	Monroe, LA MSA	0.8263328	10	0.8082	22	0.69	27
	B. Least segregated MSAs in US						
6560	Pueblo, CO MSA	0.4168497	322	0.6532	217	0.4069	261
7160	Salt Lake City–Ogden, UT MSA	0.415838	323	0.6598	209	0.4249	243
8735	Ventura, CA PMSA	0.4148834	324	0.5457	305	0.3695	286
1125	Boulder–Longmont, CO PMSA	0.4105108	325	0.6155	261	0.3239	311
7480	St Barbara–St Maria–Lompoc, CA MSA	0.4095207	326	0.5629	295	0.3894	271
5170	Modesto, CA MSA	0.394449	327	0.572	291	0.3212	313
200	Albuquerque, NM MSA	0.3794953	328	0.5505	303	0.312	319
380	Anchorage, AK MSA	0.3729775	329	0.4489	328	0.3336	308
5945	Orange County, CA PMSA	0.3686204	330	0.5072	318	0.3391	305
7400	San Jose, CA PMSA	0.3256682	331	0.4817	323	0.2939	325
	C. Most populated MSAs in US						
4480	Los Angeles–Long Beach, CA PMSA	0.6148579	177	0.6266	252	0.5765	102
5600	New York, NY PMSA	0.6903519	97	0.7013	142	0.6714	38
1600	Chicago, IL PMSA	0.7632357	35	0.8215	15	0.7789	4
6160	Philadelphia, PA–NJ PMSA	0.7276239	63	0.7565	69	0.6897	28
8840	Washington, DC–MD–VA–WV PMSA	0.651122	144	0.6449	227	0.5958	80
2160	Detroit, MI PMSA	0.8701484	3	0.8655	3	0.8405	1
3360	Houston, TX PMSA	0.7056391	81	0.6578	210	0.5716	106
520	Atlanta, GA MSA	0.6759976	115	0.6949	157	0.6148	66
1920	Dallas, TX PMSA	0.6365489	156	0.628	250	0.5396	133
1120	Boston, MA–NH PMSA	0.6009404	191	0.7084	132	0.6364	54

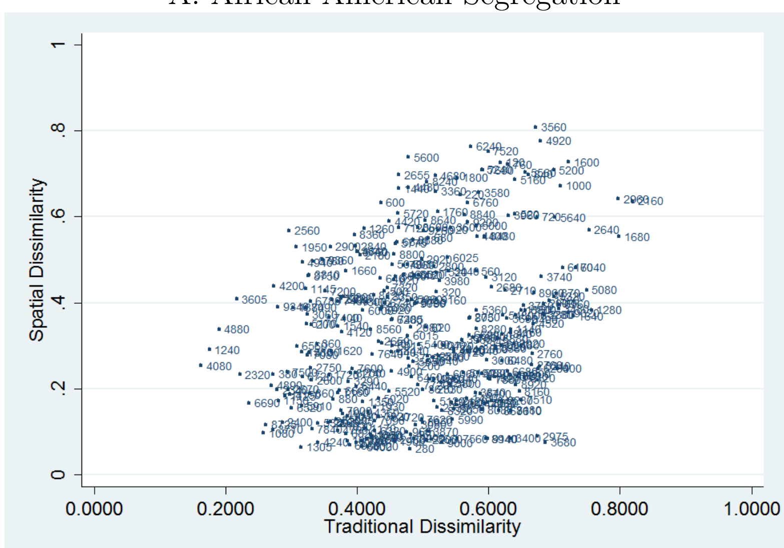
Spatial Dissimilarity is the average of the individual spatial dissimilarity. The traditional dissimilarity is computed using Census blocks and Census tracts data from the Summary File 1, Census 2000.

In panel A of Table 6, I focus on high school graduation. The impact of segregation is negative but not statistically significant, when using the traditional dissimilarity index; while the spatial dissimilarity is highly significant. In column 3 and 4, I also check if the effect of segregation is different for

Figure 5: Spatial Dissimilarity vs Traditional Dissimilarity



A. African American Segregation



B. Multigroup Segregation

Each point represents a Metropolitan Statistical Area (MSA). The marker of the points is the MSA FIPS code. The vertical axis measures the level of spatial dissimilarity and the horizontal axis the level of the traditional dissimilarity. The latter is computed using census tracts as subunits.

Table 3: Spatial Dissimilarity vs Traditional Dissimilarity (Multigroup)

MSA FIPS	Metropolitan Area	Spatial Dissimilarity		Dissimilarity (Blocks)		Dissimilarity (Tracts)	
		Levels	Rank	Levels	Rank	Levels	Rank
	A. Most segregated MSA in US						
2620	Flagstaff, AZ–UT MSA	0.8667412	1	0.7093	69	0.5808	38
2160	Detroit, MI PMSA	0.8286439	2	0.8198	2	0.7355	1
8080	Steubenville–Weirton, OH–WV MSA	0.8213511	3	0.7397	38	0.5177	86
1000	Birmingham, AL MSA	0.8187241	4	0.8029	5	0.6661	8
5200	Monroe, LA MSA	0.8163867	5	0.8033	4	0.669	7
5280	Muncie, IN MSA	0.8154988	6	0.6843	105	0.4757	120
1320	Canton–Massillon, OH MSA	0.8072924	7	0.7204	55	0.5089	96
2640	Flint, MI PMSA	0.7960778	8	0.799	6	0.6747	6
840	Beaumont–Port Arthur, TX MSA	0.7901801	9	0.738	41	0.6101	24
760	Baton Rouge, LA MSA	0.7901433	10	0.762	21	0.6113	22
	B. Least segregated MSAs in US						
1150	Bremerton, WA PMSA	0.3996123	322	0.437	322	0.2669	303
6560	Pueblo, CO MSA	0.3992773	323	0.4754	306	0.2864	293
4150	Lawrence, KS MSA	0.3982949	324	0.4753	307	0.264	306
1720	Colorado Springs, CO MSA	0.3969479	325	0.4575	313	0.3069	280
5170	Modesto, CA MSA	0.3946211	326	0.4457	317	0.2684	301
1880	Corpus Christi, TX MSA	0.3941522	327	0.4337	323	0.2515	311
7840	Spokane, WA MSA	0.3868918	328	0.5592	248	0.2777	298
380	Anchorage, AK MSA	0.354807	329	0.4051	328	0.2643	305
2320	El Paso, TX MSA	0.2795754	330	0.367	330	0.2017	327
4080	Laredo, TX MSA	0.2771601	331	0.3563	331	0.1072	331
	C. Most populated MSAs in US						
4480	Los Angeles–Long Beach, CA PMSA	0.4834004	270	0.4973	289	0.4091	183
5600	New York, NY PMSA	0.6053643	138	0.6286	183	0.5603	56
1600	Chicago, IL PMSA	0.6563473	90	0.7057	76	0.6141	21
6160	Philadelphia, PA–NJ PMSA	0.6965794	54	0.7306	45	0.6252	16
8840	Washington, DC–MD–VA–WV PMSA	0.5894296	149	0.5949	212	0.5028	100
2160	Detroit, MI PMSA	0.8286439	2	0.8198	2	0.7355	1
3360	Houston, TX PMSA	0.5698989	175	0.5689	237	0.4548	138
520	Atlanta, GA MSA	0.6376615	108	0.6702	126	0.5603	55
1920	Dallas, TX PMSA	0.5586969	188	0.5718	235	0.4478	144
1120	Boston, MA–NH PMSA	0.5336268	220	0.6435	166	0.5215	80

Spatial Dissimilarity is the average of the individual spatial dissimilarity. The traditional dissimilarity is computed using Census blocks and Census tracts data from the Summary File 1, Census 2000.

African Americans. When using the traditional segregation index, the estimated coefficients imply that segregation harms African Americans, since it decreases the probability of high school graduation. The estimates using the spatial dissimilarity have a slightly different interpretation: while segregation

Table 4: Individual Distribution of Spatial Dissimilarity, Quartiles (African Americans)

MSA FIPS	Metropolitan Area	Average	1st Quartile	Median	3rd Quartile
		A. Most Segregated MSAs in US			
5280	Muncie, IN MSA	0.8785051	0.4210913	0.536438	0.536438
2960	Gary, IN PMSA	0.8747661	0.59432	0.624964	0.624964
2160	Detroit, MI PMSA	0.8701484	0.6527996	0.6527996	0.6527996
8080	Steubenville–Weirton, OH–WV MSA	0.848863	0.5131925	0.5205587	0.5205587
6960	Saginaw–Bay City–Midland, MI MSA	0.8471904	0.4719803	0.5582634	0.5582634
1320	Canton–Massillon, OH MSA	0.8457054	0.4002332	0.5365979	0.5365979
2640	Flint, MI PMSA	0.8411021	0.531328	0.6290947	0.6315051
1000	Birmingham, AL MSA	0.8389853	0.6466866	0.7176122	1.0913744
840	Beaumont–Port Arthur, TX MSA	0.8273058	0.5746312	0.6678815	1.0368019
5200	Monroe, LA MSA	0.8263328	0.6165972	0.755943	1.2140182
		B. Least Segregated MSAs in US			
6560	Pueblo, CO MSA	0.4168497	0.198187	0.3978968	0.5100178
7160	Salt Lake City–Ogden, UT MSA	0.415838	0.1909232	0.3650295	0.5057651
8735	Ventura, CA PMSA	0.4148834	0.2094268	0.3707898	0.5103415
1125	Boulder–Longmont, CO PMSA	0.4105108	0.1856612	0.363827	0.504531
7480	Santa Barbara–Santa Maria–Lompoc, CA MSA	0.4095207	0.2085673	0.3972317	0.5123268
5170	Modesto, CA MSA	0.394449	0.2065722	0.3786059	0.5140098
200	Albuquerque, NM MSA	0.3794953	0.1716114	0.3434244	0.5133553
380	Anchorage, AK MSA	0.3729775	0.2025856	0.358544	0.5210976
5945	Orange County, CA PMSA	0.3686204	0.1853323	0.3445561	0.4796011
7400	San Jose, CA PMSA	0.3256682	0.1519909	0.3096957	0.4531364
		C. Most Populated MSAs in US			
4480	Los Angeles–Long Beach, CA PMSA	0.6148579	0.343943	0.4747065	0.5480155
5600	New York, NY PMSA	0.6903519	0.5258307	0.6423153	0.6737719
1600	Chicago, IL PMSA	0.7632357	0.5528326	0.618727	0.6195102
6160	Philadelphia, PA–NJ PMSA	0.7276239	0.498949	0.6210654	0.6286433
8840	Washington, DC–MD–VA–WV PMSA	0.651122	0.4241205	0.6056744	0.6833271
2160	Detroit, MI PMSA	0.8701484	0.6527996	0.6527996	0.6527996
3360	Houston, TX PMSA	0.7056391	0.4178659	0.55851	0.6096372
520	Atlanta, GA MSA	0.6759976	0.4436585	0.6349907	0.7082935
1920	Dallas, TX PMSA	0.6365489	0.3680504	0.5148539	0.5912919
1120	Boston, MA–NH PMSA	0.6009404	0.3859964	0.504475	0.5383391

The average spatial dissimilarity corresponds to the index of segregation for the entire city. Notice that the individual-level segregation can be greater than one, while the average is constrained to be between zero and one for comparability across cities.

harms blacks, it also harms the rest of the population.

The results for college graduation present a similar pattern. Segregation is not significant when measured using the traditional approach, while it is negative and highly significant when using the spatial approach. Panel C shows that for idleness, the estimated relationship between segregation and

Table 5: Individual Distribution of Spatial Dissimilarity, Quartiles (Multigroup)

MSA FIPS	Metropolitan Area	Average	1st Quartile	Median	3rd Quartile
		A. Most Segregated MSAs in US			
2620	Flagstaff, AZ–UT MSA	0.8667412	0.53719981	0.67667898	1.44508612
2160	Detroit, MI PMSA	0.8286439	0.4994092	0.54727524	0.54727524
8080	Steubenville–Weirton, OH–WV MSA	0.8213511	0.09468782	0.10193237	0.10193237
1000	Birmingham, AL MSA	0.8187241	0.57186057	0.64343456	0.94592965
5200	Monroe, LA MSA	0.8163867	0.57393668	0.69740037	1.09410571
5280	Muncie, IN MSA	0.8154988	0.14197748	0.17546687	0.17546687
1320	Canton–Massillon, OH MSA	0.8072924	0.1274268	0.16439563	0.16460785
2640	Flint, MI PMSA	0.7960778	0.37310235	0.44017288	0.4636971
840	Beaumont–Port Arthur, TX MSA	0.7901801	0.53031412	0.61211039	0.95667631
760	Baton Rouge, LA MSA	0.7901433	0.54107325	0.67690541	1.0536933
		B. Least Segregated MSAs in US			
1150	Bremerton, WA PMSA	0.3996123	0.1176552	0.17211191	0.22281765
6560	Pueblo, CO MSA	0.3992773	0.14755533	0.2813778	0.43960199
4150	Lawrence, KS MSA	0.3982949	0.10488319	0.18351015	0.24649722
1720	Colorado Springs, CO MSA	0.3969479	0.15178639	0.22656594	0.31751936
5170	Modesto, CA MSA	0.3946211	0.19590742	0.31616467	0.47983551
1880	Corpus Christi, TX MSA	0.3941522	0.17305489	0.30298449	0.46049655
7720	Sioux City, IA–NE MSA	0.3868918	0.24153557	0.31525281	0.31525281
380	Anchorage, AK MSA	0.354807	0.15469473	0.26151132	0.36095664
2320	El Paso, TX MSA	0.2795754	0.10509243	0.18472356	0.316258
4080	Laredo, TX MSA	0.2771601	0.07141443	0.1421759	0.2571718
		C. Most Populated MSAs in US			
4480	Los Angeles–Long Beach, CA PMSA	0.4834004	0.3974564	0.56927795	0.73234758
5600	New York, NY PMSA	0.6053643	0.60649702	0.74460922	0.88614442
1600	Chicago, IL PMSA	0.6563473	0.4863418	0.58390268	0.65356618
6160	Philadelphia, PA–NJ PMSA	0.6965794	0.41789015	0.50060683	0.53834268
8840	Washington, DC–MD–VA–WV PMSA	0.5894296	0.4377141	0.59586204	0.7592566
2160	Detroit, MI PMSA	0.8286439	0.4994092	0.54727524	0.54727524
3360	Houston, TX PMSA	0.5698989	0.41500124	0.55513781	0.70744205
520	Atlanta, GA MSA	0.6376615	0.44326551	0.61122652	0.71895259
1920	Dallas, TX PMSA	0.5586969	0.36111587	0.50142262	0.62570005
1120	Boston, MA–NH PMSA	0.5336268	0.21691148	0.27838146	0.32705205

The average spatial dissimilarity corresponds to the index of segregation for the entire city. Notice that the individual-level segregation can be greater than one, while the average is constrained to be between zero and one for comparability across cities.

outcomes using the two approaches has the same qualitative implications. However the magnitudes of the estimated effects are different.

The general result is that the estimated correlation between segregation and outcomes is different when using the spatial approach. To provide some insights, I analyze the differences between traditional and spatial dissimilar-

Table 6: Segregation and Outcomes

	A. High School Graduation			
	trad	spatial	trad	spatial
Segregation	-0.012	-0.086	0.016	-0.068
	(0.023)	(0.026)***	(0.024)	(0.028)**
Segregation * black			-0.251	-0.182
			(0.045)***	(0.054)***
Observations	139634	139634	139634	139634
R-squared	0.037	0.037	0.037	0.038
	B. College Graduation			
	trad	spatial	trad	spatial
Segregation	-0.019	-0.148	-0.014	-0.151
	(0.062)	(0.066)**	(0.067)	(0.070)**
Segregation * black			-0.05	0.025
			(0.051)	(0.052)
Observations	139634	139634	139634	139634
R-squared	0.041	0.042	0.041	0.042
	C. Idleness			
	trad	spatial	trad	spatial
Segregation	0.036	0.031	0.005	0.012
	(0.025)	(0.023)	(0.025)	(0.023)
Segregation * black			0.271	0.191
			(0.039)***	(0.053)***
Observations	139634	139634	139634	139634
R-squared	0.05	0.05	0.051	0.05

* significant at 10; ** significant at 5; *** significant at 1.

Standard errors corrected for clustering at the MSA level in parentheses. The sample contains all 25-30 years old individuals born in US. I consider only the MSAs for which the fiscal variables instruments are available. Controls included but not shown: fraction of blacks in MSA, dummies for race (black, asian, hispanic and other nonwhite), dummy for female, age dummies, log of population in MSA, log of median income in MSA, manufacturing share of MSA. The last three variables are also included interacted with the black dummy.

ity using regression analysis. In Table F.11, I regress the absolute difference between traditional and spatial dissimilarity on several metropolitan area characteristics. The difference decreases with income levels, population and geographic density; it increases with the fraction of blacks. For the dissimilarity (columns 1-3) the difference increases with the fraction of other race and the number of tracts, while it decreases with the fraction of workers in manufacturing and the fraction of people living in the urban area.

These results indicate that the spatial dissimilarity and the traditional

Table 7: Segregation and Outcomes, Instrumental Variables

	A. Individuals 25-30 years old					
	High School Graduation		*College Graduation*		*Idleness*	
	trad	spatial	trad	spatial	trad	spatial
Segregation	0.077	0.129	0.388	0.582	-0.142	-0.184
	(0.086)	(0.197)	(0.185)**	(0.503)	(0.071)**	(0.15)
Segregation * black	-0.222	-0.499	-0.305	-0.52	0.399	0.788
	(0.130)*	(0.333)	(0.154)**	(0.471)	(0.126)***	(0.365)**
Observations	138847	138847	138847	138847	138847	138847

	B. Individuals 20-24 years old					
	High School Graduation		*College Graduation*		*Idleness*	
	trad	spatial	trad	spatial	trad	spatial
Segregation	0.142	0.285	0.332	0.473	-0.116	-0.216
	(0.091)	(0.259)	(0.113)***	(0.346)	(0.051)**	(0.149)
Segregation * black	-0.528	-1.086	-0.382	-0.623	0.277	0.975
	(0.149)***	(0.517)**	(0.101)***	(0.355)*	(0.172)	(0.387)**
Observations	97338	97338	97338	97338	97338	97338

* significant at 10; ** significant at 5; *** significant at 1.

Standard errors corrected for clustering at the MSA level in parentheses. The sample contains all 25-30 years old (Panel A) and 20-24 years old (Panel B) individuals born in US from the 1% PUMS 1990. Controls included but not shown: fraction of blacks in MSA, dummies for race (black, asian, hispanic and other nonwhite), dummy for female, age dummies, log of population in MSA, log of median income in MSA, manufacturing share of MSA. The last three variables are also included interacted with the black dummy.

dissimilarity could provide different pictures of the segregation in metropolitan areas with low density and high fraction of blacks. Understanding the structural reasons of these differences requires a more structural approach and it is beyond the scope of the present paper.

To further explore the explanatory power of the spatial index, Table F.12 contains a simple exercise. I regress high school graduation on the traditional dissimilarity, controlling for the spatial dissimilarity. If the spatial dissimilarity is not significant, one would conclude that it does not add additional information to the one already contained in the traditional dissimilarity.

The spatial dissimilarity is significant, therefore suggesting that the spatial approach provides additional information with respect to the traditional neighborhood approach. The estimates obtained using the spatial exposure index confirm the latter result. The spatial index is highly significant in regressions using college graduation as dependent variable.[27]

The tests reported above are only suggestive. If the additional explanatory power of the spatial approach is the result of measurement error, an instrumental variable approach should correct for the differences estimated in Table 6. To answer this question, I provide instrumental variables estimates of the relationship between segregation and outcomes. I use the number of intercounty and intracounty rivers in the metropolitan area as instruments for segregation.[28]. Rivers provide geographical barriers, dividing the metropolitan area into subunits and creating a natural landscape for segregation. The more rivers the higher is the expected segregation. The instrumental variable approach was developed by Hoxby (2000) and used as instrument for segregation in Cutler and Glaeser (1997). I follow the latter for the implementation of the IV estimates, including a quadratic term to control for nonlinearities.

The estimates are contained in Table 7. While there is general accordance in the signs of the effects, most estimates are not significant under the spatial approach. In Panel A, for individuals aged 25 to 30, the effect of segregation on education is not significant when using the spatial approach, while the estimates using the traditional approach imply that segregation decreases the probability of graduation especially for African Americans. Higher residential segregation increases the probability of being idle for African Americans, while decreasing it for the rest of the population, when using the traditional index. The effect of the spatial dissimilarity is not significant. Similar differences are shown for the sample of individuals 20-24 years old in Panel B. These findings are confirmed in Table 8, where I compare regressions using the traditional exposure (isolation) and the spatial exposure indices. The instrumental variable estimates show that the differences in spatial and traditional coefficients are not due to measurement error only.

Tables 4 and 5 show that the individual segregation distribution is not bell-shaped for most cities: few very highly segregated individuals are responsible for the high average segregation measured in several metropolitan

[27]However, in the case of idleness, all the coefficients lose significance.

[28]The data are from Rothstein (2007)

Table 8: Segregation and Outcomes, Instrumental Variables (Exposure)

	A. Individuals 25-30 years old					
	High School Graduation		*College Graduation*		*Idleness*	
	trad	spatial	trad	spatial	trad	spatial
Segregation	0.037	0.172	0.281	0.578	-0.107	-0.213
	(0.067)	(0.176)	(0.143)**	(0.43)	(0.057)*	(0.143)
Segregation * black	-0.141	-0.593	-0.193	-0.687	0.277	0.879
	(0.099)	(0.349)*	(0.116)*	(0.488)	(0.091)***	(0.382)**
Observations	138847	138847	138847	138847	138847	138847

	B. Individuals 20-24 years old					
	High School Graduation		*College Graduation*		*Idleness*	
	trad	spatial	trad	spatial	trad	spatial
Segregation	0.073	0.285	0.236	0.401	-0.078	-0.234
	(0.071)	(0.231)	(0.084)***	(0.274)	(0.039)**	(0.148)
Segregation * black	-0.348	-1.176	-0.266	-0.677	0.129	1.066
	(0.102)***	(0.536)**	(0.069)***	(0.354)*	(0.133)	(0.416)**
Observations	97338	97338	97338	97338	97338	97338

* significant at 10; ** significant at 5; *** significant at 1.

Standard errors corrected for clustering at the MSA level in parentheses. The sample contains all 25-30 years old (Panel A) and 20-24 years old (Panel B) individuals born in US from the 1% PUMS 1990. Controls included but not shown: fraction of blacks in MSA, dummies for race (black, asian, hispanic and other nonwhite), dummy for female, age dummies, log of population in MSA, log of median income in MSA, manufacturing share of MSA. The last three variables are also included interacted with the black dummy.

areas. A more robust indicator of segregation at the city level is the median segregation. The estimated effect of median spatial dissimilarity on outcomes is reported in Table 9. The results indicate that segregation does not have any impact on the socioeconomic outcomes analyzed here. In addition to the previous evidence, this table points to a general non-robustness of the effect of segregation on outcomes.

In conclusion, the reduced form estimates presented here do not clarify what is the mechanism through which segregation may affect outcomes. A full fledged structural model would shed some light on the reasons of the huge discrepancies among the spatial approach and the traditional approach.

Table 9: Segregation and Outcomes, Median Segregation (IV)

	A. Individuals 25-30 years old					
	High School Graduation		*College Graduation*		*Idleness*	
Median Segr	-0.438	-0.433	0.749	0.651	-0.509	-0.405
	(0.393)	(0.391)	(0.949)	(0.88)	(0.442)	(0.367)
Median Segr * black		0.764		0.297		-3.497
		(1.379)		(1.507)		(3.847)
Observations	138847	138847	138847	138847	138847	138847

	B. Individuals 20-24 years old					
	High School Graduation		*College Graduation*		*Idleness*	
Median Segr	-0.588	-0.601	0.813	0.83	-0.12	0.206
	(0.57)	(0.539)	(0.842)	(0.812)	(0.275)	(0.265)
Median Segr * black		3.552		-0.734		-2.318
		(3.11)		(1.061)		(3.092)
Observations	97338	97338	97338	97338	97338	97338

* significant at 10; ** significant at 5; *** significant at 1.

Standard errors corrected for clustering at the MSA level in parentheses. Segregation is measured as the median segregation level of the metropolitan area. The sample contains all 25-30 years old (Panel A) and 20-24 years old (Panel B) individuals born in US from the 1% PUMS 1990. Controls included but not shown: fraction of blacks in MSA, dummies for race (black, asian, hispanic and other nonwhite), dummy for female, age dummies, log of population in MSA, log of median income in MSA, manufacturing share of MSA. The last three variables are also included interacted with the black dummy.

6. Conclusion

This paper provides new evidence on the effect of residential segregation on socioeconomic outcomes, using a new family of indices derived from spatial statistics. The proposed index of segregation takes individual locations and their racial groups as primitives and constructs the entire distribution of segregation in the metropolitan area. I proxy for the segregation of the city using the average individual segregation levels. I construct an index of spatial dissimilarity and an index of spatial exposure.

Using Census data, I compare the spatial approach to the traditional indices of segregation, showing that there are differences in the measured segregation. The difference are more pronounced in metropolitan areas with smaller population, lower population density, higher fraction of blacks and lower income levels.

I study the effect of segregation on education and idleness, comparing the traditional and spatial approach. My results show that the two approaches provide different results. These differences are not driven by pure measurement error: I correct for the endogeneity of racial sorting in the city using instrumental variables (inter- and intracounty rivers in the metropolitan area), finding that the differences among the approaches persist.

Since the individual segregation distribution is not bell-shaped, few very highly segregated individuals drive the high segregation levels measured in several metropolitan areas. A more robust indicator of segregation at the city level is the median segregation. When I use the median individual segregation as a proxy for city-level segregation, the effect of segregation on outcomes disappears (it is not significant).

This empirical work provides suggestive evidence that reduced-form estimates of the effect of segregation on outcomes are not robust to the spatial approach. A micro-founded economic model, in which segregation and outcomes are determined as equilibrium quantities, could help to shed light on the economic reasons behind these results.

References

Alesina, A., Zhuravskaya, E., August 2011. Segregation and the quality of government in a cross section of countries. American Economic Review 101, 1872–1911.

Ananat, E., 2011. The wrong side(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., Washington, E., 2009. Segregation an black political efficacy. Journal of Public Economics 5-6, 807–822.

Arbia, G., Copetti, M., Diggle, P., 2008. Growth in Interconnected Territories. Springer-Verlag, Ch. Modelling individual behaciour of firms in the study of spatial concentration.

Berman, M., Diggle, P., 1989. Estimating weighted integrals of the second-order intensity of a spatial point process. Journal of the Royal Stastistical Society, Series B 51 (1), 81–92.

Card, D., Rothstein, J., 2007. Racial segregation and the black-white test score gap. Journal of Public Economics 91 (11-12), 2158–2184.

Collins, W. J., Margo, R. A., 2000. Residential segregation and socioeconomic outcomes: When did ghettos go bad. Economic Letters 69, 239–243.

Conley, T. G., 1999. Gmm estimation with cross sectional dependence. Journal of Econometrics 92 (1), 1–45.

Cutler, D., Glaeser, E., Vigdor, J., 1999. The rise and decline of the american ghetto. Journal of Political Economy 107 (3), 455–506.

Cutler, D. M., Glaeser, E. L., 1997. Are ghettos good or bad? Quarterly Journal of Economics 112 (3), 827–872.

Diggle, P., 1985. A kernel method for smoothing point process data. Applied Statistics 34 (2), 138–147.

Diggle, P., 2003. Statistical analysis of spatial point patterns. Academic Press, London, Second Edition.

Diggle, P., Zheng, P., Durr, P., 2005. Nonparametric estimation of spatial segregation in a multivariate point process: bovine tubercolosis in cornwall, uk. Applied Statistics 54 (3), 645–658.

Diggle, P. J., Eglen, S. J., Troy, J. B., 2006. Case Studies in Spatial Point Process Modelling. Springer Lecture notes in Statistics 185, Ch. Modelling the Bivariate Spatial Distribution of Amacrine Cells, pp. 215–233.

Echenique, F., Fryer, R., 2007. A measure of segregation based on social interactions. Quarterly Journal of Economics 122 (2), 441–485.

Ferrara, E. L., Mele, A., 2011. Racial segregation and public school expenditure, unpublished manuscript, IGIER Bocconi University.

Frankel, D., Volij, O., 2011. Measuring school segregation. Journal of Economic Theory 146, 1–38.

Hoxby, C. M., December 2000. Does competition among public schools benefit students and taxpayers? American Economic Review 90 (5), 1209–1238.

Hutchens, R., 2004. One measure of segregation. International Economic Review 45 (2), 555–578.

Kelsall, J. E., Diggle, P. J., 1998. Spatial variation in risk of diseases: A nonparametric binary regression approach. Applied Statistics.

Koenker, R., Mizera, I., 2004. Penalized triograms: Total variation regularization for bivariate smoothing. Journal of Royal Statistical Society, Series B (Statistical Methodology) 66 (1), 145–163.

Massey, D., Denton, N., 1988. The dimensions of residential segregation. Social Forces 67 (2), 281–315.

Massey, D., Denton, N., 1993. American Apartheid: Segregation and the Making of the Underclass. Cambridge: Harvard University Press.

Moller, J., Waagepetersen, R. P., 2004. Statistical inference and simulation for spatial point processes. Monographs on Statistics and Applied Probability 100, Chapman and Hall.

Quah, D., Simpson, H., 2003. Spatial cluster empirics, working Paper, London School of Economics.

Reardon, S. F., O'Sullivan, D., 2004. Measures of spatial segregation. Sociological Methodology 34.

Rothstein, J., December 2007. Does competition among public schools benefit students and taxpayers? a comment on hoxby. American Economic Review 5 (97), 2026–2037.

Stoyan, D., Kendall, W., Mecke, J., 1987. Stochasti Geometry and its Applications. John Wiley and Sons.

Stoyan, D., Stoyan, H., 1994. Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics. Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons.

Zhuang, J., Ogata, Y., Vere-Jones, D., 2006. Case Studies in Spatial Point Process Modelling. Springer Lecture notes in Statistics 185, Ch. Diagnostic Analysis of Space-Time Branching Processes for Earthquakes, pp. 275–290.

Appendix A. Background Theory

In this section I briefly review the fundamental concepts and definitions needed to develop my main theoretical results.[29] I provide proofs of some results in Appendix B. The interested reader can refer to the books listed in the references for more details, while the reader familiar with spatial Poisson point processes can skip this appendix.

Appendix A.1. Notation, Basic Properties and Definitions

A spatial point process X is a stochastic mechanism that maps points over a set $\mathcal{S} \subseteq \mathbb{R}^2$. Alternatively it can be defined as a random counting measure over bounded sets $A \subseteq \mathcal{S}$. I denote the random set as $X = \{x_1, ..., x_n\}$, where x_i denotes the generic point of the process. The random variable $N(A)$ indicates the number of points in bounded set $A \subseteq \mathcal{S}$. I denote the realizations of X as x and the realizations of N as n. I write ξ or η to indicate a generic point (coordinate) in $\mathcal{S}$ and x_i for the generic realized point of the process. The area of region A is $|A|$ and $d\xi$ refers to the infinitesimal region containing ξ.

I consider only finite point processes, with realizations x in the set $N_{1f} = \{x \subseteq \mathcal{S} : n(x \cap A) < \infty\}$, for any bounded $A \subseteq \mathcal{S}$. A point process is *stationary* if all the probability statements about the process in any bounded set A of the plane are invariant under arbitrary translations. This implies that all the statistics are invariant under translation, e.g. $\mathbb{E}N(A) = \mathbb{E}N_p(A)$, where $N_p(A)$ is the process X translated by the vector p. A point process is *isotropic* if the invariance holds under arbitrary rotations. A process that is stationary and isotropic is called *motion-invariant.* For convenience I will also assume that the process is *simple* (or *orderly)*, i.e that multiple coincident events cannot occur.

In this paper I consider *simple nonstationary and anisotropic* processes.

Appendix A.2. First and Second Order Properties

Let X be a spatial point process defined over $\mathcal{S} \subseteq \mathbb{R}^2$. The *intensity function* is a locally integrable function[30] $\lambda : \mathcal{S} \to [0, \infty)$, defined as the limit of the expected number of points per infinitesimal area

[29] Diggle (2003), Stoyan, Kendall and Mecke (1987), Stoyan and Stoyan (1994), Moller and Waagepetersen (2004) are the basic references.

[30] A function is locally integrable if $\int_A \lambda(\xi)\, d\xi < \infty$ for all bounded $A \subseteq \mathcal{S}$

$$\lambda(\xi) = \lim_{|d\xi|\to 0} \left\{ \frac{\mathbb{E}[N(d\xi)]}{|d\xi|} \right\} \tag{A.1}$$

A stationary process has constant intensity $\lambda(\xi) = \lambda$ for all ξ. T*he intensity measure* of a point process X is defined for $A \subseteq \mathcal{S}$ as

$$\Lambda(A) = \mathbb{E}N(A) = \int_A \lambda(\xi)\, d\xi \tag{A.2}$$

and measures the expected number of points of the process in the set A. I follow the literature and assume that $\Lambda(A)$ is *locally finite*, i.e. $\Lambda(A) < \infty$ for all bounded $A \subseteq \mathcal{S}$, and *diffuse*, i.e. $\Lambda(\{\xi\}) = 0$, for $\xi \in \mathcal{S}$ (or alternatively $\nexists \xi \in \mathcal{S}$ s.t. $\Lambda(\{\xi\}) > 0$). The fact that $\Lambda(A)$ is diffuse implies that $\mathbb{P}[N(d\xi) > 1] = o(|d\xi|)$: in words, there are no coincident points, and the process is simple.[31]

Appendix A.3. Poisson Processes and Marked Poisson Processes

The Poisson point process is the simplest point process and is widely used in practical applications. The definition of the process consists of two conditions, that also provide a practical algorithm for simulation.

DEFINITION 3. ***(Poisson Point Process)*** *A point process X on $\mathcal{S}$ is a Poisson Point Process with intensity $\lambda(\xi)$ if the following two conditions are satisfied:*

1. *for any bounded $A \subseteq \mathcal{S}$ with $\Lambda(A) < \infty$*

$$\mathbb{P}[N(A) = n] = [\Lambda(A)]^n \frac{\exp[-\Lambda(A)]}{n!}, \qquad n = 0, 1, 2, ... \tag{A.3}$$

[31]The intensity function has also an infinitesimal interpretation, since the fact that $\mathbb{P}[N(d\xi) > 1] = o(|d\xi|)$ implies that $\mathbb{E}[N(d\xi)]$ converges to $\mathbb{P}[N(d\xi) = 1]$ as $|d\xi| \to 0$. It follows that the quantity $\lambda(\xi)\, d\xi$ can be interpreted as the probability of an event in the infinitesimal region $d\xi$, i.e $\lambda(\xi)\, d\xi \approx \mathbb{P}[N(d\xi) = 1]$. Analogously notice that $\mathbb{E}[N(d\eta)\, N(d\xi)] \approx \mathbb{P}[N(d\eta) = N(d\xi) = 1]$, for ξ and η close, and we can interpret the quantity $\lambda_2(\xi, \eta)\, d\xi d\eta$ as the probability of observing two events in the infinitesimal regions $d\xi$ and $d\eta$.

2. *for any* $n \in \mathbb{N}$ *and any bounded* $A \subseteq \mathcal{S}$ *with* $0 < \Lambda(A) < \infty$ *, conditional on* $N(A) = n$ *the point are i.i.d. over* $\mathcal{S}$ *with density*

$$f(\xi) = \frac{\lambda(\xi)}{\int_A \lambda(\xi)\, d\xi} \tag{A.4}$$

We will write $X \sim Poi(\mathcal{S}, \lambda(\xi))$.

The first condition requires that for any bounded set the number of points of the process is a draw from the Poisson distribution with mean $\Lambda(A) = \int_A \lambda(\xi)\, d\xi$, implying $\mathbb{E}N(A) = \Lambda(A)$ for any bounded $A \subseteq \mathcal{S}$. The second condition requires that, conditioning on the number of points, the locations are i.i.d. draws from a density function proportional to the intensity function. Therefore the intensity function entirely characterizes the process.

Sometimes condition (A.4) is replaced by the *independent scattering* property: if $X \sim Poi(\mathcal{S}, \lambda(\xi))$, then for *disjoint* sets $A_1, A_2, A_3, ...A_K \subseteq A$ the random variables $N(A_1), N(A_2), ..., N(A_K)$ are stochastically independent Poisson random variables, i.e.

$$\mathbb{P}[N(A_1) = n_1, ..., N(A_K) = n_K] = \prod_{k=1}^{K} [\Lambda(A_k)]^{n_k} \frac{\exp[-\Lambda(A_k)]}{n_k!} \tag{A.5}$$

for $n = n_1 + n_2 + ... + n_k$. In Appendix B, I prove that conditions (A.3) and (A.4) imply (A.5).

In this paper I consider only Inhomogeneous Poisson Point Processes (IPP): these processes are nonstationary and anisotropic, with spatially varying intensity function.[32] The IPP is a very simple and parsimonious model

[32] A Poisson Point Process is said Homogeneous (or stationary) if $\lambda(\xi) = \lambda$, for all $\xi \in \mathcal{S}$ and $f(\xi) = |A|^{-1}$, for any bounded $A \subseteq \mathcal{S}$. It follows that for an Homogeneous Poisson Process (HPP) $\mathbb{E}N(A) = \lambda|A|$. The HPP is considered the ideal of *complete spatial randomness* in literature. Complete spatial randomness means that we do not expect the intensity of the process to vary over the region we are considering and that there are no interactions amongst different events. Indeed, by condition (A.3) and the fact that $\lambda(\xi) = \lambda$, an HPP shows stationarity and isotropy, cause $N(A) \sim Poisson(\lambda|A|)$, and thus the expected number of events does not vary over the planar region A; by condition (A.4) and $f(\xi) = |A|^{-1}$, we have no clustering or inhibition (the presence of a point in ξ does not make more or less likely the occurrence of an event η in the neighborhood of ξ).

for clustered points. Notice that the clustering of locations arises only *exogenously*, being a consequence of the intensity specification: there is no behavioral interpretation of points clusters.

In Appendix B, I show that a point process X is Poisson *if and only if* its probability law is[33]

$$\mathbb{P}\left[(X \cap A) \in F\right] = \sum_{n=0}^{\infty} \frac{\exp\left[-\Lambda(A)\right]}{n!} \int_A \cdot \cdot \int_A \mathbf{1}_{[\{x_1,...,x_n\} \in F]} \prod_{i=1}^{n} \lambda(x_i)\, dx_1 \cdot \cdot dx_n \tag{A.6}$$

for all $A \subseteq \mathcal{S}$, with $\Lambda(A) = \int_A \lambda(\xi)\, d\xi < \infty$, and for all $F \subseteq N_{1f}$. By convention for $n = 0$, I write $\mathbf{1}_{[\emptyset \in F]}$. The probability over $\mathcal{S} \subseteq \mathbb{R}^2$ is obtained by substituting A with $\mathcal{S}$.

It is possible to enrich the Poisson model, assigning to each point a random variable (or vector) representing an attribute: this random variable is called *mark* and the process is called Marked Poisson Process.

More formally, let X_0 be a spatial point process defined over the space $\mathcal{S} \subseteq \mathbb{R}^2$. If there is a random mark $m(\xi) \in \mathcal{M}$ attached to each point $\xi \in X_0$ then the process

$$X = \{\{\xi, m(\xi)\} |\, \xi \in X_0\}$$

is called *Marked Point Process* with events in $\mathcal{S}$ and marks in $\mathcal{M}$. The mark space $\mathcal{M}$ may be a finite set, i.e. $\mathcal{M} = \{1, 2, ..., M\}$, in which case X is called a *multitype process*, or a more general set $\mathcal{M} \subseteq \mathbb{R}^q$, $q \geq 1$.

DEFINITION 4. ***(Marked Poisson Process)*** *The process $X = \{\{\xi, m(\xi)\} |\, \xi \in X_0\}$ is a Marked Poisson Process if*

1. *X_0 is a Poisson Point Process over $\mathcal{S}$ with intensity function $\lambda_0(\xi)$ (with $\int_A \lambda_0(\xi)\, d\xi < \infty$ for all bounded $A \subseteq \mathcal{S}$)*
2. *conditional on X_0 the marks $\{m(\xi) |\, \xi \in X_0\}$ are mutually independent*

The framework developed in the paper is based on the simple processes described above.

[33] See also Proposition 3.1 in Moller and Waagepetersen (2004).

Appendix B. Point Processes Theory

Appendix B.1. Independent Scattering Property

PROPOSITION If $X \sim Poi\left(\mathcal{S}, \lambda\left(\xi\right)\right)$, then for disjoint sets $A_1, A_2, A_3, ...A_k \subseteq A$ the random variables $N\left(A_1\right), N\left(A_2\right), N\left(A_3\right), ...$ are stochastically independent, i.e.

$$\mathbb{P}\left[N\left(A_1\right)=n_1,...,N\left(A_k\right)=n_k\right]=\prod_{j=1}^{k}\left[\Lambda\left(A_j\right)\right]^{n_j}\frac{\exp\left[-\Lambda\left(A_j\right)\right]}{n_j!} \qquad \text{(B.1)}$$

for $n = n_1 + n_2 + ... + n_k$.

Proof. Consider the case in which we have only two disjoint sets, i.e. $A = A_1 \cup A_2$. The extension to k sets is done by induction. Conditional on $N\left(A\right) = n_1 + n_2 = n$, $\mathbb{P}\left[\xi \in \left(X \cap A\right)\right] = f\left(\xi\right) = \lambda\left(\xi\right)/\Lambda\left(A\right)$. Then given $N\left(A\right) = n$,

$$\mathbb{P}\left[N\left(A_1\right)=1|\, N\left(A\right)=n\right]=\int_{A_1} f\left(\xi\right)d\xi=\frac{\Lambda\left(A_1\right)}{\Lambda\left(A\right)}$$

and by condition (1) of the definition of a Poisson process, $\mathbb{P}\left[N\left(A_1\right)=n_1|\, N\left(A\right)=n\right]=\left[\frac{\Lambda(A_1)}{\Lambda(A)}\right]^{n_1}$ and also

$$\begin{aligned}\mathbb{P}\left[N\left(A_1\right)=n_1,N\left(A_2\right)=n_2|\, N\left(A\right)=n\right] &= \binom{n_1+n_2}{n_1}\left[\frac{\Lambda\left(A_1\right)}{\Lambda\left(A\right)}\right]^{n_1}\left[\frac{\Lambda\left(A_2\right)}{\Lambda\left(A\right)}\right]^{n_2}\\ &= \frac{n!}{n_1!\left(n-n_1\right)!}\frac{\left[\Lambda\left(A_1\right)\right]^{n_1}\left[\Lambda\left(A_2\right)\right]^{n-n_1}}{\Lambda\left(A\right)^{n}}\end{aligned}$$

and thus condition (2) of the definition of a Poisson process implies that the unconditional probability is

$$\begin{aligned}\mathbb{P}\left[N\left(A_1\right)=n_1,N\left(A_2\right)=n_2\right] &= \frac{n!}{n_1!\left(n-n_1\right)!}\frac{\left[\Lambda\left(A_1\right)\right]^{n_1}\left[\Lambda\left(A_2\right)\right]^{n-n_1}}{\left[\Lambda\left(A\right)\right]^{n}}\left[\Lambda\left(A\right)\right]^{n}\frac{\exp\left[-\Lambda\left(A\right)\right]}{n!}\\ &= \left[\Lambda\left(A_1\right)\right]^{n_1}\frac{\exp\left[-\Lambda\left(A_1\right)\right]}{n_1!}\left[\Lambda\left(A_2\right)\right]^{n-n_1}\frac{\exp\left[-\Lambda\left(A_2\right)\right]}{\left(n-n_1\right)!}\end{aligned}$$

■

Appendix B.2. Probability Law of a Poisson Point Process

PROPOSITION A point process X is a Poisson Point Process, i.e $X \sim Poi(\mathcal{S}, \lambda(\xi))$, if and only if for all $A \subseteq \mathcal{S}$, with $\Lambda(A) = \int_A \lambda(\xi)\, d\xi < \infty$, and for all $F \subseteq N_{1f}$

$$\mathbb{P}[(X \cap A) \in F] = \sum_{n=0}^{\infty} \frac{\exp[-\Lambda(A)]}{n!} \int_A \cdot \cdot \int_A \mathbf{1}_{[\{x_1,\ldots,x_n\} \in F]} \prod_{i=1}^{n} \lambda(x_i)\, dx_1 \cdot \cdot dx_n \tag{B.2}$$

where by convention for $n = 0$ we have $\mathbf{1}_{[\emptyset \in F]}$

Proof. Conditioning on $N(A) = n$, a specific realization $\{x_1, ..., x_n\}$ over A has probability $\prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \left[\frac{\lambda(x_i)}{\int_A \lambda(\xi) d\xi}\right]$. Therefore all the possible realizations $\{x_1, ..., x_n\} \in F$ have probability

$$\mathbb{P}[(X \cap A) \in F | N(A) = n] = \int_A \cdot \cdot \int_A \mathbf{1}_{[\{x_1,\ldots,x_n\} \in F]} \prod_{i=1}^{n} \left[\frac{\lambda(x_i)}{\Lambda(A)}\right] dx_1 \cdot \cdot dx_n.$$

In order to get the unconditional probability we just need to multiply by $\mathbb{P}[N(A) = n] = \frac{\exp[-\Lambda(A)]}{n!} \Lambda(A)^n$ and sum for all n, obtaining (B.2).

For the necessary part of the proof just multiply (B.2) inside the sum by $\frac{\Lambda(A)^n}{\Lambda(A)^n}$ and notice you can rewrite the probability as

$$\begin{aligned}
\mathbb{P}[(X \cap A) \in F] &= \sum_{n=0}^{\infty} \frac{\exp[-\Lambda(A)]}{n!} \Lambda(A)^n \int_A \cdot \cdot \int_A \mathbf{1}_{[\{x_1,\ldots,x_n\} \in F]} \prod_{i=1}^{n} \left[\frac{\lambda(x_i)}{\Lambda(A)}\right] dx_1 \cdot \cdot dx_n \\
&= \sum_{n=0}^{\infty} \mathbb{P}[N(A) = n] \times \mathbb{P}[(X \cap A) \in F | N(A) = n]
\end{aligned}$$

where $\mathbb{P}[N(A) = n]$ is a Poisson distribution and $\mathbb{P}[(X \cap A) \in F | N(A) = n]$ is a binomial point process. ■

The probability law of the process over $\mathcal{S} \subseteq \mathbb{R}^2$ is obtained from (B.2), by substituting A with $\mathcal{S}$.

Appendix B.3. The process under A1,A2 and A3 is Poisson

In our framework we use the Marked Poisson Process extensively and we exploit a property that we prove in the following lemma (see also Proposition 3.9 in Moller and Waagepetersen (2004), p. 26).

LEMMA 1 If X satisfies Assumptions 1-3 with $\mathcal{M} \subseteq \mathbb{R}^q$, $q \geq 1$ then $X \sim Poi\left(\mathcal{S} \times \mathcal{M}, \lambda\left(\xi, m\right)\right)$

Proof. Notice that Assumptions 1 and 2 are simply the definition of a Marked Poisson Process. If we add Assumption 2, the probability of a pair (ξ, m) is $f\left(\xi\right)\rho\left(\xi, m\right) = \frac{\lambda_0(\xi)}{\Lambda_0(A)}\rho\left(\xi, m\right)$ for any bounded $A \subseteq \mathcal{S}$. Therefore, conditioning on $N\left(A\right) = n$ we have

$$
\begin{aligned}
& P\left[\left(X \cap A\right) \in F \middle| N\left(A\right) = n\right] \\
= & \int_A \cdot\cdot \int_A \int_{\mathcal{M}} \cdot\cdot \int_{\mathcal{M}} \mathbf{1}_{\left[\left\{(x_1,m_1),\ldots,(x_n,m_n)\right\}\in F\right]} \prod_{i=1}^{n} \left[\frac{\lambda_0\left(x_i\right)}{\Lambda_0\left(A\right)}\rho\left(x_i, m_i\right)\right] dx_1 \cdot\cdot dx_n dm_1 \cdot\cdot dm_n \\
= & \int_{A\times\mathcal{M}} \cdot\cdot \int_{A\times\mathcal{M}} \mathbf{1}_{\left[\left\{(x_1,m_1),\ldots,(x_n,m_n)\right\}\in F\right]} \prod_{i=1}^{n} \left[\frac{\lambda\left(x_i, m_i\right)}{\Lambda_0\left(A\right)}\right] dx_1 \cdot\cdot dx_n dm_1 \cdot\cdot dm_n
\end{aligned}
$$

Therefore the unconditional distribution is

$$
\begin{aligned}
& P\left[\left(X \cap A\right) \in F\right] \\
= & \sum_{n=0}^{\infty} \frac{\exp\left[-\Lambda_0\left(A\right)\right]}{n!} \int_{A\times\mathcal{M}} \cdot\cdot \int_{A\times\mathcal{M}} \mathbf{1}_{\left[\left\{(x_1,m_1),\ldots,(x_n,m_n)\right\}\in F\right]} \prod_{i=1}^{n} \left[\lambda\left(x_i, m_i\right)\right] dx_1 \cdot\cdot dx_n dm_1 \cdot\cdot dm_n
\end{aligned}
$$

Notice that $\int\limits_{A\times\mathcal{M}} \lambda\left(\xi, m\right) d\xi dm = \int\limits_{A} \lambda_0\left(\xi\right)\left[\int\limits_{\mathcal{M}} \rho\left(\xi, m\right) dm\right] d\xi = \int\limits_{A} \lambda_0\left(\xi\right) d\xi =$ $\Lambda_0\left(A\right)$ for any A and define $t = (\xi, m)$ with values in $T = \mathcal{S} \times \mathcal{M}$ and $\lambda\left(t\right) = \lambda_0\left(\xi\right)\rho\left(\xi, m\right)$ to get

$$
\mathbb{P}\left[\left(X \cap A\right) \in F\right] = \sum_{n=0}^{\infty} \frac{\exp\left[-\int_{A\times\mathcal{M}} \lambda\left(t\right) dt\right]}{n!} \int_{A\times\mathcal{M}} \cdot\cdot \int_{A\times\mathcal{M}} \mathbf{1}_{\left[\left\{t_1,\ldots,t_n\right\}\in F\right]} \prod_{i=1}^{n} \left[\lambda\left(t_i\right)\right] dt_1 \cdot\cdot dt_n
$$

It follows from (B.2) that $X \sim Poi\left(T, \lambda\left(t\right)\right)$ or $X \sim Poi\left(\mathcal{S} \times \mathcal{M}, \lambda\left(\xi, m\right)\right)$

■

Appendix B.4. The case of Multitype Point Process

If the process is a multitype point process then the previous proposition can be specialized in the following

LEMMA 2 If a Marked Point Process X with discrete mark space $\mathcal{M} = \{1, 2, ..., M\}$ satisfies Assumptions 1-3, it is equivalent to a multivariate Poisson Process $(X_1, X_2, ..., X_M)$, i.e $X_m \sim Poi\,(S, \lambda_m\,(\xi))$ are mutually independent and $\lambda_m\,(\xi) = \lambda_0\,(\xi)\,\rho_m\,(\xi)$, $m = 1, ..., M$.

Proof. Assumptions 1 and 2 together form the definition of a Multitype Poisson Process. The (IF) part of the proof then just requires to prove that Assumption 3 implies the multivariate poisson process, i.e. that $P\,(m\,(\xi) = m|\, X_0 = x_0) = \rho_m\,(\xi)$ implies $X_m \sim Poi\,(\mathcal{S}, \lambda_m\,(\xi))$ and mutually independent.

(IF) A Poisson Point Process is uniquely determined by its void probabilities (Theorem 3.1 p. 16 in Moller and Waagepetersen (2004))

$$v\,(A) = \mathbb{P}\,[N\,(A) = 0] = \mathbb{P}\,[X \cap A = \emptyset] = \exp\,[-\Lambda\,(A)]$$

Therefore for independent Poisson Processes X_1 and X_2 with intensity measure $\Lambda_1\,(\cdot)$ and $\Lambda_2\,(\cdot)$, their joint distribution is uniquely determined by the joint void probabilities

$$\mathbb{P}\,[X_1 \cap A = \emptyset, X_2 \cap B = \emptyset] = \exp\,[-\Lambda_1\,(A) - \Lambda_2\,(A)]$$

for any bounded $A, B \subseteq \mathcal{S}$. For simplicity consider a multitype point process with $\mathcal{M} = \{1, 2\}$ only: the extension to M types can be proven by induction. Let the intensity functions of the univariate processes be $\lambda_m\,(\xi) = \lambda_0\,(\xi)\,\rho_m\,(\xi)$ with intensity measures $\Lambda_m\,(A) = \int_A \lambda_m\,(\xi)\,d\xi$. The univariate process X_1 can be thought of as obtained from the multitype process X_0 by including $\xi \in X$ in X_1 with probability $P\,(m\,(\xi) = 1|\, X_0 = x_0) = \rho_1\,(\xi)$. Such a process is called an *independent thinning* of X_0 with *retention* probabilities $\rho_1\,(\xi)$. The events are excluded or included independently of each other. Formally the process X_1 can be thought of as the process

$$X_1 = \{\xi \in X_0 : U\,(\xi) \leq \rho_1\,(\xi)\}$$

where $U\,(\xi) \sim U\,[0, 1]$.

Notice that $\Lambda_0(A) = \Lambda_1(A) + \Lambda_2(A)$ and that *conditional* on $\xi \in X_0$, for $\xi \in A$

$$\mathbb{P}[\xi \in X_1] = \int_A \rho_1(\xi) \frac{\lambda_0(\xi)}{\Lambda_0(A)} d\xi$$

The definition of Poisson process then implies that

$$\begin{aligned}
\mathbb{P}[X_1 \cap A = \emptyset] &= \sum_{n=0}^{\infty} \mathbb{P}[N(X_0 \cap A) = n] \times \mathbb{P}[X_1 \cap A = \emptyset | N(X_0 \cap A) = n] \\
&= \sum_{n=0}^{\infty} \frac{\exp[-\Lambda_0(A)]}{n!} \Lambda_0(A)^n \times \\
&\times \int_A \cdot \cdot \int_A \left(\prod_{i=1}^{n} [1 - \rho_1(x_i)] \frac{\lambda_0(x_i)}{\Lambda_0(A)} \right) dx_1 \cdot \cdot dx_n \\
&= \sum_{n=0}^{\infty} \frac{\exp[-\Lambda_0(A)]}{n!} \left[\int_A [1 - \rho_1(\xi)] \lambda_0(\xi) d\xi \right]^n \\
&= \exp[-\Lambda_0(A)] \sum_{n=0}^{\infty} \frac{\left[\int_A \lambda_0(\xi) d\xi - \int_A \rho_1(\xi) \lambda_0(\xi) d\xi \right]^n}{n!} \\
&= \exp[-\Lambda_0(A)] \sum_{n=0}^{\infty} \frac{[\Lambda_0(A) - \Lambda_1(A)]^n}{n!} \\
&= \exp[-\Lambda_0(A)] \exp[\Lambda_0(A) - \Lambda_1(A)] \\
&= \exp[-\Lambda_1(A)]
\end{aligned}$$

Using the same argument we can show that

$$\mathbb{P}[X_2 \cap A = \emptyset] = \mathbb{P}[X_0 \setminus X_1 \cap A = \emptyset] = \exp[-\Lambda_0(A) + \Lambda_1(A)]$$

Therefore we have proven that X_1 and X_2 are Poisson processes. It remains to be shown that they are independent. Rewrite the joint probability of X_1 and X_2 for $A, B \subseteq \mathcal{S}$ as

$$\mathbb{P}[X_1 \cap A = \emptyset, X_2 \cap B = \emptyset] = \mathbb{P}[X \cap (A \cap B) = \emptyset, X_1 \cap A \setminus B = \emptyset, X_2 \cap B \setminus A = \emptyset]$$

Using the independent scattering property of the Poisson Process, for

$A, B \subseteq \mathcal{S}$

$$
\begin{aligned}
& \mathbb{P}\left[X \cap (A \cap B) = \emptyset, X_1 \cap A \diagdown B = \emptyset, X_2 \cap B \diagdown A = \emptyset\right] \\
= \; & \mathbb{P}\left[X \cap (A \cap B) = \emptyset\right] \mathbb{P}\left[X_1 \cap A \diagdown B = \emptyset\right] \mathbb{P}\left[X \diagdown X_1 \cap B \diagdown A = \emptyset\right] \\
= \; & \exp\left[-\Lambda_0 (A \cap B)\right] \exp\left[-\Lambda_1 (A \diagdown B)\right] \exp\left[-\Lambda_0 (B \diagdown A) + \Lambda_1 (B \diagdown A)\right] \\
= \; & \exp\left[-\Lambda_0 (A \cap B) - \Lambda_1 (A \diagdown B) - \Lambda_0 (B \diagdown A) + \Lambda_1 (B \diagdown A) + \Lambda_1 (A \cap B) - \Lambda_1 (A \cap B)\right] \\
= \; & \exp\left[-\Lambda_1 (A) - \Lambda_0 (B) + \Lambda_1 (B)\right] \\
= \; & \exp\left[-\Lambda_1 (A)\right] \exp\left[-\Lambda_0 (B) + \Lambda_1 (B)\right] \\
= \; & \mathbb{P}\left[X_1 \cap A = \emptyset\right] \mathbb{P}\left[X_2 \cap B = \emptyset\right]
\end{aligned}
$$

Then X_1 and X_2 are independent Poisson Processes with intensity $\lambda_m(\xi) = \lambda_0(\xi)\rho_m(\xi)$, $m = 1, 2$. We can extend the argument to $m = 1, .., M$ by induction.

$(ONLY\ IF)$ Remember that the union of independent Poisson Processes is a Poisson Process with the intensity function equal to the sum of the single processes intensities. Therefore $\left(\bigcup_{m=1}^{M} X_m\right) \sim Poi\left(\mathcal{S}, \sum_{m=1}^{M} \lambda_m(\xi)\right) = Poi(\mathcal{S}, \lambda_0(\xi)) = X_0$. This means that the process satisfies Assumption 1. The proof follows from the fact that conditioning on the sum of M independent Poisson variables we obtain a multinomial distribution

$$
\begin{aligned}
\mathbb{P}\left(m(\xi) = m \middle| X_0 = x_0\right) &= \mathbb{P}\left[\xi \in X_m \middle| \xi \in \bigcup_{m=1}^{M} X_m\right] \\
&= \mathbb{P}\left[(\xi \in X_m) \cap \left(\xi \in \bigcup_{m=1}^{M} X_m\right)\right] \times \left(\mathbb{P}\left[\xi \in \bigcup_{m=1}^{M} X_m\right]\right)^{-1} \\
&= \frac{\lambda_m(\xi)}{\sum_{m=1}^{M} \lambda_m(\xi)} = \frac{\lambda_0(\xi)\rho_m(\xi)}{\sum_{m=1}^{M} \lambda_0(\xi)\rho_m(\xi)} = \rho_m(\xi)
\end{aligned}
$$

Therefore also Assumption 3 is satisfied and since Assumption 3 implies Assumption 2, the proof is complete. ■

When the conditional mark distribution does not depend on location, $\rho(\xi, m) = \rho(m)$ for all ξ, then we have *random labelling.*

Appendix C. Statistical Properties of the Indices

If the process X satisfies Assumptions 1-3 it is possible to derive the moments of any index $\mathcal{T}(X)$. The following theorem applies to any possible index based on the above definition: it is therefore a very general result.

THEOREM 1. *If X is a point process satisfying Assumptions 1-3, then the expected value of any index $\mathcal{T}(X)$ is*

$$\begin{aligned}\mathbb{E}[\mathcal{T}(X)] &= \sum_{n=0}^{\infty} \frac{\exp[-\Lambda(\mathcal{S}\times\mathcal{M})]}{n!} \times \\ &\times \int_{\mathcal{S}\times\mathcal{M}} \cdot\cdot \int_{\mathcal{S}\times\mathcal{M}} \mathcal{T}(\{x_i, m_i\}_{i=1}^{n}) \prod_{i=1}^{n} \lambda(x_i, m_i)\, dx_1 \cdot\cdot dx_n dm_1 \cdot\cdot dm_n\end{aligned}$$

More generally the r-th raw moment of $\mathcal{T}(X)$ is

$$\begin{aligned}\mathbb{E}[\mathcal{T}^r(X)] &= \sum_{n=0}^{\infty} \frac{\exp[-\Lambda(\mathcal{S}\times\mathcal{M})]}{n!} \times \\ &\times \int_{\mathcal{S}\times\mathcal{M}} \cdot\cdot \int_{\mathcal{S}\times\mathcal{M}} \mathcal{T}^r(\{x_i, m_i\}_{i=1}^{n}) \prod_{i=1}^{n} \lambda(x_i, m_i)\, dx_1 \cdot\cdot dx_n dm_1 \cdot\cdot dm_n\end{aligned}$$

Proof. If the process satisfies Assumptions 1-3, then it is Poisson over $\mathcal{S}\times\mathcal{M}$ by Lemma 1. Therefore the probability law of X is given by (**??**). Notice that $\mathcal{T}(X)$ is a nonnegative function. Since any nonnegative function can be expressed as a weighted sum of indicator functions, the result follows. The same argument delivers the results for all the moments. ■

THEOREM 2. *Assume X follows a point process satisfying Assumptions 1-3 and the index satisfies Assumption 4. Then*

$$\mathbb{E}[\mathcal{T}(X)] = \mathbb{E}[\phi(\xi)] = \int_{\mathcal{S}} \phi(\xi) \frac{\lambda_0(\xi)}{\Lambda(\mathcal{S})} d\xi \tag{C.1}$$

$$\mathbb{V}[\mathcal{T}(X)] = \mathbb{E}\left[\frac{1}{N(\mathcal{S})}\right] \mathbb{V}[\phi(\xi)] \tag{C.2}$$

Proof. The Poisson assumption allows us to compute the expectation in the following way

$$\mathbb{E}\left[\mathcal{T}\left(X\right)\right]=\sum_{n=0}^{\infty}\mathbb{E}\left[\mathcal{T}\left(X\right)\middle|\,N\left(\mathcal{S}\right)=n\right]\times\mathbb{P}\left[N\left(\mathcal{S}\right)=n\right]$$

It follows that

$$\begin{aligned}
\mathbb{E}\left[\mathcal{T}\left(X\right)\right] &= \mathbb{E}\left[\frac{1}{N\left(\mathcal{S}\right)}\sum_{\xi\in X_0}\phi\left(\xi\right)\right] \\
&= \sum_{n=0}^{\infty}\mathbb{E}\left[\frac{1}{n}\sum_{\xi\in X_0}\phi\left(\xi\right)\middle|\,N\left(\mathcal{S}\right)=n\right]\times\mathbb{P}\left[N\left(\mathcal{S}\right)=n\right] \\
&= \sum_{n=0}^{\infty}\frac{1}{n}\sum_{\xi\in X_0}\mathbb{E}\left[\phi\left(\xi\right)\middle|\,N\left(\mathcal{S}\right)=n\right]\times\mathbb{P}\left[N\left(\mathcal{S}\right)=n\right] \\
&= \sum_{n=0}^{\infty}\frac{1}{n}\left[n\int_{\mathcal{S}}\phi\left(\xi\right)\frac{\lambda_0\left(\xi\right)}{\Lambda_0\left(\mathcal{S}\right)}d\xi\right]\times\mathbb{P}\left[N\left(\mathcal{S}\right)=n\right] \\
&= \int_{\mathcal{S}}\phi\left(\xi\right)\frac{\lambda_0\left(\xi\right)}{\Lambda_0\left(\mathcal{S}\right)}d\xi\sum_{n=0}^{\infty}\mathbb{P}\left[N\left(\mathcal{S}\right)=n\right] \\
&= \int_{\mathcal{S}}\phi\left(\xi\right)\frac{\lambda_0\left(\xi\right)}{\Lambda_0\left(\mathcal{S}\right)}d\xi \\
&= \mathbb{E}\left[\phi\left(\xi\right)\right]
\end{aligned}$$

where the fourth equality follows from the fact that the locations of the poisson process are i.i.d points with density $\frac{\lambda_0(\xi)}{\Lambda_0(\mathcal{S})}$

The variance of the index is computed in several steps

$$
\begin{aligned}
\mathbb{V}\left[\mathcal{T}(X)\right] &= \mathbb{V}\left[\frac{1}{N(\mathcal{S})}\sum_{\xi\in X_0}\phi(\xi)\right] \\
&= \mathbb{E}\left[\left(\frac{1}{N(\mathcal{S})}\sum_{\xi\in X_0}\phi(\xi)\right)^2\right] - \left(\mathbb{E}\left[\frac{1}{N(\mathcal{S})}\sum_{\xi\in X_0}\phi(\xi)\right]\right)^2 \\
&= \mathbb{E}\left[\frac{1}{N(\mathcal{S})^2}\sum_{\xi\in X_0}\phi(\xi)^2\right] + \mathbb{E}\left[\frac{1}{N(\mathcal{S})^2}\sum_{\xi\in X_0}\sum_{\substack{\eta\in X_0\\ \eta\neq\xi}}\phi(\xi)\,\phi(\eta)\right] \\
&- \left(\mathbb{E}\left[\frac{1}{N(\mathcal{S})}\sum_{\xi\in X_0}\phi(\xi)\right]\right)^2
\end{aligned}
$$

The first component of the sum above is

$$
\begin{aligned}
\mathbb{E}\left[\frac{1}{N(\mathcal{S})^2}\sum_{\xi\in X_0}\phi(\xi)^2\right] &= \sum_{n=0}^{\infty}\frac{1}{n^2}\left[n\int_{\mathcal{S}}\phi(\xi)^2\,\frac{\lambda_0(\xi)}{\Lambda_0(\mathcal{S})}d\xi\right]\times\mathbb{P}\left[N(\mathcal{S})=n\right] \\
&= \mathbb{E}\left[\frac{1}{N(\mathcal{S})}\right]\int_{\mathcal{S}}\phi(\xi)^2\,\frac{\lambda_0(\xi)}{\Lambda_0(\mathcal{S})}d\xi \\
&= \mathbb{E}\left[\frac{1}{N(\mathcal{S})}\right]\mathbb{E}\left[\phi(\xi)^2\right]
\end{aligned}
$$

The second component of the sum is

$$
\begin{aligned}
\mathbb{E}\left[\frac{1}{N(\mathcal{S})^2}\sum_{\xi\in X_0}\sum_{\substack{\eta\in X_0\\ \eta\neq\xi}}\phi(\xi)\,\phi(\eta)\right] &= \sum_{n=0}^{\infty}\frac{1}{n^2}\left[n(n-1)\int_{\mathcal{S}}\int_{\mathcal{S}}\phi(\xi)\,\phi(\eta)\,\frac{\lambda_0(\xi)\,\lambda_0(\eta)}{\Lambda_0(\mathcal{S})^2}d\eta d\xi\right] \\
&\times\mathbb{P}\left[N(\mathcal{S})=n\right] \\
&= \mathbb{E}\left[\frac{n-1}{n}\left(\int_{\mathcal{S}}\phi(\xi)\,\frac{\lambda_0(\xi)}{\Lambda_0(\mathcal{S})}d\xi\right)^2\right] \\
&= \left(1-\mathbb{E}\left[\frac{1}{N(\mathcal{S})}\right]\right)\mathbb{E}\left[\phi(\xi)\right]^2
\end{aligned}
$$

where the second equality follows from the i.i.d. condition of the Poisson process, so ξ and η are independent points. Therefore the variance is

$$\begin{aligned}
\mathbb{V}\left[\mathcal{T}(X)\right] &= \mathbb{E}\left[\frac{1}{N(\mathcal{S})}\right]\mathbb{E}\left[\phi(\xi)^2\right] + \\
&\quad + \left(1 - \mathbb{E}\left[\frac{1}{N(\mathcal{S})}\right]\right)\mathbb{E}\left[\phi(\xi)\right]^2 \\
&\quad - \mathbb{E}\left[\phi(\xi)\right]^2 \\
&= \mathbb{E}\left[\frac{1}{N(\mathcal{S})}\right]\left[\mathbb{E}\left[\phi(\xi)^2\right] - \mathbb{E}\left[\phi(\xi)\right]^2\right] \\
&= \mathbb{E}\left[\frac{1}{N(\mathcal{S})}\right]\mathbb{V}\left[\phi(\xi)\right]
\end{aligned}$$

■

PROOF OF PROPOSITION 1

Consider the quantity $\sum_{m\in\mathcal{M}} |\rho_m(\xi) - \rho_m|$. Under complete segregation, for all $\xi \in X_0$, $\exists m^* \in \mathcal{M}$ such that $\rho_{m^*}(\xi) = 1$ and $\rho_m(\xi) = 0$ for any $m \neq m^*$. The probability of m^* is ρ_{m^*}, therefore

$$\begin{aligned}
\sum_{m\in\mathcal{M}} |\rho_m(\xi) - \rho_m| &= \rho_1 |1 - \rho_1| + (1 - \rho_1)|0 - \rho_1| + ... \\
&\quad ... + \rho_M |1 - \rho_M| + (1 - \rho_M)|0 - \rho_M| \\
&= 2\rho_1(1 - \rho_1) + ... + 2\rho_M(1 - \rho_M) \\
&= 2\sum_{m\in\mathcal{M}} \rho_m(1 - \rho_m) \\
&= 2I
\end{aligned}$$

The second part follows the same lines. Consider the quantity $\sum_{m\in\mathcal{M}} (\rho_m(\xi) - \rho_m)^2$. Under complete segregation, for all $\xi \in X_0$, $\exists m^* \in \mathcal{M}$ such that $\rho_{m^*}(\xi) = 1$ and $\rho_m(\xi) = 0$ for any $m \neq m^*$. The probability of m^* is ρ_{m^*}, therefore

$$
\begin{aligned}
d\left(\xi^{s}\right) &= \sum_{m\in\mathcal{M}}\left(\rho_{m}\left(\xi^{s}\right)-\rho_{m}\right)^{2} \\
&= \rho_{1}\left(1-\rho_{1}\right)^{2}+\left(1-\rho_{1}\right)\left(0-\rho_{1}\right)^{2}+ \\
&\quad ...+\rho_{M}\left(1-\rho_{M}\right)^{2}+\left(1-\rho_{M}\right)\left(0-\rho_{M}\right)^{2} \\
&= \rho_{1}\left(1-\rho_{1}\right)\left(1-\rho_{1}+\rho_{1}\right)+...+\rho_{M}\left(1-\rho_{M}\right)\left(1-\rho_{M}+\rho_{M}\right) \\
&= \sum_{m\in\mathcal{M}}\rho_{m}\left(1-\rho_{m}\right)=I
\end{aligned}
$$

Appendix C.1. Extensions to Continuous Marks

Throughout the paper I maintained the assumption that the marks were discrete, since I focused on the measurement of racial segregation. Here I show how to extend the basic definitions and results to continuous and multivariate segregation. Assume the researcher is interested in measuring income segregation.

The definition of extreme spatial separation slightly changes.

DEFINITION 5. *The process X is* ***completely unsegregated*** *if and only if $\rho(\xi, m) = \rho(m)$ for all $\xi \in X_0$, $m \in \mathcal{M}$. The process X is* ***completely segregated*** *if and only if for all $\xi \in x_0$, there is an $m^* = m^*(\xi) \in \mathcal{M}$ such that $\rho(\xi, m) = \delta(m - m^*)$, where $\delta(u)$ is the Dirac-Delta function.*

To measure the level of income segregation (or any nonnegative continuous variable) the mark space is assumed to be $\mathcal{M} = [0, \infty)$. The spatial dissimilarity index is derived analogously to the racial segregation case. Consider the quantity

$$d(\xi) = \int_{\mathcal{M}} |\rho(\xi, m) - \rho(m)|\, dm \qquad \text{(C.3)}$$

PROPOSITION 2. *If the mark space is $\mathcal{M} = [0, \infty)$ then under Complete Segregation*

$$d\left(\xi^{s}\right) = 2$$

Proof. Consider the quantity $\int_{\mathcal{M}} |\rho(\xi, m) - \rho(m)|\, dm$. For a given ξ and under complete segregation, $\exists m^* = m^*(\xi) \in \mathcal{M}$ such that $\rho(\xi, m) = \delta(m - m^*)$. The density associated with the realization of m^* is $\rho(m^*)$. Therefore we get

$$\int_{\mathcal{M}} |\rho(\xi, m) - \rho(m)|\, dm = \int_0^{\infty} \rho(m^*) \left[\int_0^{\infty} |\delta(m - m^*) - \rho(m)|\, dm \right] dm^*$$

We can solve the integral inside to get

$$\begin{aligned}
\int_0^{\infty} |\delta(m - m^*) - \rho(m)|\, dm &= \lim_{\varepsilon \to 0} \int_0^{m^* - \frac{\varepsilon}{2}} \rho(m)\, dm \\
&+ \lim_{\varepsilon \to 0} \int_{m^* - \frac{\varepsilon}{2}}^{m^* + \frac{\varepsilon}{2}} |\delta(m - m^*) - \rho(m)|\, dm + \\
&\quad + \lim_{\varepsilon \to 0} \int_{m^* + \frac{\varepsilon}{2}}^{\infty} \rho(m)\, dm \\
&= \lim_{\varepsilon \to 0} \int_0^{m^* - \frac{\varepsilon}{2}} \rho(m)\, dm + \lim_{\varepsilon \to 0} \int_{m^* - \frac{\varepsilon}{2}}^{m^* + \frac{\varepsilon}{2}} \delta(m - m^*)\, dm \\
&\quad - \lim_{\varepsilon \to 0} \int_{m^* - \frac{\varepsilon}{2}}^{m^* + \frac{\varepsilon}{2}} \rho(m)\, dm + \lim_{\varepsilon \to 0} \int_{m^* + \frac{\varepsilon}{2}}^{\infty} \rho(m)\, dm
\end{aligned}$$

By taking the limit for $\varepsilon \to 0$, using the fact that for Dirac-Delta

$$\lim_{\varepsilon \to 0} \int_{m^* - \frac{\varepsilon}{2}}^{m^* + \frac{\varepsilon}{2}} \delta(m - m^*)\, dm = 1$$

and

$$\lim_{\varepsilon \to 0} \int_{m^* - \frac{\varepsilon}{2}}^{m^* + \frac{\varepsilon}{2}} \rho(m)\, dm = 0$$

we can show that

$$\int_0^{\infty} |\delta(m - m^*) - \rho(m)|\, dm = 1 + \int_0^{m^*} \rho(m)\, dm + \int_{m^*}^{\infty} \rho(m)\, dm = 2$$

It follows that

$$\int_{\mathcal{M}} |\rho(\xi, m) - \rho(m)|\, dm = \int_0^{\infty} 2\rho(m^*)\, dm^* = 2$$

■

Therefore the individual Spatial Dissimilarity index for income segregation is defined as

$$\phi_{D\_Inc}(\xi) = \frac{1}{2} \int\limits_{\mathcal{M}} |\rho(\xi, m) - \rho(m)| \, dm$$

Appendix D. Artificial Cities

In Figure D.6, I show six artificial cities: A(symptotia), B(ayesia), C(lassica), D(eMoivria), E(mpirica) and F(isheria). Each city contains 800 individuals, distributed over the square $[0, 4] \times [0, 4]$. There are 25% blacks (the black circles) and 75% whites (the red circles). The grid represents the partition in neighborhoods.

For Cities A, B and C, I simulated an homogeneous Poisson Process with 50 points on a unit square, one for blacks and a different one for whites; I used the unit squares as neighborhoods of the cities, assigning 4 of them to be black and 12 of them to be white. City D was constructed by simulating white locations as an HPP with 600 points over the square $[0, 4] \times [0, 4]$. Then I simulated blacks locations as an HPP with 200 points in the circle of radius one, where the center of the circle coincided with the center of the city. City E was constructed by simulating an HPP with 600 points over the square $[0, 4] \times [0, 4]$ for the whites. Then I simulated two HPP with 100 points each over the circle of radius 1 for the black population. This creates an irregular black neighborhood in the city, while allowing whites to be inside the ghetto too. Finally, city F is the result of a simulation of an HPP with 600 points over the square $[0, 4] \times [0, 4]$ for the whites and an HPP with 200 points over the square $[0, 4] \times [0, 4]$ for the blacks. This is the perfect integrated case, according to our framework.

I report results for the spatial dissimilarity index estimation. In Table A1 I report the results of estimation for the artificial cities. The bandwidth is chosen using the Diggle and Berman (1989) procedure.

Figure D.6: Artificial Cities

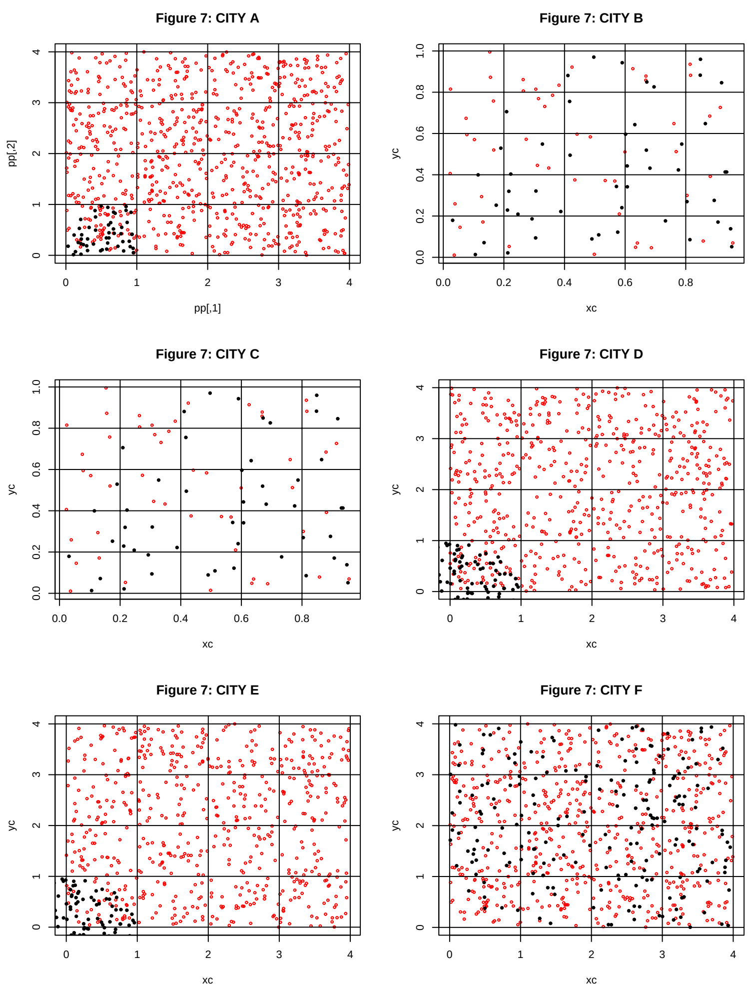


Table A1: Traditional vs Spatial Dissimilarity

	Bandwidth	Spatial Dism	Trad. Dism
City A	2.83	0.9225333	1
City B	2.605	0.900698	1
City C	0.37	0.9061751	1
City D	2.445	0.803017	0.7816667
City E	2.85	0.8993939	0.8816667
City F	2.73	0.03108531	0.1216667

For cities A, B and C the estimated spatial dissimilarity is smaller than the traditional, since the conditional probabilities surfaces make the estimate smoother. For cities D and E spatial and traditional index are very close. Of course if we change the neighborhoods definition this does not have to hold.[34] For the perfectly integrated city F, the spatial dissimilarity measures less segregation than the standard measure.

Appendix E. Spatial Indices of Segregation and Diversity

Appendix E.1. Spatial Dissimilarity Index

The spatial dissimilarity is constructed by using the absolute deviation as distance function between distributions

$$d(\xi) = \sum_{m \in \mathcal{M}} |\rho_m(\xi) - \rho_m| \tag{E.1}$$

Using the results of Proposition 1 we construct the global **Spatial Dissimilarity Index** is

$$\mathcal{T}_D(X) = \frac{1}{N(\mathcal{S})} \sum_{\xi \in X_0} \phi_D(\xi) \tag{E.2}$$

The main difference is that in the traditional dissimilarity the conditional probability $\rho_m(\xi)$ is assumed to be the same for all locations in the same neighborhood, while the spatial dissimilarity does not impose such within-neighborhood restriction on the spatial segregation.

Using the results in Theorem 2, one can derive the theoretical expected

[34] I computed the dissimilarity index for several different partitions of cities D and E: 4, 16, and 64 neighborhoods respectively.

For city E there is a clear increase of the index as we increase the number of neighborhoods. Surprisingly, for city D, the value of the index is not necessarily monotonically increasing in the number of neighborhoods: from 4 neighborhoods to 16 the index increases, while it decreases from 16 neighborhoods to 64.

This suggests another potential problem of the neighborhood-based approach: the relationship between the scale of the partition and the index is not necessarily monotonic.

These results are available from the author value of the index.

$$\mathbb{E}\left[\mathcal{T}_D\left(X\right)\right] = \left[2\Lambda_0\left(\mathcal{S}\right)\sum_{m\in\mathcal{M}}\rho_m\left(1-\rho_m\right)\right]^{-1}\int_{\mathcal{S}}\left[\sum_{m\in\mathcal{M}}\left|\rho_m\left(\xi\right)-\rho_m\right|\right]\lambda_0\left(\xi\right)d\xi \tag{E.3}$$

In most of the literature, the dissimilarity index is used to measure the segregation of a minority group from the rest of the population: this is the dichotomous version, in which the racial groups are assumed to be two, the minority and the rest of the population. In its *dichotomous* version, the spatial dissimilarity can be simplified, by using the fact that $\rho_{nb} = 1 - \rho_b$ (where b=blacks and nb=nonblacks), with $\phi_{Dic}\left(\xi\right) = \frac{|\rho_b(\xi)-\rho_b|}{2\rho_b(1-\rho_b)}$

$$\mathcal{T}_{Dic}\left(X\right) = \frac{1}{N\left(\mathcal{S}\right)}\sum_{\xi\in X_0}\phi_{Dic}\left(\xi\right)$$

Appendix E.2. Spatial Exposure Indices

The spatial exposure indices are derived using the squared deviation as distance function between spatial densities

$$d\left(\xi\right) = \sum_{m\in\mathcal{M}}\left[\rho_m\left(\xi\right)-\rho_m\right]^2 \tag{E.4}$$

The value of the index under perfect segregation is derived in proposition 1 The individual Spatial Exposure Index is defined as the location-specific squared deviation from perfect integration, normalized using (11).

$$\phi_{Exp}\left(\xi\right) = \frac{\sum_{m\in\mathcal{M}}\left[\rho_m\left(\xi\right)-\rho_m\right]^2}{\sum_{m\in\mathcal{M}}\rho_m\left(1-\rho_m\right)} \tag{E.5}$$

and the global **Spatial Exposure Index** is defined as

$$T_{Exp}\left(X\right) = \frac{1}{N\left(\mathcal{S}\right)}\sum_{\xi\in X_0}\phi_{Exp}\left(\xi\right) \tag{E.6}$$

An alternative approach to construct an exposure index is suggested in Reardon and Firerbaugh (2002). One can consider the dichotomous version of the index (E.5) for each group m, that is

$$\phi_{V,m}\left(\xi\right) = \frac{\left[\rho_m\left(\xi\right)-\rho_m\right]^2}{\rho_m\left(1-\rho_m\right)} \tag{E.7}$$

giving the dichotomous version of (E.6)

$$T_{V,m}(X) = \frac{1}{N(\mathcal{S})} \sum_{\xi \in X_0} \phi_{V,m}(\xi) \tag{E.8}$$

This index corresponds to a spatial version of Eta^2 (see White (1986) for a description) and it is a measure of how isolated a racial group is from the rest of the population. This is an index varying between 0 and 1, therefore a normalized index is constructed as the weighted sum of (E.8), where the weights are the ρ_m's. The **Spatial Normalized Exposure Index** is derived as

$$\begin{aligned} T_P(X) &= \sum_{m \in \mathcal{M}} \rho_m T_{V,m}(X) \\ &= \frac{1}{N(\mathcal{S})} \sum_{\xi \in X_0} \sum_{m \in \mathcal{M}} \frac{[\rho_m(\xi) - \rho_m]^2}{(1-\rho_m)} \end{aligned} \tag{E.9}$$

Notice that this is not equivalent to index (E.6).

Appendix E.3. Spatial Fractionalization Indices

Many studies relate ethnic and racial heterogeneity to economic outcomes.[35] The level of heterogeneity in these studies is usually measured with the Fractionalization Index. The latter measures the probability that two randomly drawn individuals belong to different racial groups. The index is derived from the Herfindhal index of homogeneity and it is equal to

$$I = 1 - \sum_{m \in \mathcal{M}} \rho_m^2 = \sum_{m \in \mathcal{M}} \rho_m (1 - \rho_m) \tag{E.10}$$

In the sociological literature the index is also known as the Simpson Interaction index. An index of zero indicates perfect homogeneity, in which only one racial group is present. Increasing values of the index imply increasing heterogeneity.

[35] Alesina, Baqir and Easterly (1999) show that more fractionalization is correlated with lower provision of local public goods. Easterly and Levine (1997) argue that more racially heterogenous societies show slower economic growth. Alesina and La Ferrara (2000) that participation in social activities is lower in more unequal and in more racially or ethnically heterogeneous localities. Mauro (1994) associates racial heterogeneity to more corruption.

In a recent contribution, D'Ambrosio, Bossaert and La Ferrara (2008) develop a more general version of the index in which the primitives are assumed to be individuals and their similarity. I follow a similar idea and develop a *spatial version* of the fractionalization index, in which the primitives of the aggregate index are the individual location-specific heterogeneity indices. The location-specific index is the level of fractionalization in location ξ

$$I(\xi) = \sum_{m \in \mathcal{M}} \rho_m(\xi)(1 - \rho_m(\xi))$$

and therefore the aggregate **Spatial Fractionalization Index** is

$$\mathcal{T}_I(X) = \frac{1}{N(\mathcal{S})} \sum_{\xi \in X_0} I(\xi) \tag{E.11}$$

This index measures the racial heterogeneity in the city incorporating the spatial location of individuals. Moreover the index can be disaggregated at the individual level, to examine the distribution of heterogeneity in the population. It can also be disaggregated over space showing which regions of the metropolitan area are more diverse.

An index of segregation can be derived from the spatial fractionalization using the distance

$$d(\xi) = |I(\xi) - I|$$

It is straightforward to show that under complete segregation $d(\xi^s) = I$: in each location there is maximum homogeneity therefore $I(\xi) = 0$ for any ξ. Define

$$\phi_F(\xi) = \frac{|I(\xi) - I|}{I}$$

to be the individual spatial relative fractionalization, which measures the absolute deviation from spatial homogeneity. The global **Spatial Relative Fractionalization Index** is

$$T_F(X) = \frac{1}{N(\mathcal{S})} \sum_{\xi \in X_0} \phi_F(\xi) \tag{E.12}$$

Appendix E.4. Spatial Entropy Indices

An alternative to the fractionalization indices is the Theil Entropy (or Information) Index (see Theil (1972) and Theil and Finezza (1971)). The entropy index for the metropolitan area is

$$E = \sum_{m \in \mathcal{M}} \rho_m \ln \left(\frac{1}{\rho_m} \right) \tag{E.13}$$

and it can be thought of as a measure of heterogeneity of the city since it is equal to zero if there is only one group and it reaches its maximum when all the groups have equal probability. I define a location-specific entropy index as

$$E(\xi) = \sum_{m \in \mathcal{M}} \rho_m(\xi) \ln \left(\frac{1}{\rho_m(\xi)} \right)$$

The **Spatial Entropy Index** is

$$T_E(X) = \frac{1}{N(\mathcal{S})} \sum_{\xi \in X_0} E(\xi) \tag{E.14}$$

This index measures the average racial heterogeneity in the city but incorporates the spatial location of each individual as a primitive. As for the fractionalization index it can be disaggregated at the individual and spatial level.

A simple index of segregation based on the spatial entropy can be constructed by defining a distance function

$$d(\xi) = |E(\xi) - E|$$

It is straightforward to show that under complete segregation $d(\xi^s) = E$: in fact complete segregation implies $E(\xi) = 0$ for all ξ. Define the individual location-specific spatial relative entropy as

$$\phi_H(\xi) = \frac{|E(\xi) - E|}{E}$$

This is the value of the absolute deviation from spatial homogeneity as measured by the entropy of the metropolitan area. The **Spatial Relative Entropy Index** formula is

$$T_H(X) = \frac{1}{N(\mathcal{S})} \sum_{\xi \in X_0} \phi_H(\xi) \tag{E.15}$$

and measures the average absolute deviation from spatial homogeneity.

Appendix F. Additional Tables

Table F.10: Correlations with traditional indices

	Spat. Dissim.	SSI	Dissim	Isol	Info
	Panel A: Blacks				
SSI	0.7044				
Dissimilarity	0.6675	0.5740			
Isolation	0.7371	0.9000	0.7810		
Information	0.7290	0.7926	0.9210	0.9545	
Gini	0.6749	0.5905	0.9897	0.7797	0.9180
	Panel B: Multigroup				
Dissimilarity	0.7484				
Isolation	0.7241		0.8821		
Information	0.7470		0.9530	0.9544	
Gini	0.7430		0.9860	0.8442	0.9402

The Spatial Dissimilarity is the average individual spatial dissimilarity. The SSI is the Spectral Segregation Index of Echenique and Fryer (2006). The Isolation, Information and Gini indices of segregation are described in Massey and Denton (1988) and Reardon and Firebaugh (2002). The spatial dissimilarity and the SSI are computed using block level data from the Summary File 1, Census 2000. The remaining indices are computed using Census Tracts data from the Census 2000. Correlations with indices computed using blocks are similar and available from the author.

Table F.11: Differences among traditional and spatial indices

	Dissim			Exposure		
Median hh income (log)	-0.195	-0.167	-0.144	-0.139	-0.113	-0.101
	(0.038)***	(0.042)***	(0.042)***	(0.048)***	(0.053)**	(0.055)*
Population (log)	-0.025	-0.021	-0.028	-0.015	-0.016	-0.012
	(0.005)***	(0.005)***	(0.008)***	(0.006)**	(0.007)**	(0.010)
Fraction of BA	0.169	0.125	0.101	0.022	-0.054	-0.076
	(0.081)**	(0.092)	(0.091)	(0.101)	(0.119)	(0.118)
Gini	-0.127	-0.062	0.122	0.299	0.318	0.509
	(0.272)	(0.279)	(0.285)	(0.342)	(0.359)	(0.368)
Fraction of Blacks	0.326	0.326	0.308	0.506	0.486	0.462
	(0.054)***	(0.055)***	(0.054)***	(0.067)***	(0.070)***	(0.070)***
Fraction of Asian/Pac. Isl	0.001	0.021	0.196	-0.055	-0.064	0.131
	(0.215)	(0.218)	(0.223)	(0.270)	(0.280)	(0.288)
Fraction of Other	0.190	0.222	0.209	-0.042	-0.094	-0.058
	(0.107)*	(0.115)*	(0.117)*	(0.135)	(0.148)	(0.150)
Manuf. Share		-0.136	-0.129		-0.116	-0.123
		(0.075)*	(0.075)*		(0.096)	(0.097)
Urban		-0.106	-0.082		-0.001	0.016
		(0.044)**	(0.045)*		(0.057)	(0.058)
Density$^{(a)}$			-0.013			-0.015
			(0.0045)***			(0.0058)***
Area$^{(a)}$			0.0002			-0.0045
			(0.0025)			(0.0032)
Number of Tracts$^{(a)}$			0.0588			0.0371
			(0.0344)*			(0.0443)
Constant	2.419	2.172	1.934	1.555	1.352	1.108
	(0.444)***	(0.470)***	(0.479)***	(0.558)***	(0.605)**	(0.618)*
Obs	308	293	293	308	293	293
R-squared	0.369	0.387	0.406	0.351	0.333	0.353

$^{(a)}$ coefficient and standard error multiplied by 10^3; * significant at 10; ** significant at 5; *** significant at 1.
Standard errors corrected for clustering at the MSA level in parentheses. The sample contains all 25-30 years old (Panel A) and 20-24 years old (Panel B) individuals born in US from the 1% PUMS 1990. Controls included but not shown: fraction of blacks in MSA, dummies for race (black, asian, hispanic and other nonwhite), dummy for female, age dummies, log of population in MSA, log of median income in MSA, manufacturing share of MSA. The last three variables are also included interacted with the black dummy.

Table F.12: Segregation and Outcomes: spatial vs non-spatial effects

	A. Individuals 25-30 years old			
	Dissimilarity		Exposure	
Spatial Segregation	-0.155	-0.168	-0.098	-0.106
	(0.028)***	(0.029)***	(0.017)***	(0.018)***
Spatial Segregation * black		0.055		0.074
		(0.07)		(0.069)
Trad. Segregation	0.088	0.127	0.047	0.077
	(0.030)***	(0.031)***	(0.023)**	(0.024)***
Trad. Segregation * black		-0.299		-0.237
		(0.064)***		(0.050)***
Observations	139634	139634	139634	139634
R-squared	0.038	0.038	0.038	0.038

	B. Individuals 20-24 years old			
	Dissimilarity		Exposure	
Spatial Segregation	-0.184	-0.209	-0.124	-0.139
	(0.036)***	(0.040)***	(0.024)***	(0.026)***
Spatial Segregation * black		0.127		0.146
		(0.088)		(0.083)*
Trad. Segregation	0.095	0.156	0.061	0.107
	(0.039)**	(0.041)***	(0.030)**	(0.031)***
Trad. Segregation * black		-0.416		-0.326
		(0.065)***		(0.047)***
Observations	97932	97932	97932	97932
R-squared	0.04	0.041	0.04	0.041

* significant at 10; ** significant at 5; *** significant at 1.
Standard errors corrected for clustering at the MSA level in parentheses. The sample contains all 25-30 years old (Panel A) and 20-24 years old (Panel B) individuals born in US from the 1% PUMS 1990. Controls included but not shown: fraction of blacks in MSA, dummies for race (black, asian, hispanic and other nonwhite), dummy for female, age dummies, log of population in MSA, log of median income in MSA, manufacturing share of MSA. The last three variables are also included interacted with the black dummy.