Poisson Indices of Segregation

Angelo, Mele

Department of Economics - University of Illinois at Urbana-Champaign

7 February 2009

Poisson Indices of Segregation*

Angelo Mele[†]

February 7, 2009

Abstract

Existing indices of residential segregation are based on an arbitrary partition of the city in neighborhoods: given a spatial distribution of racial groups, the index provides different levels of segregation for different partitions.

This paper proposes a method in which individual locations are mapped to aggregate levels of segregation, avoiding arbitrary partitions. Assuming a simple spatial process driving the locations of different racial groups, I define a location-specific segregation index and measure the city-level segregation as average of the individual index. After deriving several distributional results for this family of indices, I apply the idea to US Census data, using nonparametric estimation techniques. This approach provides different levels and rankings of cities' segregation than traditional indices. I show that high aggregate levels of spatial separation are the result of very few locations with extremely high local segregation.

I replicate the study of Cutler and Glaeser (1997) showing that their results change when segregation is measured using my approach. These findings potentially challenge the robustness of previous studies about the impact of segregation on socioeconomic outcomes.

JEL Classification: C14, C21, J15

Keywords: spatial segregation, spatial processes, nonparametric estimation

*I thank Roger Koenker, for generous advice and guidance. I am grateful to Andrea Asoni, Luigi Balletta, Alberto Bisin, Rosa Ferrer, Antonio Galvao, Darren Grant, Ron Laschever, Darren Lubotsky, Antonio Mele, Micah Pollak, Oleg Smirnov, Giorgio Topa, Antonella Tutino, Jungmo Yoon, Meyzam Zare and participants to the Econometrics Lunch at UIUC, MEA Conference 2008, SOLE Meeting 2008 and SEA Conference 2008 for helpful suggestions. All remaining errors are mine. *This paper won the Young Spatial Econometrician Award at the SEA Conference 2008 for the best theoretical paper in spatial econometrics.*

[†]Department of Economics, University of Illinois at Urbana-Champaign, 419 David Kinley Hall, 1407 W. Gregory, Urbana, IL 61801. Email: amele2@uiuc.edu. Web: http://netfiles.uiuc.edu/amele2/www/

1 Introduction

The spatial separation of racial groups in US metropolitan areas is a well documented fact, being the topic of an enormous 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 segregation in the metropolitan area.

However, all the existing indices of segregation are based on a partition of the city in neighborhoods, that makes the index directly dependent on the specific partition adopted. In particular, given a spatial distribution of racial groups, the index measures different segregation levels for alternative neighborhood definitions.[2] This mismeasurement problem raises concerns about the robustness of the estimated relationship between segregation and outcomes.

To overcome these issues, this paper proposes a method mapping individual locations to the level of aggregate segregation in a city and analyze how this affects the estimated correlation between racial segregation and socioeconomic outcomes. Assuming that the spatial distribution of socioeconomics characteristics is a realization of a spatial stochastic process that generates (exogenous) clustering by race, I define an individual location-specific index of segregation. The primitives of my index are the individual coordinates and their segregation levels. The metropolitan area segregation is measured as average of the individual indices.

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

[1] See for example Massey and Denton (1988 and 1993), Cutler and Glaeser (1997), Cutler, Glaser and Vigdor (1999), Ananat (2007), Echenique and Fryer (2007), Oscar and Volij (2008), Card and Rothstein (2007), Collins and Margo (2000), La Ferrara and Mele (2009), Ananat and Washington (2008).

[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.

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 method has several advantages with respect to the traditional neighborhood-based approach. First, the index does not depend on arbitrary partitions of the city in neighborhoods. I obtain the probability of location for each racial group for all possible locations in the metropolitan area, without relying on arbitrary neighborhood definitions. If the neighborhood definition changes over time my measure of segregation is unaffected.[3]

Second, this method provides the entire distribution of segregation among individuals and over space, allowing the researcher to identify which individuals or spatial regions are driving the spatial separation of groups. Indeed in the empirical section I argue that in many cases the synthetic index alone may be misleading: the estimated distributions are very skewed and very few extremely highly segregated individuals drive the average segregation, while most of the population experiences moderate levels of spatial separation.

Third, the estimation method relies on simple nonparametric techniques, available in standard statistical software. Therefore the computational burden is minimal and the time needed for estimation is reasonable.[4] In this paper the spatial distribution and the index are estimated using kernel estimation methods. In principle, as long as the researcher has access to a spatial random sample of individuals/locations for the metropolitan area, it is still possible to estimate the average segregation. This could possibly reduce the computational burden even further.

The paper describes several distributional properties of the spatial approach. After deriving the theoretical moments for *any possible* index of segregation, I restrict my attention to the family of *additive* indices, where each individual location contributes additively to the aggregate index. For each individual I define a location-specific index of segregation and measure the aggregate level of spatial separation as the average individual segregation. I characterize the expectation and variance for this family of indices.

Using alternative distance functions to measure the difference between spatial densities, I build several indices of diversity and segregation based on the spatial approach. The di-

[3]Most researcher define a neighborhood as a Census Tract. The US Census Bureau periodically revises the definitions of census tracts. Therefore the neighborhood partitions change over time, making comparability of the indices even more problematic.

[4]The only available individual level measure of segregation, the Spectral Segregation Index, is computationally very challenging for big cities. This is because the index is based on the network of each individual, requiring the computation of eigenvalues of an association matrix. This computation itself may require several hours for cities like New York. My index for New York can be computed in less than a minute.

versity indices are the Spatial Fractionalization Index and the Spatial Entropy Index, that measure the average population heterogeneity in the metropolitan area, taking into account the location of individuals and their local diversity. Several measures of segregation are also derived: a Spatial Dissimilarity, a Spatial Relative Fractionalization, a Spatial Relative Entropy, a Spatial Exposure and a Spatial Normalized Exposure. All these indices measure the segregation of the average individual in the metropolitan area, but differ in the specific distance function used as primitive. Other traditional indices of segregation can be reformulated in this framework.

The methodology is applied to the study of racial segregation in US metropolitan areas using Census 1990 and 2000 data. The estimate of the spatial distribution is obtained using standard nonparametric kernel estimation techniques for spatial point processes.[5]

I estimate *actual* segregation levels for all the metropolitan areas in the US using the average individual segregation. I compare the segregation levels measured by the spatial dissimilarity and the traditional dissimilarity. The levels of segregation and ranking of cities are very different when using my approach. For example, Muncie (IN) is the metropolitan area with highest segregation for African Americans according to the spatial dissimilarity, while according to the traditional dissimilarity is 141st. Correlations between the spatial dissimilarity index and the traditional indices are between 0.65 and 0.75. An analysis of individual segregation suggests that in several cities the high levels of spatial separation are driven by very few locations with extremely segregated individuals.

The differences between the two alternative approaches have significant economic implications. Using data from the 1% PUMS 1990 and Summary Tape File 1B of the 1990 Census, I replicate part of Cutler and Glaeser's (1997) study. They find that racial segregation undermines the socioeconomic performance of blacks in education, unemployment and earnings. Furthermore, segregation does not affect all the individuals, but mostly African Americans.

I compare results obtained using the Traditional Dissimilarity Index and the Spatial Dissimilarity Index,[6] using the same sample and variable definitions of the original work. My results confirm that racial segregation of African Americans is negatively related to blacks' individual socioeconomic outcomes.

However, I find that in the least squares estimates, *segregation is negatively correlated with the outcomes of all individuals*, not only blacks. By instrumenting racial segregation

[5] See Diggle (2003), Diggle, Zheng and Durr (2005) and Cressie (1993) for details.

[6] Echenique and Fryer (2007) replicate the ordinary least squares results of Cutler and Glaeser (1997) using the Spectral Segregation Index, confirming the original results.

with the number of local governments in 1962 and the transfer of federal revenues in 1962 as in the original paper, I find that *the magnitude of the coefficients is amplified*, implying an even stronger negative impact of segregation on socioeconomic outcomes.

These empirical findings suggest that the conclusions of previous studies may not be robust: when segregation is measured in a more precise way, i.e. taking into account the spatial location of each individual, the estimated correlation between segregation and outcomes may be different.

Finally, I show simple extensions of the methodology that can be used to measure segregation of continuous variables (e.g. income) or vectors. The definition of segregation slightly changes but the main theorems still hold. Furthermore, this approach is not confined to measuring residential segregation, but it can be applied in other fields of economics as well. For example, the spatial approach can be used to measure clustering of economic activities or spatial concentration of industries.[7]

2 Motivation and Related Literature

Residential separation by race (or other socioeconomic variables) is commonly observed in US metropolitan areas. The spatial separation has important economic implications: many studies show that there is a negative correlation between segregation and socioeconomic performance of minorities. Massey and Denton (1993) argue that residential segregation is responsible for the poor socioeconomic outcomes of blacks in US cities. Cutler and Glaeser (1997) is one of the most influential papers in the economics literature. They 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. Ananat (2007) provides similar results using a better instrumental variable technique for correcting the endogeneity of segregation. In particular she finds a mix of positive and negative effects on whites. Echenique and Fryer (2007) develop and use the Spectral Segregation Index to replicate the least squares regressions in Cutler and Glaeser (1997). They find that results are qualitatively the same as in the original paper, with slightly different point estimates. Collins and Margo (2000), suggest that the negative impact of residential segregation on African Americans outcomes is relatively recent, starting

[7]Several recent works follow the spatial approach. Arbia, Copetti and Diggle (2008) present methods similar to those used here for the analysis of spatial concentration of economic activity. Quah and Simpson (2003) build a model of spatial location of economic activity whose implication are empirically tested using techniques from spatial statistics.

from 1980. Card and Rothstein (2007) find that the black-white test score gap is higher in more segregated cities.[8]

In all these studies, the level of segregation of (say) blacks is measured with a synthetic index. The city is partitioned in K neighborhoods and for each neighborhood k we 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. If there is no segregation, the fraction of blacks in each neighborhood B_k/P_k should be equal to the fraction of blacks in the whole city, B/P. An index of segregation is then a synthetic measure of the *difference* between the actual distribution of races across neighborhoods, i.e. the distribution $(B_1/P_1, ..., B_K/P_K)$, and the distribution arising when there is no segregation, $(B/P, ..., B/P)$. The index is normalized to obtain a number from 0 to 1 that is comparable across cities. Different distance functions used by the researcher to measure this *difference* will lead to alternative indices.

To be concrete, consider the dissimilarity index, which is commonly used in empirical work. The distance between the distribution is computed using the absolute deviation $|B_k/P_k - B/P|$. The index is

$$D = \sum_{k=1}^{K} \frac{P_k}{P} \frac{|B_k/P_k - B/P|}{2\,(B/P)\,(1 - B/P)}$$

and it measures the proportion of blacks that should change neighborhood in order to achieve a perfectly integrated city.

We can also interpret the index as mean deviation from evenness. Define $\phi_k = \frac{|B_k/P_k - B/P|}{2(B/P)(1-B/P)}$ for each neighborhood k: this can be interpreted as the neighborhood-level segregation index. The global segregation D can be interpreted as average neighborhood segregation, weighting each neighborhood with the population proportion

$$D = \sum_{k=1}^{K} \frac{P_k}{P} \phi_k = \frac{1}{P} \sum_{k=1}^{K} P_k \phi_k \tag{1}$$

However, any index constructed according to the neighborhood-based approach presents some flaws, illustrated in Figure 1. The figure shows four stylized cities with the same spatial distribution of racial groups but a different partition in neighborhoods.

[Insert Figure 1 here]

[8] Recently Alesina and Zhuravskaya (2008) 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.

First, the index is based on an arbitrary partition of the metropolitan area in neighborhoods (as argued by Echenique and Fryer (2007)), usually census tracts or blocks, making the measurement directly dependent on the specific partition adopted. If segregation is measured using the standard dissimilarity then city A and C are perfectly segregated, city B is perfectly integrated and city C has an intermediate level of segregation. However, the spatial distribution of the racial groups is the same in the four cities: the difference in the measured segregation is just the outcome 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 (even worse) different ranking of the cities, a problem known in spatial analysis as Modifiable Area Unit Problem (MAUP). In Figure 1, the neighborhood partition in city A is obtained by partitioning each of the neighborhoods in city B in four sub-areas of same size. This results in a dissimilarity of 1 in city A, while in B segregation is 0.

Third, the majority of the indices does not take into account the spatial location of the individuals over the urban area, thus completely ignoring the within neighborhood spatial distribution. 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 relying on an arbitrary partition in neighborhoods, all these flaws do not apply. This is the main motivation of the present work.

To make the argument clear, let's assign to each individual i, $i = 1, .., n$, an individual index of segregation $\phi_i = \frac{|B_i/P_i - B/P|}{2(B/P)(1-B/P)}$, where B_i/P_i is the fraction of blacks in a small area around individual i.[9] The aggregate level of segregation is the average of individual segregation

$$D_{ind} = \frac{1}{P}\sum_{k=1}^{K}\sum_{i=1}^{P_k}\phi_i \tag{2}$$

By comparing (1) and (2) we notice that the traditional dissimilarity imposes a restriction on the individual level segregation, i.e.

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

In other words the traditional dissimilarity assumes no intra-neighborhood variation of

[9] I will be more precise about the definition of small area around the individual in the theoretical section.

spatial segregation. The approach presented here does not impose such a restriction and explicitly considers the spatial distribution of racial groups within neighborhoods.

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 (2008a and b) and Hutchens (2000) 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.

Echenique and Fryer (2007) is an exception in the axiomatic approach: they develop a segregation index based on individuals' social networks, satisfying three axioms. The Spectral Segregation Index measures segregation based on social interactions with same race neighbors, where neighbors are defined as agents living within 1 km.

I borrow several concepts and results from the literature on point processes.[10,11] In particular, this paper is related to Diggle, Zheng and Durr (2005), that 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

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

[11] Statistical models of point patterns are used in spatial epidemiology (Diggle, Zheng and Durr (2005), Kelsall and Diggle (1998)), Neuroscience (Diggle, Eglen and Troy (2006)), Astrophysics, Ecology, Geology (Zhuang, Ogata and Vere-Jones (2006)) and Image Recognition.

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.[12,13]

The use of spatial techniques in economics is very recent. Arbia, Copetti and Diggle (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 Theoretical Results

3.1 Notation, Basic Properties and Definitions

A spatial point process X is a stochastic process that maps points over a set $\mathcal{S} \subseteq \mathbb{R}^2$. Alternatively it can be defined as a random counting measure over bounded sets $A \subseteq \mathcal{S}$.[14,15] 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 a 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 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 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. The process is *simple* (or *orderly)* if there are no coincident points. In this paper I consider *simple nonstationary and anisotropic* processes.

[12] 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.

[13] 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.

[14] See Conley (1999) for a more technical explanation of point processes in the context of spatial GMM.

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

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[16] $\lambda : \mathcal{S} \rightarrow [0, \infty)$, defined as the limit of the expected number of points per infinitesimal area

$$\lambda(\xi) = \lim_{|d\xi| \rightarrow 0} \left\{ \frac{\mathbb{E}[N(d\xi)]}{|d\xi|} \right\} \tag{3}$$

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

$$\Lambda(A) = \mathbb{E}N(A) = \int_A \lambda(\xi)\, d\xi \tag{4}$$

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. Examples of marks are racial groups, income groups, income levels, education levels, or a mix of them.

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

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

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

therefore

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

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

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

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

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

This assumption provides a simple but flexible model for the spatial distribution of households in the urban area, which exhibits clustering. Notice that the clustering of locations in a certain region is *exogenous* and depends on the intensity only. In other words, this assumption imposes no behavioral or equilibrium restriction on how people choose their residential locations. In this context, I am not interested in studying the determinants of residential segregation: the important point is being able to estimate the spatial distribution of racial groups. The assumption of spatial Poisson locations allows this estimation in a simple way (as shown in the estimation section), while allowing the process to exhibit complex clustering properties of individual 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*

This implies that the presence of a specific attribute at a specific location does not influence the attributes at other locations. On the other hand, the assumption does not rule out clustering of marks.

Let $\rho\left(\xi, m, X_0 \diagdown \xi\right) \equiv \mathbb{P}\left(m\left(\xi\right) = m \middle| X_0\right)$ be the probability that an individual located in ξ has mark m, conditional on the realization of 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 $m \in \mathcal{M}$*

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

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.

Under these three assumptions it is possible to derive several distributional results, which I prove in Lemmas 1 and 2 in Appendix B. Lemma 1 characterizes the probability law of the process under the three assumptions. For a bounded region $A \subseteq \mathcal{S}$ and a configuration of points F it is possible to show that the probability law of the process is

$$\begin{aligned} &\mathbb{P}\left[(X \cap A) \in F\right] \qquad (5) \\ = &\sum_{n=0}^{\infty} \frac{\exp\left[-\Lambda_0(A)\right]}{n!} \int\limits_{A\times\mathcal{M}} \cdot\cdot \int\limits_{A\times\mathcal{M}} \mathbf{1}_{[\{(x_1,m_1),\ldots,(x_n,m_n)\}\in F]} \prod_{i=1}^{n} \left[\lambda\left(x_i, m_i\right)\right] dx_1 \cdot\cdot dx_n dm_1 \cdot\cdot dm_n \end{aligned}$$

To make exposition more concise I will focus on the case of *discrete marks*, which is the appropriate framework for racial segregation. In the last section of the paper I show how the definitions and theorems can be extended if marks are continuous or multivariate. Notice that both the main theorems are general and do not depend on the mark space being discrete.

Lemma 2 analyzes the stochastic process when the mark space is discrete: in this setting I use notation $\rho_m(\xi)$ to indicate the probability of mark m occurring at location ξ. The Lemma proves that the spatial process is equivalent to a multivariate Inhomogeneous Poisson process $X = \bigcup_{m=1}^{M} X_m$ with intensities $\lambda_m(\xi) = \lambda_0(\xi)\rho_m(\xi)$, $m = 1, 2, ..., M$ respectively, where the X_m's are stochastically independent.

The definition of segregated spatial distribution is operationalized using the conditional mark distributions. Intuitively, 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 labelling.* Therefore the marked poisson process is defined as completely unsegregated if there is random labelling of the events. 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 at the same location.[17]

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 labelling, i.e. $\rho_m(\xi) = \rho_m$ for all individuals $\xi \in X_0$, 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 $m \neq m^*$.*

[17] See Diggle, Zheng and Durr (2005) for a similar definition. The same idea is proposed in Arbia, Copetti and Diggle (2008).

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

$$\mathbb{E}[\mathcal{T}(X)] = \sum_{n=0}^{\infty} \frac{\exp[-\Lambda(\mathcal{S}\times\mathcal{M})]}{n!} \int_{\mathcal{S}\times\mathcal{M}} \cdot\cdot \int_{\mathcal{S}\times\mathcal{M}} \mathcal{T}(\{x_i, m_i\}_{i=1}^{n}) \prod_{i=1}^{n} \lambda(x_i, m_i)\, dx_1 \cdot\cdot dx_n dm_1 \cdot\cdot dm_n \tag{6}$$

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

$$\mathbb{E}[\mathcal{T}^r(X)] = \sum_{n=0}^{\infty} \frac{\exp[-\Lambda(\mathcal{S}\times\mathcal{M})]}{n!} \int_{\mathcal{S}\times\mathcal{M}} \cdot\cdot \int_{\mathcal{S}\times\mathcal{M}} \mathcal{T}^r(\{x_i, m_i\}_{i=1}^{n}) \prod_{i=1}^{n} \lambda(x_i, m_i)\, dx_1 \cdot\cdot dx_n dm_1 \cdot\cdot dm_n \tag{7}$$

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 (5). 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. ■

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 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. I assume that 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 $\phi(\xi)$*

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

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 the following theorem.

THEOREM 2 *Assume X follows a point process satisfying Assumptions 1-4. Then*

$$\mathbb{E}[\mathcal{T}(X)] = \mathbb{E}[\phi(\xi)] = \int_{\mathcal{S}} \phi(\xi) \frac{\lambda_0(\xi)}{\Lambda(\mathcal{S})} d\xi \tag{8}$$

$$\mathbb{V}[\mathcal{T}(X)] = \mathbb{E}\left[\frac{1}{N(\mathcal{S})}\right] \mathbb{V}[\phi(\xi)] \tag{9}$$

Proof. In appendix C ■

The results in Theorem 2 show that there is no difference between the expectation of global or individual level segregation. This follows from the independence assumption in the Poisson process. The variance of the global index is proportional to the variance of the individual level segregation. Moreover, this variance should be smaller for cities with higher population, or in terms of the poisson process, in cities with higher intensity measure.[18]

If we condition on the realized $N(\mathcal{S}) = n$, we obtain the following corollary

[18] To the extent that intensity measures for metropolitan areas are of the order of 100 thousands, we have that $\mathbb{E}\left[\frac{1}{N(\mathcal{S})}\right] \approx \frac{1}{\mathbb{E}N(\mathcal{S})}$ and therefore bigger cities will have smaller variances.

COROLLARY 1 *Under the assumptions of Theorem 2,*

$$\mathbb{E}\left[\mathcal{T}(X)\right| N(\mathcal{S}) = n] = \mathbb{E}\left[\phi(\xi)\right] \tag{10}$$

and

$$\mathbb{V}\left[\mathcal{T}(X)\right| N(\mathcal{S}) = n] = \frac{1}{n}\mathbb{V}\left[\phi(\xi)\right] \tag{11}$$

The work of Diggle, Zheng and Durr (2005) is based on the conditional specification of the spatial process.

I now provide several examples of indices of segregation. It is worth noting that most of the existing indices can be adapted to this approach by redefining the neighborhoods as individuals.

4 Spatial Indices of Segregation and Diversity

4.1 Spatial Dissimilarity Index

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

$$d(\xi) = \sum_{m \in \mathcal{M}} \left|\rho_m(\xi) - \rho_m\right| \tag{12}$$

In order to derive the distributional results, it is necessary to know the value of (12) under complete segregation. The following result applies to any index using a discrete set of marks. Let ξ^s be a generic point of a perfectly segregated process.

PROPOSITION 1 *If the mark space is discrete the value of (12) under* ***complete segregation*** *is*

$$d(\xi^s) = 2\sum_{m \in \mathcal{M}} \rho_m(1 - \rho_m) \tag{13}$$

Proof. *In Appendix C* ■

Incidentally notice that $d(\xi^s)$ is equivalent to twice the fractionalization of the city as defined below in (24). The individual-level segregation index is then measured by the function

$$\phi_D(\xi) = \frac{\sum_{m \in \mathcal{M}} \left|\rho_m(\xi) - \rho_m\right|}{2\sum_{m \in \mathcal{M}} \rho_m(1 - \rho_m)} \tag{14}$$

and the global **Spatial Dissimilarity Index** is

$$\mathcal{T}_D(X) = \frac{1}{N(\mathcal{S})} \sum_{\xi \in X_0} \phi_D(\xi) \tag{15}$$

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 value of the index.

$$\mathbb{E}[\mathcal{T}_D(X)] = \left[2\Lambda_0(\mathcal{S}) \sum_{m \in \mathcal{M}} \rho_m(1-\rho_m)\right]^{-1} \int_{\mathcal{S}} \left[\sum_{m \in \mathcal{M}} |\rho_m(\xi) - \rho_m|\right] \lambda_0(\xi)\, d\xi \tag{16}$$

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 $\phi_{Dic}(\xi) = \frac{|\rho_b(\xi) - \rho_b|}{2\rho_b(1-\rho_b)}$

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

4.2 Spatial Exposure Indices

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

$$d(\xi) = \sum_{m \in \mathcal{M}} [\rho_m(\xi) - \rho_m]^2 \tag{17}$$

The value of the index under perfect segregation is derived in the following proposition

PROPOSITION 2 *If the mark space is discrete the value of (17) under* ***complete segregation*** *is*

$$d(\xi^s) = \sum_{m \in \mathcal{M}} \rho_m(1-\rho_m) \tag{18}$$

Proof. *In Appendix C* ■

The individual Spatial Exposure Index is defined as the location-specific squared deviation from perfect integration, normalized using (18).

$$\phi_{Exp}(\xi) = \frac{\sum_{m\in\mathcal{M}} \left[\rho_m(\xi) - \rho_m\right]^2}{\sum_{m\in\mathcal{M}} \rho_m (1-\rho_m)} \tag{19}$$

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

$$T_{Exp}(X) = \frac{1}{N(\mathcal{S})} \sum_{\xi\in X_0} \phi_{Exp}(\xi) \tag{20}$$

An alternative approach to construct an exposure index is suggested in Reardon and Firerbaugh (2002). One can consider the dichotomous version of the index (19) for each group m, that is

$$\phi_{V,m}(\xi) = \frac{\left[\rho_m(\xi) - \rho_m\right]^2}{\rho_m (1-\rho_m)} \tag{21}$$

giving the dichotomous version of (20)

$$T_{V,m}(X) = \frac{1}{N(\mathcal{S})} \sum_{\xi\in X_0} \phi_{V,m}(\xi) \tag{22}$$

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 (22), where the weights are the ρ_m's. The **Spatial Normalized Exposure Index** is derived as

$$\begin{aligned} T_P(X) &= \sum_{m\in\mathcal{M}} \rho_m T_{V,m}(X) \\ &= \frac{1}{N(\mathcal{S})} \sum_{\xi\in X_0} \sum_{m\in\mathcal{M}} \frac{\left[\rho_m(\xi) - \rho_m\right]^2}{(1-\rho_m)} \end{aligned} \tag{23}$$

Notice that this is not equivalent to index (20).

4.3 Spatial Fractionalization Indices

Many studies relate ethnic and racial heterogeneity to economic outcomes.[19] 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

$$I = 1 - \sum_{m \in \mathcal{M}} \rho_m^2 = \sum_{m \in \mathcal{M}} \rho_m \left(1 - \rho_m\right) \tag{24}$$

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.

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

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

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

$$\mathcal{T}_I\left(X\right) = \frac{1}{N\left(\mathcal{S}\right)} \sum_{\xi \in X_0} I\left(\xi\right) \tag{25}$$

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

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

[19] 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.

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

$$T_F(X) = \frac{1}{N(\mathcal{S})} \sum_{\xi \in X_0} \phi_F(\xi) \tag{26}$$

4.4 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

$$E = \sum_{m \in \mathcal{M}} \rho_m \ln\left(\frac{1}{\rho_m}\right) \tag{27}$$

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

$$T_E(X) = \frac{1}{N(\mathcal{S})} \sum_{\xi \in X_0} E(\xi) \tag{28}$$

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

$$T_H(X) = \frac{1}{N(\mathcal{S})} \sum_{\xi \in X_0} \phi_H(\xi) \tag{29}$$

and measures the average absolute deviation from spatial homogeneity.

5 Empirical Methodology

All the data analysis was performed with **R**[20] by using some available packages for the analysis of spatial point patterns and by custom functions written by the author in **R** and C.[21]

5.1 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.

[Insert Figure 2 here]

I have data for all the 331 MSA's (Metropolitan Statistical Areas) and PMSA's (Primary Metropolitan Statistical Areas) for years 1990 and 2000. In order to maintain comparability

[20] http://www.r-project.org/

[21] In particular I used the packages **Splancs** and **SpatStat**. I also used a modified version of the package **spatialkernel** developed by Diggle, Zheng and Durr (2005). I created some additional C routines in order to compute the indices using the kernel regression approach explained below.

[22] I have an application pending at the Census Bureau in order to gain access to such data.

across census years, I adopt the racial categories in Census 1990: Whites/Caucasians, African Americans, Asian/Pacific Islanders, Native American, Other.

Figure 2(a) plots all blocks centroids locations in the New York PMSA for the 2000: the black dots represent blocks in which the majority is black while red dots are blocks in which the majority is nonblack. The pattern of geographic separation is clear: African Americans are concentrated in Harlem, Bronx and Bedford-Stuyvesant. Figure 2(b) 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]

5.2 Estimation Strategy with Exact Location Data

The estimation strategy consists of estimating the intensity function using nonparametric techniques. When individual location data are available there are standard methods used in spatial statistics to estimate the intensity of the process.[24]

Lemma 2 in Appendix B states that a multitype point process can be reformulated as a multivariate Poisson process with independent univariate processes, therefore one can estimate the intensities of each univariate process separately. This observation leads to a convenient estimate of $\widehat{\rho}_m(\xi)$

$$\widehat{\rho}_m(\xi) = \frac{\widehat{\lambda}_m(\xi)}{\widehat{\lambda}_0(\xi)} \tag{30}$$

where $\widehat{\lambda}_m(\xi)$ is the estimate of the intensity function for the univariate process X_m, corresponding to the spatial pattern of group m. Diggle (1985) and Berman and Diggle (1989) suggested a nonparametric estimator based on the definition of intensity function, $\widehat{\lambda}(\xi) = N(\xi, h)/\pi h^2$, 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.[25] 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 (2004))[26]

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

[24] See Diggle (2003), Diggle, Zheng and Durr (2005).

[25] This can be intepreted as a kernel estimator in which the kernel is

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

[26] There are alternative ways to estimate the conditional mark probability. For example, Diggle, Zheng

$$\widehat{\lambda}(\xi) = \sum_{i=1}^{n} \frac{\mathcal{K}_h(\xi - x_i)}{\int_{\mathcal{S}} \mathcal{K}_h(\xi - x_i)\, d\xi} \tag{31}$$

where $\mathcal{K}_h(u) = \frac{1}{h^2}\mathcal{K}(u/h)$. In my computations I will use a multiplicative quartic kernel in order to speed up the estimation procedure.[27]

It is known in the spatial statistics literature that the choice of the bandwidth is more important than the choice of the kernel function. The optimal h should be different for each city, since it should take into account the specific geographic density. The bandwidth 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 Squared Error (MSE) minimization procedure suggested in Diggle (1985) and Berman and Diggle (1989). The formula for the $MSE(h)$ is[28,29]

$$MSE(h) = \mu(0) + \Lambda(A)\frac{1 - 2K(h)}{\pi h^2} + \left(\pi h^2\right)^{-2} \int\int \mu(\|\xi - \eta\|)\, d\eta d\xi \tag{32}$$

where $\mu(\|\xi - \eta\|)$ is the *second-order intensity function* defined as

$$\mu(\xi, \eta) = \lim_{|d\xi|, |d\eta| \longrightarrow 0} \left\{ \frac{\mathbb{E}[N(d\eta)\, N(d\xi)]}{|d\eta|\, |d\xi|} \right\} \tag{33}$$

which is a measure of the spatial association of the process. 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. If we assume stationarity and isotropy then $\mu(\xi, \eta) = \mu(\|\xi - \eta\|)$, i.e it is a function of the euclidean distance among the two points. The quantity $K(h)$ is

---

and Durr (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).

[27] I have tried with a gaussian kernel, but the computational time is increased without differences in the estimated probabilities.

[28] A Cox Process is a point process such that:

1) $\left\{\Lambda(\xi) : \xi \in \mathbb{R}^2\right\}$ 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 $\lambda(\xi)$.

We can see an Inhomogeneous Poisson Point process as a particular Cox process in which the distribution of $\Lambda(\xi)$ is degenerate at $\lambda(\xi)$.

[29] This is a simple method of computing the optimal bandwidth. The literature on Point Processes usually relies on *ad hoc* criteria. Diggle, Zheng and Durr (2005) use cross-validated likelihood methods.

$$K(h) = \lambda^{-1} E[N_o(h)] = 2\pi\lambda^{-2} \int_0^h \mu(\xi)\, \xi d\xi \tag{34}$$

and it is defined as 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 $u_{ij} = \|x_i - x_j\|$.

$$\widehat{K}(h) = \frac{1}{n(n-1)} |\mathcal{S}| \sum_{i=1}^{n} \sum_{j \neq i} w_{ij}^{-1} I_h(u_{ij}) \tag{35}$$

where $I_h(u_{ij}) = I(u_{ij} \leq h)$ is an indicator function. This gives edge-corrected estimates of the $K(h)$ function. For the remaining part of (32), $\mu(0)$ does not depend on h, while for the integral we use the weighted integral suggested by Berman and Diggle (1989). By plugging these estimates in (32) we obtain an estimated $\widehat{MSE}(h)$.

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 show how the technique works using artificial data.

5.3 Estimation Strategy with Block Level Data

In many cases the exact location data are not available, thus I develop an approximated estimation technique to deal with data at the block level. I assume the researcher has the number of individuals of each racial group for each block and the location of the block centroid, as it 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 counting variables $N_0(\mathcal{S}_k)$ 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 $\overline{\xi}_k \in \mathcal{S}_k$ such that $\int_{\mathcal{S}_k} \lambda_0(\xi)\, d\xi = \lambda_0\left(\overline{\xi}_k\right) |\mathcal{S}_k|$ and thus

$$N_0(\mathcal{S}_k) = \lambda_0\left(\overline{\xi}_k\right) |\mathcal{S}_k| + u_k \tag{36}$$

Notice that $\overline{\xi}_k$ is not necessarily the centroid of the block. An approximation of (36) 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

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

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

$$\widehat{\lambda}_0(\xi)|S_k| = \sum_{k=1}^{K} \frac{\mathcal{K}_h(\xi - x_k)}{\sum_{j=1}^{K} \mathcal{K}_h(\xi - x_j)} n_{0k} \tag{37}$$

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 $\widehat{\rho}_m(\xi)$

$$\widehat{\rho}_m(\xi) = \frac{\widehat{\lambda}_m(\xi)}{\widehat{\lambda}_0(\xi)} = \frac{\sum_{k=1}^{K} \mathcal{K}_h(\xi - x_k)\, n_{mk}}{\sum_{k=1}^{K} \mathcal{K}_h(\xi - x_k)\, n_{0k}} \tag{38}$$

where n_{0k} is the number of people living in block k and n_{mk} is the number of people belonging to race m and living in block k; I use the estimated conditional probabilities evaluated at the block centroid to compute the index.

6 Empirical Results

6.1 Global Segregation In US Cities

I have estimated the Spatial Dissimilarity Index for all the racial groups and all the US metropolitan areas in 1990 and 2000. In this section I present results based on the 2000 data while in next section I use the indices for 1990 to estimate the impact of segregation on individual outcomes. In the tables I show only several metropolitan areas for ease of exposition.[30]

In Figure 3 I show the estimated conditional probability of African Americans in the New York PMSA. The bandwidth for the Kernel estimator obtained using the Berman and Diggle (1989) procedure is 0.348 km.

[Insert Figure 3 here]

[30] The complete tables in Excel files are available from the author.

The three main black areas in Bronx, Harlem and Bedford-Stuyvesant shown in Figure 2 above, correspond to the whiter areas in Figure 3, where the conditional probability is close or equal to 1.[31] The spatial dissimilarity of African Americans for New York is estimated to be 0.69.

In Figure 4(a) I plot the spatial dissimilarity and the neighborhood-based dissimilarity (computed using blocks) for *African Americans*. Figure 4(b) shows the same plot but the traditional dissimilarity is computed using census tract data. Each point represents a metropolitan area, indicated with the MSA FIPS code.

[Insert Figure 4 here]

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.

The segregation levels are shown in Table 1 for several cities. I compare the segregation levels obtained with the spatial dissimilarity and those obtained with the traditional approach, using blocks and census tracts, in column 3, 5 and 7. I also present the different ranking of the cities in columns 4, 6 and 8. Panel A and B are the ten most and least segregated MSAs respectively. Panel C shows the results for the most populated MSAs.

[Insert Table 1 here]

Not all the metropolitan areas show strikingly different levels of segregation when using the two approaches. For example, Detroit (MI) and Flint (MI) have comparable levels of segregation according to spatial and traditional dissimilarity. Muncie (IN) and Beaumont (TX) instead have dramatic differences in both level and rank. The least segregated city for African Americans is San Jose (CA). In Panel C of Table 1, I show the estimates for the most populated MSAs. The pattern seems confirmed: both levels and ranks are different.

[31] The reader should be aware that Figure 3 is realized with a grid 1000×1000. In the computation of the index I estimate the conditional probability only at the observed locations. This is more precise and computationally faster than imposing the grid.

[Insert Figure 5 here]

Figure 5(a) and (b) plot the multigroup Spatial Dissimilarity index against the traditional Dissimilarity. A similar pattern is present and not surprisingly the most segregated cities for African Americans are also the most segregated when considering all racial groups. This suggests that multigroup segregation levels are mainly driven by blacks' segregation.

[Insert Table 2 here]

This is confirmed in Table 2, where the multigroup spatial dissimilarity is compared to its neighborhood-based version. The most segregated MSA in US is Flagstaff (AZ-UT), with a level of 0.866741 while the lease segregated is Laredo (TX), with a spatial dissimilarity of 0.27716. The most populated cities display the same behavior of the previous table.

I computed the correlation among the spatial dissimilarity and several neighborhood-based indices in Table 3. I present correlations with the standard dissimilarity, the isolation index, the information index and the Gini index (see Massey and Denton (1988) or Reardon and Firebaugh (2002) for a detailed description). For blacks I also show the correlation with the Spectral Segregation Index of Echenique and Fryer (2007), which is the only index based on individuals locations available in the literature.

[Insert Table 3 here]

For the dichotomous version of the index (blacks) in Panel A, the correlation with the standard dissimilarity is 0.6675. Similarly the correlation with the Gini is 0.6749. Notice that Gini and Dissimilarity are almost perfectly correlated. The correlation with the Information and Isolation indices is slightly higher but still far from one.

The correlation with the Spectral Segregation Index (SSI) is similar, 0.7044. As explained in Echenique and Fryer (2007), the SSI is more correlated with the isolation index, since it is a more precise measure of exposure to same race neighbors. The correlations for the multigroup indices in Panel B are slightly higher but the pattern is similar. Rank correlations not shown here confirm the same pattern.[32]

6.2 Individual Segregation Distribution

One of the main advantages of using the spatial approach is that the researcher can examine the entire distribution of segregation. Figure 6 shows the distribution of individual-level

[32] Rank correlation results are available from the author.

segregation for African Americans, smoothed using a kernel estimate.

[Insert Figure 6 here]

The four distributions show that using the average to measure the level of segregation of a metropolitan area can be misleading. The red vertical line is the average segregation, which I use to measure segregation for the entire city. The green vertical line is the median segregation, while the blue lines are the 10th and 90th percentiles. Notice that in all four MSAs the average is above the median. The distribution is very skewed and there are very few blocks with extremely high levels of segregation that drive the average up.

[Insert Table 4 here]

I show the quartiles of the distributions for several MSAs in Table 4. For comparison I report the average segregation levels as in Table 1. Muncie (IN), which is the most segregated metropolitan area according to the average segregation, shows that this high average is due to very few blocks that are highly segregated, while most of the individuals are exposed to moderate levels of segregation. The same is valid for Detroit (MI), since most of the blocks have moderate levels of segregation. The average segregation is 0.87 while the 3rd quartile is about 0.65.

[Insert Table 5 here]

I repeat the same exercise for the multigroup version of the index, in Table 5. The distributional pattern is confirmed.

6.3 The Impact of Segregation on Socioeconomic Outcomes

The literature on the effect of segregation on socioeconomic outcomes usually shows a negative correlation of spatial separation and individual performance. In particular, Cutler and Glaeser (1997) is one of the most influential papers in economics. They regress measures of individual socioeconomic performance on the (traditional) dissimilarity and the interaction of dissimilarity and a dummy for African Americans, showing that racial segregation undermines the socioeconomic performance of African Americans in education, unemployment, earnings and single motherhood. They also find that whites are not affected significantly by segregation, even after controlling for the possible endogeneity of segregation. Ananat (2007) provides similar results using an alternative instrumental variable strategy. In particular she finds a mix of positive and negative effects on whites. Echenique and Fryer (2007) use the

Spectral Segregation Index to replicate the least squares regressions in Cutler and Glaeser (1997), finding qualitatively the same results, with slightly different point estimates.

I use the 1% PUMS 1990 and the Summary Tape File 1B to replicate the Cutler and Glaeser (1997) study. I analyze the same sample and the same specifications of the original paper, while substituting the traditional dissimilarity with the spatial dissimilarity. The samples contain all 20-24 years old and 25-30 years old individuals born in US. I consider only the MSAs for which the fiscal variables instruments are available.

The estimated linear probability model has the following specification

$$y_{ic} = \alpha + \beta Seg_c + \gamma Seg_c \times black_i + \delta X_i + \varepsilon_{ic} \tag{39}$$

where i indicates an individual and c a MSA/PMSA, y_{ic} is a socioeconomic outcome, $black_i$ is a dummy indicating if the individual is black, Seg_c is the segregation level of the MSA, and the controls X_i's are: 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.

The dependent variables are: the probability of high school graduation, the probability of college graduation, the probability of being idle (not in school nor at work) and the log of total earnings.

[Insert Table 6 here]

Table 6 compares the replication of Cutler and Glaeser's least squares estimates (upper panel) with the ones I obtained using the spatial dissimilarity (bottom panel), for the sample of 20-24 years old individuals.

There are several differences. When using the spatial dissimilarity, the segregation is harmful *per se*, decreasing the probability of high school graduation for all individuals and not only for blacks. The coefficients for college graduation are not significant in my analysis, implying that segregation is not correlated with this outcome. The results on idleness are similar, but I find that the log of total earnings is negatively correlated with segregation and strongly significant. These findings suggests that the conclusions of previous studies may be driven by the mismeasurement of segregation and may not be very robust.

To check the robustness of my estimates, I ran all the regressions using the Spatial Dissimilarity computed using a fixed bandwidth of .5 and 1 km. The point estimates are slightly different but the qualitative implications of Table 6 are confirmed. I also use the specification in Echenique and Fryer (2007), where variables are normalized, and the results do not change.

In order to correct for the endogeneity of segregation I use the fiscal instruments of Cutler and Glaeser. It is not clear if these are good instruments for the spatial dissimilarity and I leave the search of a good instrument for the spatial dissimilarity to future research. Nonetheless this exercise provides some insights.

The results of these regressions are reported in Table 7.

[Insert Table 7 here]

The results confirm the qualitative conclusions of the original paper. On the other hand the coefficients have higher magnitudes, implying a stronger negative impact of segregation on socioeconomic performance of blacks. Assuming that the instrumental variables are valid for the spatial dissimilarity, this suggests that segregation's negative impact may have been underestimated in previous studies.

The same pattern is confirmed by the 25-30 years old sample. The least squares estimates are in Table 8 and the instrumental variable estimates are in Table 9

[Insert Table 8 and 9 here]

Even if the causal interpretation of these results is unclear, the estimation exercise provides some evidence that when segregation is measured with more precision, i.e. taking into account the intra-neighborhood distribution of racial groups, the estimated impact on socioeconomic outcome is different.[33]

7 Extensions to Continuous and Multiple 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 3 *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*

[33] In results not shown I repeat the estimation using standardized variables as in Echenique and Fryer (2007). The qualitative results are unchanged. These results are available from the author.

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

$$d(\xi) = \int_{\mathcal{M}} |\rho(\xi, m) - \rho(m)| \, dm \tag{40}$$

PROPOSITION 3 *If the mark space is* $\mathcal{M} = [0, \infty)$ *then under Complete Segregation*

$$d(\xi^s) = 2$$

Proof. *In Appendix C* ■

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

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

and the average index for the metropolitan area is

$$\mathcal{T}_{D\_Inc}(X) = \frac{1}{2N(\mathcal{S})} \sum_{\xi \in X_0} \int_{\mathcal{M}} |\rho(\xi, m) - \rho(m)| \, dm \tag{41}$$

As emphasized in the theoretical section, Theorem 1 and 2 are general and do not depend on the mark space. The expectation computed using Theorem 2 is

$$\mathbb{E}\left[\mathcal{T}_{D\_Inc}(X)\right] = \frac{1}{2\Lambda_0(\mathcal{S})} \int_{\mathcal{S}} \int_{\mathcal{M}} |\rho(\xi, m) - \rho(m)| \, \lambda_0(\xi) \, dm d\xi \tag{42}$$

Furthermore, the researcher may be interested in computing segregation levels for more than one variable, for example residential segregation by race and income. This is easily done in this framework: define the mark as a vector $r = (m, y)$, where m is the racial group and y is the income level, and compute the joint conditional spatial probability of the marks $\rho(\xi, r)$. All the previous results apply.

Moreover the researcher can allow for correlation among marks of different type. In the case just mentioned, there is no restriction on the correlation between racial group and income levels. In other words, the mark vectors r must be independent, but there is no restriction about the joint distribution of m and y, i.e $\rho(\xi, m, y) \neq \rho_m(\xi) \rho(\xi, y)$

8 Conclusions

In this work I propose an alternative method for measuring spatial segregation of socioeconomic variables that considers individuals and their locations as primitives. Existing indices of segregation are based on an arbitrary partition of the metropolitan area in neighborhoods: given the same spatial distribution of racial groups, the index will measure different segregation levels for different neighborhood definitions.

The proposed method assumes that the locations of racial groups are the realization of a simple spatial process that generates a spatial density of racial groups characterized by (exogenous) clustering. For each coordinate of the metropolitan area, one can measure the probability that an individual living at that specific location belongs to a specific racial group. If there is no segregation the spatial density does not vary over space, it is flat. The segregation level of an individual is defined as difference between the actual spatial density and the flat spatial density *at her location*. The level of segregation of the entire metropolitan area is the segregation level of the average individual.

This method has several advantages with respect to the traditional neighborhood-based approach. First, the index does not depend on arbitrary partitions of the city in neighborhoods. Second, this method provides the entire distribution of segregation among individuals and over space, and therefore it is more informative than a synthetic index. I show that for most cities, the extremely high average level of segregation is driven by very few locations with excessive segregation, while most of the location are exposed to moderate levels of spatial separation. Third, the estimation method relies on simple nonparametric estimation techniques, available in standard statistical software. Furthermore, I derive several distributional properties of the indices derived under the spatial approach that could be used to develop rigorous statistical tests for segregation.

The methodology is not confined to indices of racial segregation, but it can be extended to measure segregation of continuous variables or vectors of variables as I show in the last section of the paper. Other applications include the study of clustering of economic activities or the spatial concentration of industries. These are topic of interest in the rapidly growing literature of economic geography, but there are very few applications of spatial statistics techniques in this field.[34]

This method delivers different segregation levels than the ones measured by traditional indices. Using 1990 and 2000 Census data I compute a spatial dissimilarity and compare it with the traditional dissimilarity. The resulting levels of segregation and ranking of cities in

[34] An exception is Arbia, Copetti and Diggle (2008).

terms of spatial separation are significantly different.

These differences have important economic implications. I replicate Cutler and Glaeser (1997) analysis of the impact of segregation on socioeconomic outcomes, showing that results change when the traditional dissimilarity is replaced by the spatial dissimilarity. I conclude that my empirical findings may potentially challenge the robustness of the estimated impact of racial segregation on individual outcomes: when segregation is measured in a more precise way, i.e. taking into account the spatial location of each individual, the conclusions of previous studies may be different.

References

[1] Alesina, Alberto, Baqir, Reza and Easterly, William (1999), "Public Goods and Ethnic Divisions", The Quarterly Journal of Economics 114(4):1243-1284

[2] Alesina Alberto and La Ferrara, Eliana (2000), "Participation In Heterogeneous Communities," The Quarterly Journal of Economics 115(3):847-904

[3] Alesina, Alberto and Zhuravskaya, Ekaterina (2008), "Segregation and the Quality of Government in a Cross-Section of Countries", mimeo, Harvard University

[4] Ananat Oltmans, Elizabeth (2007), "The Wrong Side(s) of the Tracks: The Effect of Racial Segregation on City Outcomes", NBER WP 13343

[5] Ananat Oltmans, Elizabeth and Washington, Ebonya (2008), "Segregation and Black Political Efficacy", NBER Working Paper No. 13606.

[6] Arbia, Giuseppe, Copetti, Massimiliano and Diggle, Peter (2008), "Modelling individual behaviour of firms in the study of spatial concentration", in Fratesi U and Senn L (eds.) "Growth in Interconnected Territories", Springer-Verlag (2008)

[7] Baddeley, Adrian and Turner, Rolf (2005), "spatstat: An R Package for Analyzing Spatial Point Patterns", Journal of Statistical Software, 12(6):1-42

[8] Berman, Mark and Diggle, Peter (1989), "Estimating Weighted Integrals of the Second-Order Intensity of a Spatial Point Process", Journal of the Royal Statistical Society, Series B, 51(1):81-92

[9] Card, David and Rothstein, Jesse (2007), "Racial Segregation and the Black-White Test Score Gap", Journal of Public Economics 91(11-12):2158-2184.

[10] Collins, William J. and Margo, Robert A. (2000), "Residential segregation and socioeconomic outcomes: When did ghettos go bad?", Economic Letters 69:239-243

[11] Conley, Timothy G. (1999), "GMM Estimation with Cross Sectional Dependence" Journal of Econometrics 92(1):1-45.

[12] Cressie, Noel A.C. (1993), "Statistics for Spatial Data", John Wiley and Sons, New York

[13] Cutler , D. M. and Glaeser, E. L. (1997), "Are Ghettos Good or Bad", Quarterly Journal of Economics, 112: 827-872

[14] Cutler, D. M., Glaeser, E.L. and Vigdor, Jacob L. (1999), The Rise and Decline of the American Ghetto, Journal of Political Economy, 107(3):455-506

[15] Daley, D. J. and Vere-Jones, D (2003), "An Introduction to the Theory of Point Processes", Springer, 2nd Edition

[16] D'Ambrosio, Conchita, Bossaert, Walter and La Ferrara, Eliana (2008), "A Generalized Index of Ethno-Linguistic Fractionalization", mimeo, Bocconi University and IGIER

[17] Diggle, Peter (1983), "Statistical Analysis of Spatial Point Patterns", Academic Press, London, First Edition

[18] Diggle, Peter (1985), "A Kernel Method for Smoothing Point Process Data", Applied Statistics, 34(2):138-147

[19] Diggle, Peter (2003), "Statistical Analysis of Spatial Point Patterns", Academic Press, London, Second Edition

[20] Diggle, Peter, Zheng, Pingping and Durr, Peter (2005), "Nonparametric estimation of spatial segregation in a multivariate point process: bovine tubercolosis in Cornwall, UK", Applied Statistics, 54(3):645-658

[21] Diggle P. J, Eglen, S. J. and Troy, J. B. (2006). "Modelling the Bivariate Spatial Distribution of Amacrine Cells", In A. Baddeley et al. (Eds.) Case Studies in Spatial Point Process Modelling, Springer Lecture notes in Statistics 185:215-233

[22] Easterly W. and R. Levine (1997), "Africa's Growth Tragedy: Policies and Ethnic Divisions", Quarterly Journal of Economics, 111(4), 1203-1250.

[23] Echenique, Federico and Fryer, Roland G. Jr. (2007), "A Measure of Segregation Based on Social Interactions", Quarterly Journal of Economics 122(2):441-485

[24] Frankel, David and Volij, Oscar (2008a), "Measuring School Segregation", mimeo

[25] Frankel, David and Volij, Oscar (2008b), "Scale Invariant Measures of Segregation", mimeo

[26] Glaeser, E. L. and Vigdor, Jacob L. (2000), "Racial Segregation in 2000 Census: Promising News", Center of Urban and Metropolitan Policy, The Brookings Institution Survey Series, April

[27] Hutchens, Robert (2004), "One Measure of Segregation", International Economic Review 45(2):555-578

[28] Kelsall, Julia E. and Diggle, Peter J. (1998), "Spatial Variation in Risk of Diseases: A Nonparametric Binary Regression Approach", Applied Statistics, 47(4):559-573

[29] Koenker, Roger and Mizera, Ivan (2004), "Penalized Triograms: Total Variation Regularization for Bivariate Smoothing", Journal of Royal Statistical Society: Series B (Statistical Methodology) 66(1):145-163

[30] La Ferrara, Eliana and Mele, Angelo (2009), "Racial Segregation and Public School Expenditure", CEPR WP 5750

[31] Massey, Douglas S. and Denton, Nancy A. (1988), "The Dimensions of Residential Segregation", Social Forces, 67(2):281-315

[32] Massey, Douglas S. and Denton, Nancy A. (1993), "American Apartheid: Segregation and the Making of the Underclass", Cambridge: Harvard University Press.

[33] Mauro, Paolo (1995), “Corruption and Growth”, Quarterly Journal of Economics,110:681-712.

[34] Moller, Jesper and Waagepetersen, Rasmus Plenge (2004), "Statistical Inference and Simulation for Spatial Point Processes", Monographs on Statistics and Applied Probability 100, Chapman and Hall

[35] Quah, Danny and Simpson, Helen (2003), "Spatial Cluster Empirics", Working Paper, London School of Economics

[36] Reardon, Sean F. and Firebaugh, Glenn (2002), "Measures of Multigroup Segregation", Sociological Methodology 32:33-68

[37] Reardon, Sean F. and O'Sullivan, David (2004), "Measures of Spatial Segregation", Sociological Methodology 34:

[38] Reardon, Sean F, O'Sullivan, David, Lee, Barrett A., Firebaugh, Glenn, Farrell, Chad (2006), "The Segregation Profile: Investigating How Metropolitan Racial Segregation Varies by Spatial Scale", WP 06-01, Stanford University

[39] Rowlingson, B.S. and Diggle P.J. (1993), "Splancs: Spatial Point Patterns Analysis Code in S-Plus", Computers and Geosciences, 19:627-655

[40] Stoyan, D., Kendall, W.S. and Mecke, J. (1987), "Stochastic Geometry and Its Applications", John Wiley and Sons

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

[42] Theil, Henri (1972). "Statistical Decomposition Analysis". Amsterdam: North-Holland.

[43] Theil, Henri, and Finezza, Anthony J. (1971). "A Note on the Measurement of Racial Integration of Schools by Means of Informational Concepts." Journal of Mathematical Sociology 1:187-94.

[44] Zhuang J., Ogata Y. and Vere-Jones D. (2005), "Diagnostic Analysis of Space-Time Branching Processes for Earthquakes" Chap. 15 (Pages 275-290) of Case Studies in Spatial Point Process Models, Edited by Baddeley A., Gregori P., Mateu J., Stoica R. and Stoyan D. Springer-Verlag, New York.

A Background Theory

In this section I briefly review the fundamental concepts and definitions needed to develop my main theoretical results.[35] 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.

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.

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[36] $\lambda : \mathcal{S} \to [0, \infty)$, defined as the limit of the expected number of points per infinitesimal area

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

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

$$\lambda(\xi) = \lim_{|d\xi| \to 0} \left\{ \frac{\mathbb{E}[N(d\xi)]}{|d\xi|} \right\} \tag{43}$$

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

$$\Lambda(A) = \mathbb{E}N(A) = \int_A \lambda(\xi)\, d\xi \tag{44}$$

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.[37]

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 4 ***(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 $\Lambda(A) < \infty$*

$$\mathbb{P}[N(A) = n] = [\Lambda(A)]^n \frac{\exp[-\Lambda(A)]}{n!}, \qquad n = 0, 1, 2, ... \tag{45}$$

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*

$$f(\xi) = \frac{\lambda(\xi)}{\int_A \lambda(\xi)\, d\xi} \tag{46}$$

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

[37] 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 $d\eta$.

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 (46) 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.

$$\mathbb{P}\left[N(A_1) = n_1, ..., N(A_K) = n_K\right] = \prod_{k=1}^{K} \left[\Lambda(A_k)\right]^{n_k} \frac{\exp\left[-\Lambda(A_k)\right]}{n_k!} \tag{47}$$

for $n = n_1 + n_2 + ... + n_k$. In Appendix B, I prove that conditions (45) and (46) imply (47).

In this paper I consider only Inhomogeneous Poisson Point Processes (IPP): these processes are nonstationary and anisotropic, with spatially varying intensity function.[38] The IPP is a very simple and parsimonious model 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[39]

$$\mathbb{P}\left[(X \cap A) \in F\right] = \sum_{n=0}^{\infty} \frac{\exp\left[-\Lambda(A)\right]}{n!} \int_A \cdot\cdot \int_A \mathbf{1}_{[\{x_1,...,x_n\} \in F]} \prod_{i=1}^{n} \lambda(x_i)\, dx_1 \cdot\cdot dx_n \tag{48}$$

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.

[38] 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 (45) 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 (46) 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).

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

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 5 ***(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.

B Point Processes Theory

B.1 Independent Scattering Property

PROPOSITION 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 $N(A_1), N(A_2), N(A_3), ...$ are stochastically independent, i.e.

$$\mathbb{P}[N(A_1) = n_1, ..., N(A_k) = n_k] = \prod_{j=1}^{k} [\Lambda(A_j)]^{n_j} \frac{\exp[-\Lambda(A_j)]}{n_j!} \tag{49}$$

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 $N(A) = n_1 + n_2 = n$, $\mathbb{P}[\xi \in (X \cap A)] = f(\xi) = \lambda(\xi) / \Lambda(A)$. Then given $N(A) = n$,

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

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

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

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

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

■

B.2 Probability Law of a Poisson Point Process

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

$$\mathbb{P}\left[\left(X \cap A\right) \in F\right]=\sum_{n=0}^{\infty}\frac{\exp\left[-\Lambda\left(A\right)\right]}{n!}\int_A \cdot\cdot \int_A \mathbf{1}_{\left[\left\{x_1,\ldots,x_n\right\}\in F\right]}\prod_{i=1}^{n}\lambda\left(x_i\right)dx_1\cdot\cdot dx_n \tag{50}$$

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

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

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

In order to get the unconditional probability we just need to multiply by $\mathbb{P}\left[N\left(A\right)=n\right]=\frac{\exp[-\Lambda(A)]}{n!}\Lambda\left(A\right)^n$ and sum for all n, obtaining (50).

For the necessary part of the proof just multiply (50) inside the sum by $\frac{\Lambda(A)^n}{\Lambda(A)^n}$ and notice you can rewrite the probability as

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

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

The probability law of the process over $\mathcal{S}\subseteq\mathbb{R}^2$ is obtained from (50), by substituting A with $\mathcal{S}$.

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

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

Therefore the unconditional distribution is

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

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

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

It follows from (50) that $X \sim Poi(T, \lambda(t))$ or $X \sim Poi(\mathcal{S}\times\mathcal{M}, \lambda(\xi,m))$ ■

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

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

Using the same argument we can show that

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

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

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

Using the independent scattering property of the Poisson Process, for $A, B \subseteq \mathcal{S}$

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

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 $\left(\bigcup_{m=1}^{M} X_m\right) \sim Poi\left(\mathcal{S}, \sum_{m=1}^{M} \lambda_m(\xi)\right) = Poi\left(\mathcal{S}, \lambda_0(\xi)\right) = X_0$. 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.*

C Proofs of the Main Results

PROOF OF THEOREM 2

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

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

It follows that

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

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

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

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

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

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

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

PROOF OF PROPOSITION 2

This proof follows the same lines of the proof for Proposition 1. Consider the quantity $\sum_{m\in\mathcal{M}}\left(\rho_m(\xi)-\rho_m\right)^2$. 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

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

PROOF OF PROPOSITION 3

Consider the quantity $\int_{\mathcal{M}}\left|\rho(\xi,m)-\rho(m)\right|dm$. 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

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

We can solve the integral inside to get

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

By taking the limit for $\varepsilon\to 0$, using the fact that for Dirac-Delta $\lim_{\varepsilon\to 0}\int_{m^*-\frac{\varepsilon}{2}}^{m^*+\frac{\varepsilon}{2}}\delta\left(m-m^*\right)dm = 1$ and $\lim_{\varepsilon\to 0}\int_{m^*-\frac{\varepsilon}{2}}^{m^*+\frac{\varepsilon}{2}}\rho\left(m\right)dm = 0$

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

It follows that

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

D Artificial Cities

In Figure A1, 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.

[Insert Figure A1 here]

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.

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.[40] For the perfectly integrated city F, the spatial dissimilarity measures less segregation than the standard measure.

[40] 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

Table 1: 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 MSA in US						
5280	Muncie, IN MSA	0.878505	*1*	0.7022	*141*	0.5282	*150*
2960	Gary, IN PMSA	0.874766	*2*	0.8602	*4*	0.8093	*2*
2160	Detroit, MI PMSA	0.870148	*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	0.84719	*5*	0.8123	*19*	0.7334	*12*
1320	Canton--Massillon, OH MSA	0.845705	*6*	0.738	*89*	0.5774	*99*
2640	Flint, MI PMSA	0.841102	*7*	0.8268	*11*	0.7646	*6*
1000	Birmingham, AL MSA	0.838985	*8*	0.8157	*17*	0.6989	*20*
840	Beaumont--Port Arthur, TX MSA	0.827306	*9*	0.7513	*74*	0.6481	*47*
5200	Monroe, LA MSA	0.826333	*10*	0.8082	*22*	0.69	*27*
	B. Least segregated MSAs in US						
6560	Pueblo, CO MSA	0.41685	*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	0.414883	*324*	0.5457	*305*	0.3695	*286*
1125	Boulder--Longmont, CO PMSA	0.410511	*325*	0.6155	*261*	0.3239	*311*
7480	Santa Barbara--Santa Maria--Lompoc, CA MSA	0.409521	*326*	0.5629	*295*	0.3894	*271*
5170	Modesto, CA MSA	0.394449	*327*	0.572	*291*	0.3212	*313*
200	Albuquerque, NM MSA	0.379495	*328*	0.5505	*303*	0.312	*319*
380	Anchorage, AK MSA	0.372978	*329*	0.4489	*328*	0.3336	*308*
5945	Orange County, CA PMSA	0.36862	*330*	0.5072	*318*	0.3391	*305*
7400	San Jose, CA PMSA	0.325668	*331*	0.4817	*323*	0.2939	*325*
	C. Most populated MSAs in US						
4480	Los Angeles--Long Beach, CA PMSA	0.614858	*177*	0.6266	*252*	0.5765	*102*
5600	New York, NY PMSA	0.690352	*97*	0.7013	*142*	0.6714	*38*
1600	Chicago, IL PMSA	0.763236	*35*	0.8215	*15*	0.7789	*4*
6160	Philadelphia, PA--NJ PMSA	0.727624	*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	0.870148	*3*	0.8655	*3*	0.8405	*1*
3360	Houston, TX PMSA	0.705639	*81*	0.6578	*210*	0.5716	*106*
520	Atlanta, GA MSA	0.675998	*115*	0.6949	*157*	0.6148	*66*
1920	Dallas, TX PMSA	0.636549	*156*	0.628	*250*	0.5396	*133*
1120	Boston, MA--NH PMSA	0.60094	*191*	0.7084	*132*	0.6364	*54*

Notes: 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.

Table 2: Spatial Dissimilarity vs Traditional Dissimilarity (Multigroup)

MSA FIPS	Metropolitan Area	Spatial Dissimilarity Levels	Spatial Dissimilarity Rank	Dissimilarity (Blocks) Levels	Dissimilarity (Blocks) Rank	Dissimilarity (Tracts) Levels	Dissimilarity (Tracts) Rank
	A. Most segregated MSAs in US						
2620	Flagstaff, AZ--UT MSA	0.866741	*1*	0.7093	*69*	0.5808	*38*
2160	Detroit, MI PMSA	0.828644	*2*	0.8198	*2*	0.7355	*1*
8080	Steubenville--Weirton, OH--WV MSA	0.821351	*3*	0.7397	*38*	0.5177	*86*
1000	Birmingham, AL MSA	0.818724	*4*	0.8029	*5*	0.6661	*8*
5200	Monroe, LA MSA	0.816387	*5*	0.8033	*4*	0.669	*7*
5280	Muncie, IN MSA	0.815499	*6*	0.6843	*105*	0.4757	*120*
1320	Canton--Massillon, OH MSA	0.807292	*7*	0.7204	*55*	0.5089	*96*
2640	Flint, MI PMSA	0.796078	*8*	0.799	*6*	0.6747	*6*
840	Beaumont--Port Arthur, TX MSA	0.79018	*9*	0.738	*41*	0.6101	*24*
760	Baton Rouge, LA MSA	0.790143	*10*	0.762	*21*	0.6113	*22*
	B. Least segregated MSAs in US						
1150	Bremerton, WA PMSA	0.399612	*322*	0.437	*322*	0.2669	*303*
6560	Pueblo, CO MSA	0.399277	*323*	0.4754	*306*	0.2864	*293*
4150	Lawrence, KS MSA	0.398295	*324*	0.4753	*307*	0.264	*306*
1720	Colorado Springs, CO MSA	0.396948	*325*	0.4575	*313*	0.3069	*280*
5170	Modesto, CA MSA	0.394621	*326*	0.4457	*317*	0.2684	*301*
1880	Corpus Christi, TX MSA	0.394152	*327*	0.4337	*323*	0.2515	*311*
7840	Spokane, WA MSA	0.386892	*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	0.279575	*330*	0.367	*330*	0.2017	*327*
4080	Laredo, TX MSA	0.27716	*331*	0.3563	*331*	0.1072	*331*
	C. Most populated MSAs in US						
4480	Los Angeles--Long Beach, CA PMSA	0.4834	*270*	0.4973	*289*	0.4091	*183*
5600	New York, NY PMSA	0.605364	*138*	0.6286	*183*	0.5603	*56*
1600	Chicago, IL PMSA	0.656347	*90*	0.7057	*76*	0.6141	*21*
6160	Philadelphia, PA--NJ PMSA	0.696579	*54*	0.7306	*45*	0.6252	*16*
8840	Washington, DC--MD--VA--WV PMSA	0.58943	*149*	0.5949	*212*	0.5028	*100*
2160	Detroit, MI PMSA	0.828644	*2*	0.8198	*2*	0.7355	*1*
3360	Houston, TX PMSA	0.569899	*175*	0.5689	*237*	0.4548	*138*
520	Atlanta, GA MSA	0.637662	*108*	0.6702	*126*	0.5603	*55*
1920	Dallas, TX PMSA	0.558697	*188*	0.5718	*235*	0.4478	*144*
1120	Boston, MA--NH PMSA	0.533627	*220*	0.6435	*166*	0.5215	*80*

Notes: 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.

Table 3: Correlations with traditional indices					
Panel A: Blacks					
	Spat. Dissim.	SSI	Dissim	Isol	Info
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

Notes: 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 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	0.878505	0.4210913	0.536438	0.536438
2960	Gary, IN PMSA	0.874766	0.59432	0.624964	0.624964
2160	Detroit, MI PMSA	0.870148	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	0.84719	0.4719803	0.5582634	0.5582634
1320	Canton--Massillon, OH MSA	0.845705	0.4002332	0.5365979	0.5365979
2640	Flint, MI PMSA	0.841102	0.531328	0.6290947	0.6315051
1000	Birmingham, AL MSA	0.838985	0.6466866	0.7176122	1.0913744
840	Beaumont--Port Arthur, TX MSA	0.827306	0.5746312	0.6678815	1.0368019
5200	Monroe, LA MSA	0.826333	0.6165972	0.755943	1.2140182
	B. Least Segregated MSAs in US				
6560	Pueblo, CO MSA	0.41685	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	0.414883	0.2094268	0.3707898	0.5103415
1125	Boulder--Longmont, CO PMSA	0.410511	0.1856612	0.363827	0.504531
7480	Santa Barbara--Santa Maria--Lompoc, CA MSA	0.409521	0.2085673	0.3972317	0.5123268
5170	Modesto, CA MSA	0.394449	0.2065722	0.3786059	0.5140098
200	Albuquerque, NM MSA	0.379495	0.1716114	0.3434244	0.5133553
380	Anchorage, AK MSA	0.372978	0.2025856	0.358544	0.5210976
5945	Orange County, CA PMSA	0.36862	0.1853323	0.3445561	0.4796011
7400	San Jose, CA PMSA	0.325668	0.1519909	0.3096957	0.4531364
	C. Most Populated MSAs in US				
4480	Los Angeles--Long Beach, CA PMSA	0.614858	0.343943	0.4747065	0.5480155
5600	New York, NY PMSA	0.690352	0.5258307	0.6423153	0.6737719
1600	Chicago, IL PMSA	0.763236	0.5528326	0.618727	0.6195102
6160	Philadelphia, PA--NJ PMSA	0.727624	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	0.870148	0.6527996	0.6527996	0.6527996
3360	Houston, TX PMSA	0.705639	0.4178659	0.55851	0.6096372
520	Atlanta, GA MSA	0.675998	0.4436585	0.6349907	0.7082935
1920	Dallas, TX PMSA	0.636549	0.3680504	0.5148539	0.5912919
1120	Boston, MA--NH PMSA	0.60094	0.3859964	0.504475	0.5383391

Notes: 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.

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

Notes: 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.

Table 6: Individuals 20-24 years old, OLS results

	Cutler and Glaeser (1997)			
	hs grad	*coll grad*	*idle*	*ln(earnings)*
Dissimilarity	.0199	.0683*	−.0024	−.0738
	(.032)	(.0398)	(.0192)	(.1063)
Dissim * black	−.326***	−.0793**	.3157***	−.7099***
	(.044)	(.0354)	(.0445)	(.1651)
N	97932	97932	97932	56390
R^2	0.039	0.0931	.0510	.0885

	Mele (2008)			
	hs grad	*coll grad*	*idle*	*ln(earnings)*
Spat. Dissimilarity	−.0880**	−.0316	.0148	−.4838***
	(.0358)	(.0465)	(.0192)	(.1098)
Spat. Dissim * black	−.1984***	−.0161	.2709***	−.8068***
	(.0706)	(.0378)	(.0624)	(.1679)
N	97932	97932	97932	56390
R^2	.0395	.0928	.0503	.0913

Standard errors corrected for clustering at the MSA level. The sample contains all 20-24 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.

Table 7: Individuals 20-24 years old, IV results

	Cutler and Glaeser (1997)			
	hs grad	*coll grad*	*idle*	*ln(earnings)*
Dissimilarity	.1330***	.2144***	−.0403	.0985
	(.0445)	(.0531)	(.0255)	(.1649)
Dissim * black	−.4136***	−.1991***	.2093***	−.8156***
	(.0836)	(.0558)	(.0888)	(.2849)
N	97932	97932	97932	56390
R^2	.0384	.0911	.0508	.0882

	Mele (2008)			
	hs grad	*coll grad*	*idle*	*ln(earnings)*
Spat. Dissimilarity	.2193**	.3535***	−.0666	.1609
	(.0932)	(.1049)	(.0439)	(.2863)
Spat. Dissim * black	−.8288***	−.3269***	.6378***	−1.487***
	(.2764)	(.1111)	(.2099)	(.5667)
N	97932	97932	97932	56390
R^2	.0325	.0811	.0486	.0877

Standard errors corrected for clustering at the MSA level. The sample contains all 20-24 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.

Table 8: Individuals 25-30 years old, OLS results

	Cutler and Glaeser (1997)			
	hs grad	*coll grad*	*idle*	*ln(earnings)*
Dissimilarity	.0165	−.0137	.0048	−.0926
	(.0238)	(.0671)	(.0249)	(.1151)
Dissim * black	−.2513***	−.0496	.2707***	−.5563***
	(.0453)	(.0512)	(.0395)	(.1416)
N	139634	139634	139634	105526
R^2	.0374	.0412	.0510	.0988

	Mele (2008)			
	hs grad	*coll grad*	*idle*	*ln(earnings)*
Spat. Dissimilarity	−.0681***	−.1507**	.0125	−.4736***
	(.0276)	(.0699)	(.0228)	(.1202)
Spat. Dissim * black	−.1823***	.0248	.1915***	−.5198***
	(.0541)	(.0522)	(.0527)	(.1371)
N	139634	139634	139634	105526
R^2	.0376	.0422	.0504	.1011

Standard errors corrected for clustering at the MSA level. 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.

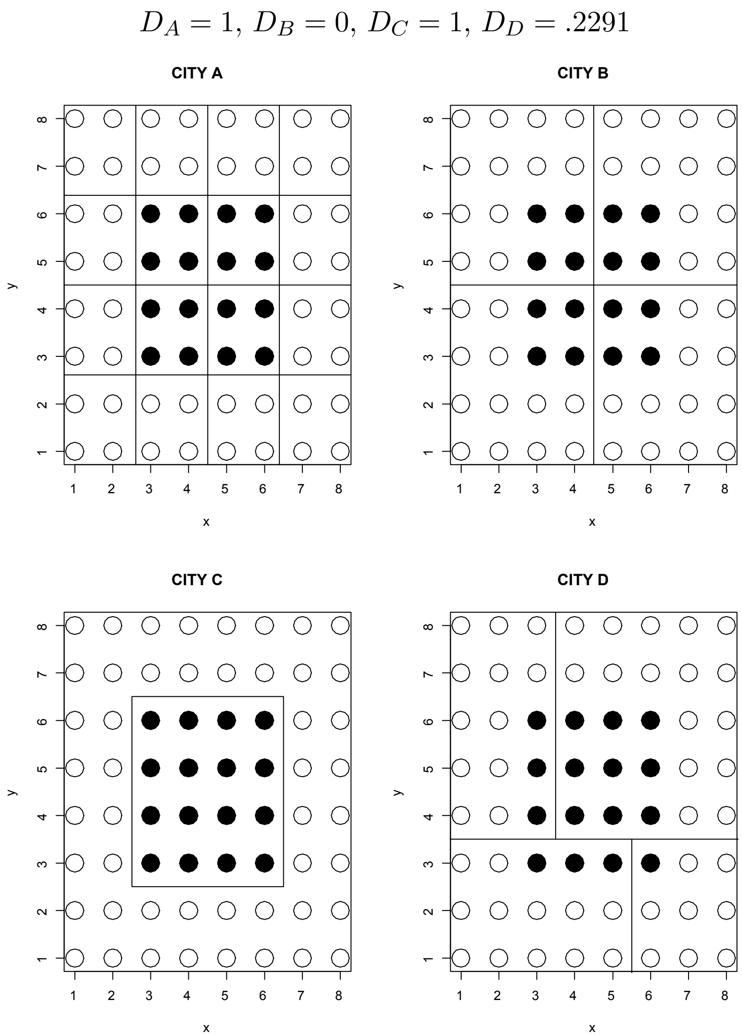
Table 9: Individuals 25-30 years old, IV results

	Cutler and Glaeser (1997)			
	hs grad	*coll grad*	*idle*	*ln(earnings)*
Dissimilarity	.0648**	.1013	.0086	.1027
	(.0317)	(.0760)	(.0278)	(.1675)
Dissim * black	−.2172***	−.1320**	.2832	−.5230
	(.0757)	(.0672)	(.0624)	(.2417)
N	139634	139634	139634	105526
R^2	.0371	.0405	.0510	.0983

	Mele (2008)			
	hs grad	*coll grad*	*idle*	*ln(earnings)*
Spat. Dissimilarity	.1061*	.1642	.0147	.1697
	(.0589)	(.1259)	(.0462)	(.2857)
Spat. Dissim * black	−.4435***	−.2403**	.5761***	−.9032*
	(.1797)	(.1193)	(.1883)	(.4992)
N	139634	139634	139634	105526
R^2	.0353	.0379	.0490	.0973

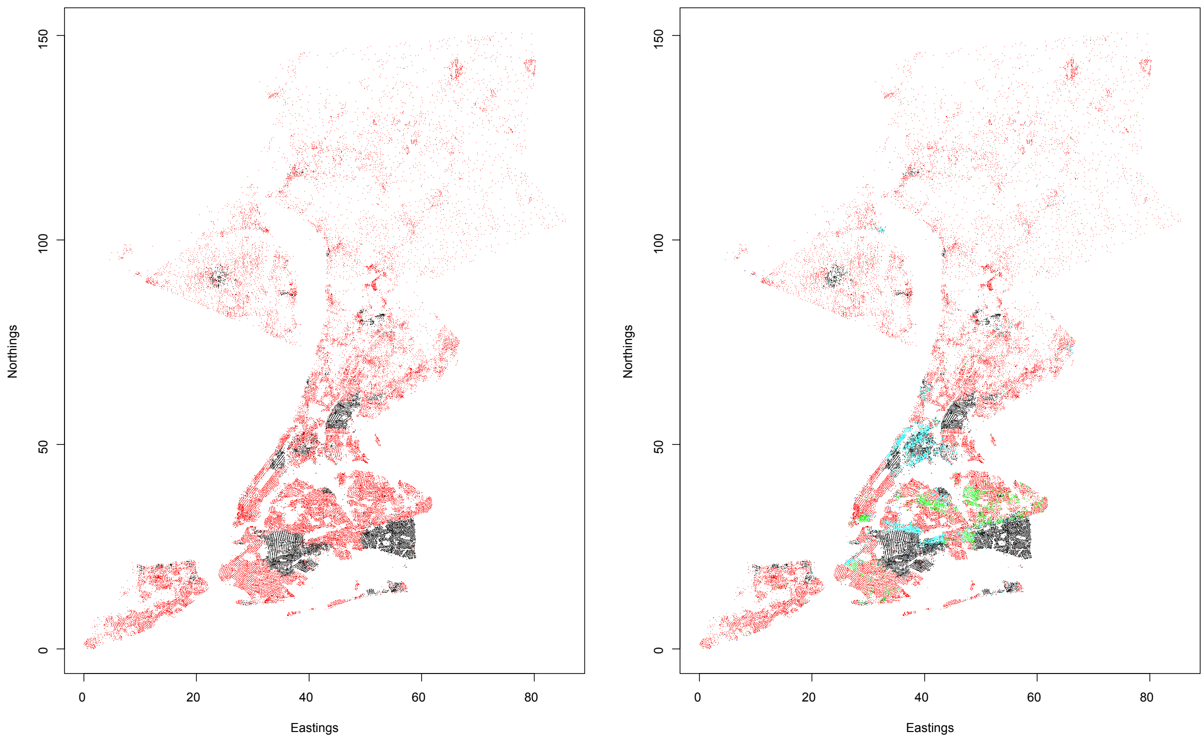
Standard errors corrected for clustering at the MSA level. The sample contains all 25-30 years old individuals born in US. I consider only the MSAs for which the fiscal variables instruments are available. Controls included but not shown: fraction of blacks in MSA, dummies for race (black, asian, hispanic 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.

Figure 1: Different partitions imply different segregation levels



Note: 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, $D_D = .2291$.

Figure 2: Spatial Distribution of Racial Groups in New York PMSA, 2000

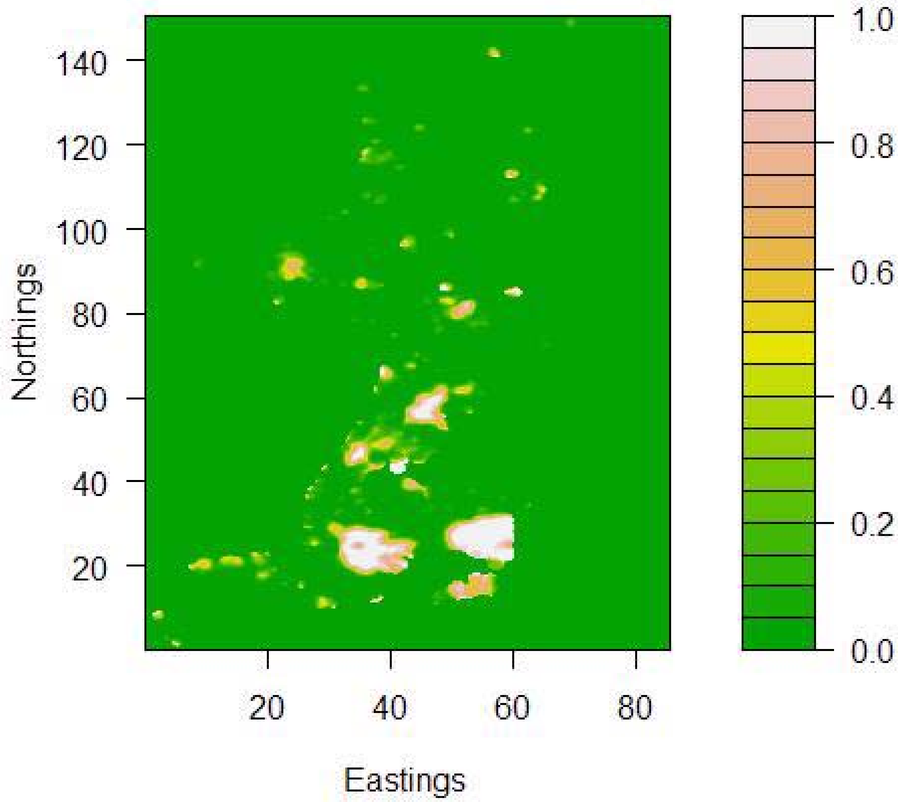


(a) African Americans (b) All Racial groups

Notes: Each point represents the centroid of a census block. A black dot represent a census block in which the majority is African American. Red represents Whites/Caucasians blocks, Green dots are Asians, Light Blue are Other racial groups. Distances are measured in Kilometers and the axis are rescaled so that the southwest corner is the origin.

Source: Summary File 1, Census of Population and Housing 2000, Us Bureau of Census.

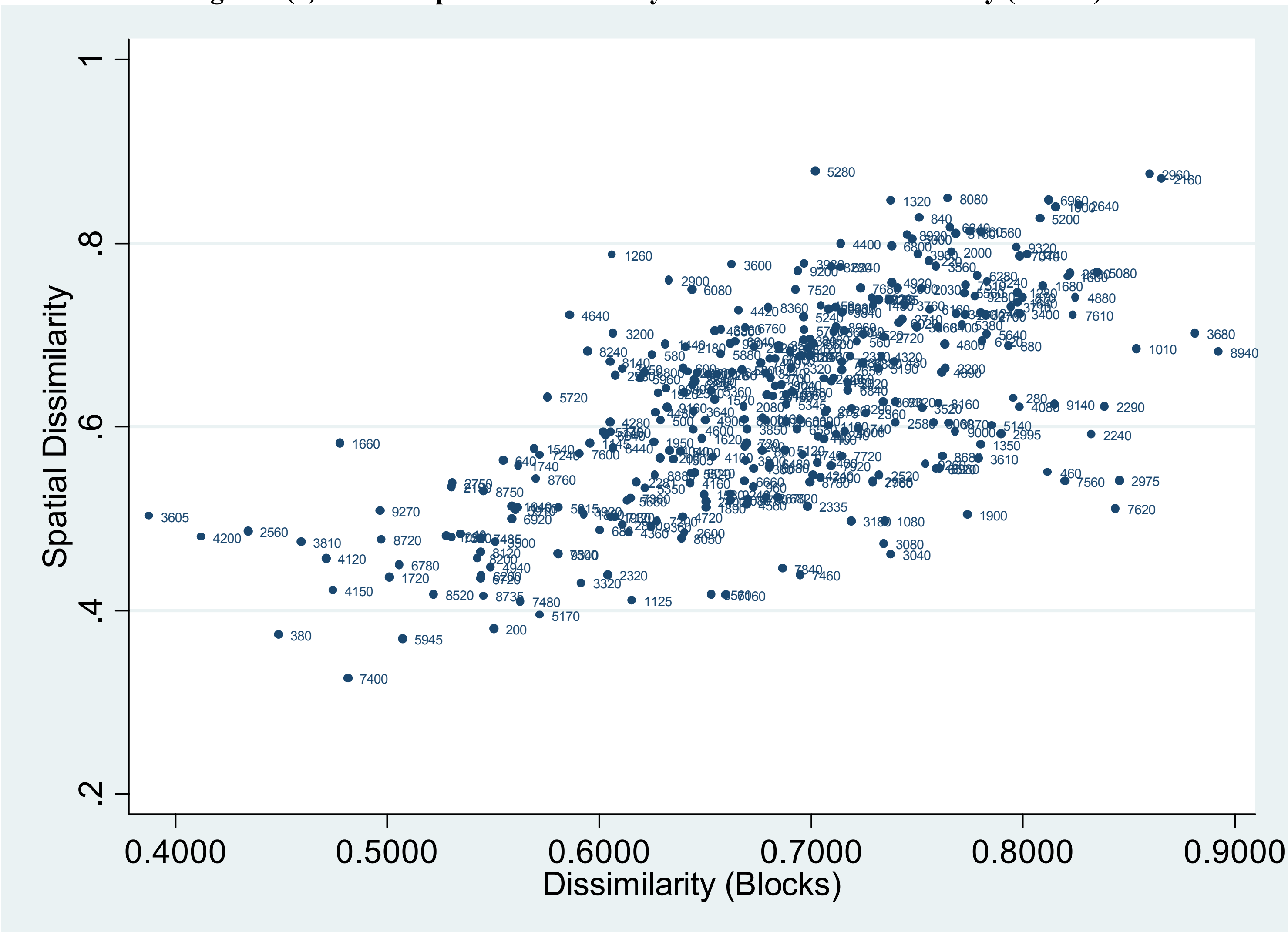
Figure 3: Estimated Conditional Probability of African American, New York PMSA 2000



Notes: Estimated conditional probability of African American location in New York PMSA in 2000. Distances are measured in Kilometers and the axis are rescaled so that the southwest corner is the origin. The areas with higher probabilities correspond to the neighborhods of Harlem, a part of the Bronx, Bedford Stuyvesant and Jamaica.

Source: Author's calculations based on Summary File 1, Census of Population and Housing 2000, Us Bureau of Census.

Figure 4(a): Blacks Spatial Dissimilarity vs Traditional Dissimilarity (Blocks)

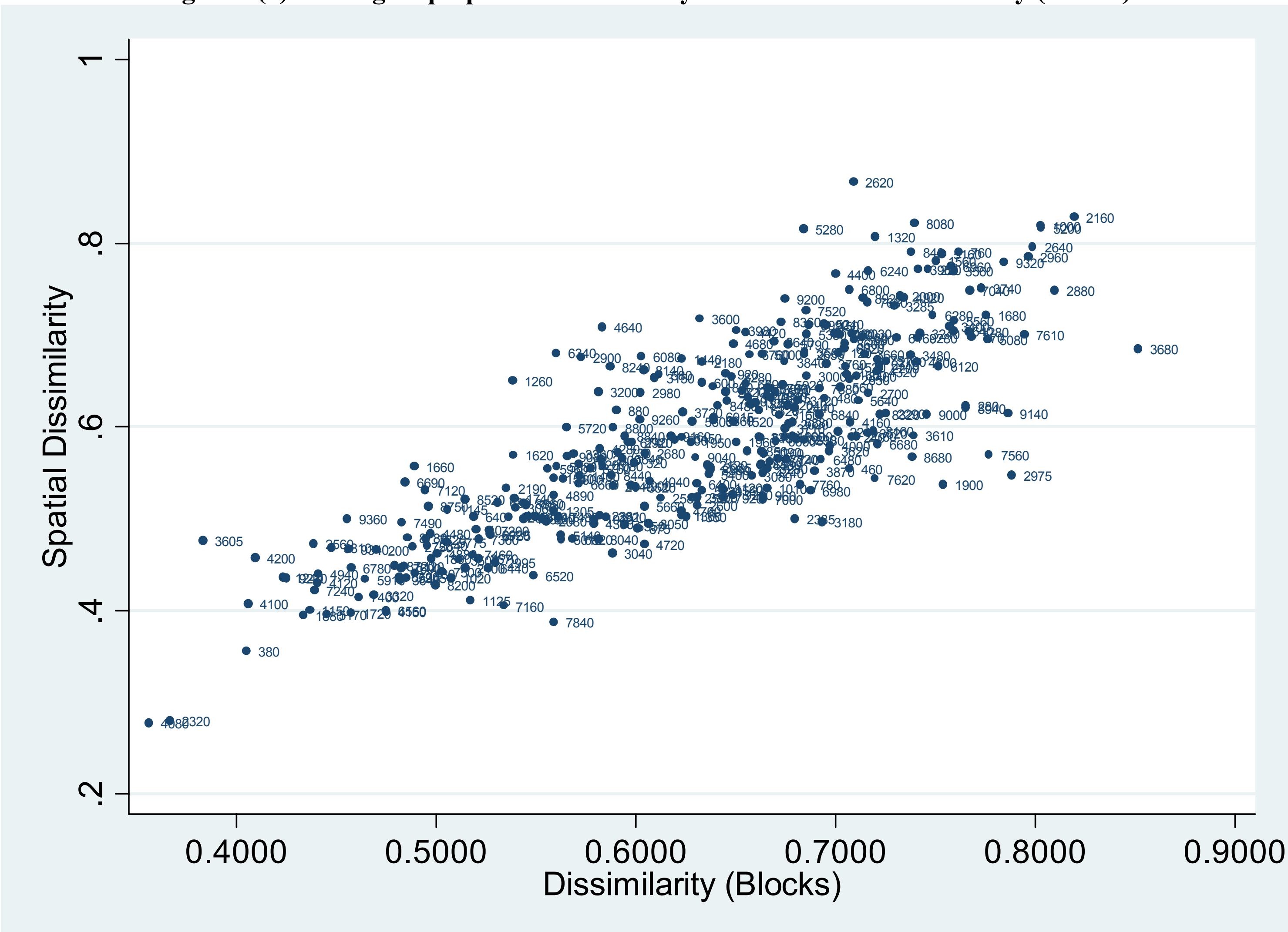


Note: Each point represents a Metropolitan Statistical Area (MSA). The marker of the points is the MSA FIPS code. The vertical axis measures the level of spatial dissimilarity for African Americans and the horizontal axis the level of traditional dissimilarity. The latter is computed using Blocks as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

Figure 4(b): Blacks Spatial Dissimilarity vs Traditional Dissimilarity (Tracts)

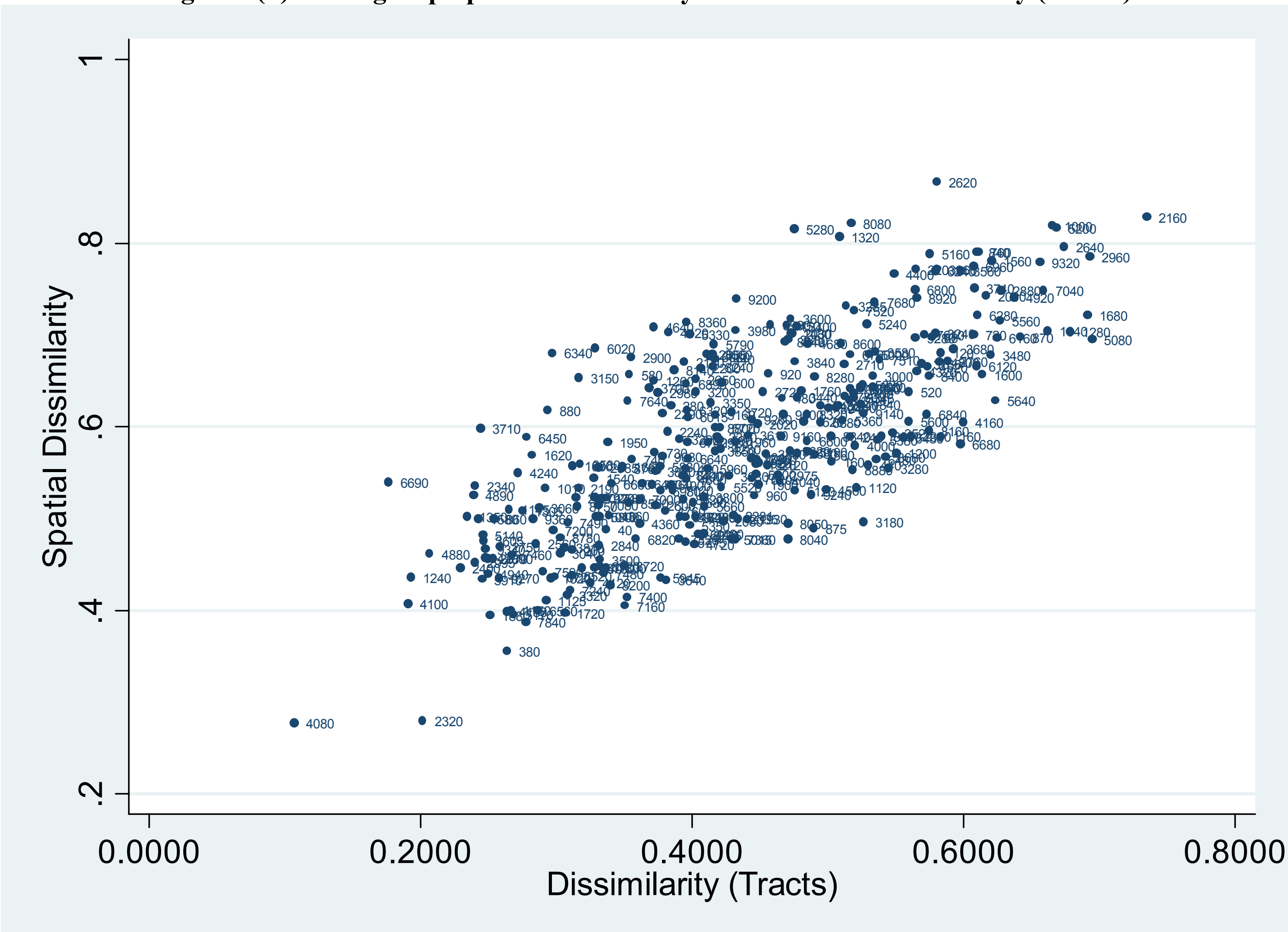
Note: Each point represents a Metropolitan Statistical Area (MSA). The marker of the points is the MSA FIPS code. The vertical axis measures the level of spatial dissimilarity for African Americans and the horizontal axis the level of traditional dissimilarity. The latter is computed using Blocks as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

Figure 5(a): Multigroup Spatial Dissimilarity vs Traditional Dissimilarity (Tracts)



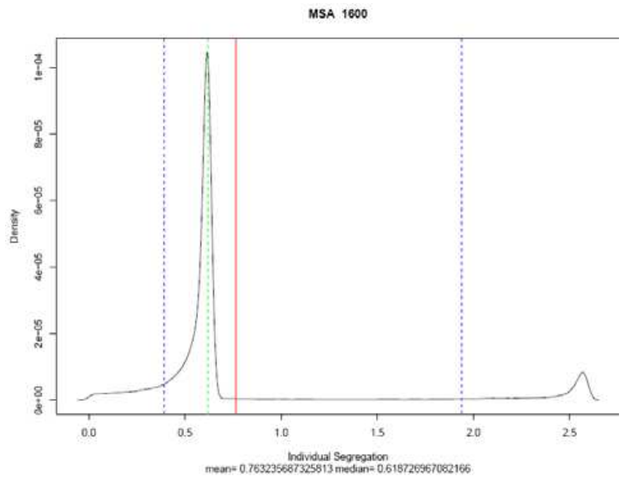
Note: Each point represents a Metropolitan Statistical Area (MSA). The marker of the points is the MSA FIPS code. The vertical axis measures the level of spatial dissimilarity for all racial groups and the horizontal axis the level of traditional dissimilarity. The latter is computed using Blocks as subunits.

Figure 5(b): Multigroup Spatial Dissimilarity vs Traditional Dissimilarity (Tracts)

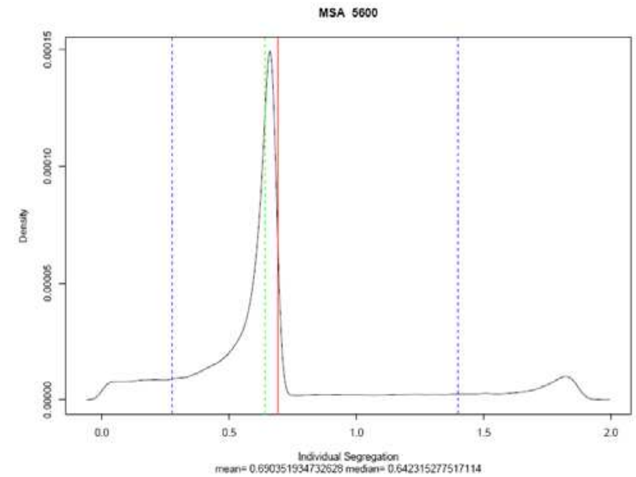


Note: Each point represents a Metropolitan Statistical Area (MSA). The marker of the points is the MSA FIPS code. The vertical axis measures the level of spatial dissimilarity for all racial groups and the horizontal axis the level of traditional dissimilarity. The latter is computed using Blocks as subunits.

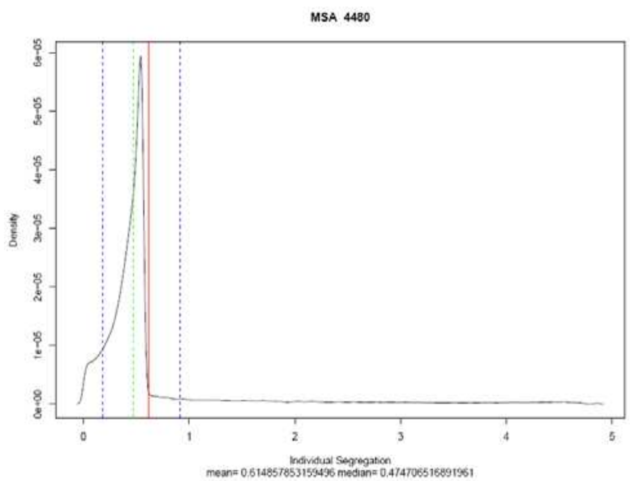
Figure 6: Individual Black Segregation Distribution



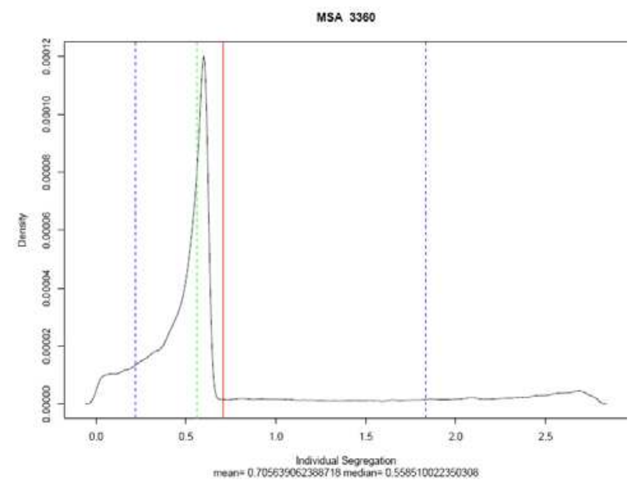
(a) Chicago PMSA



(b) New York PMSA



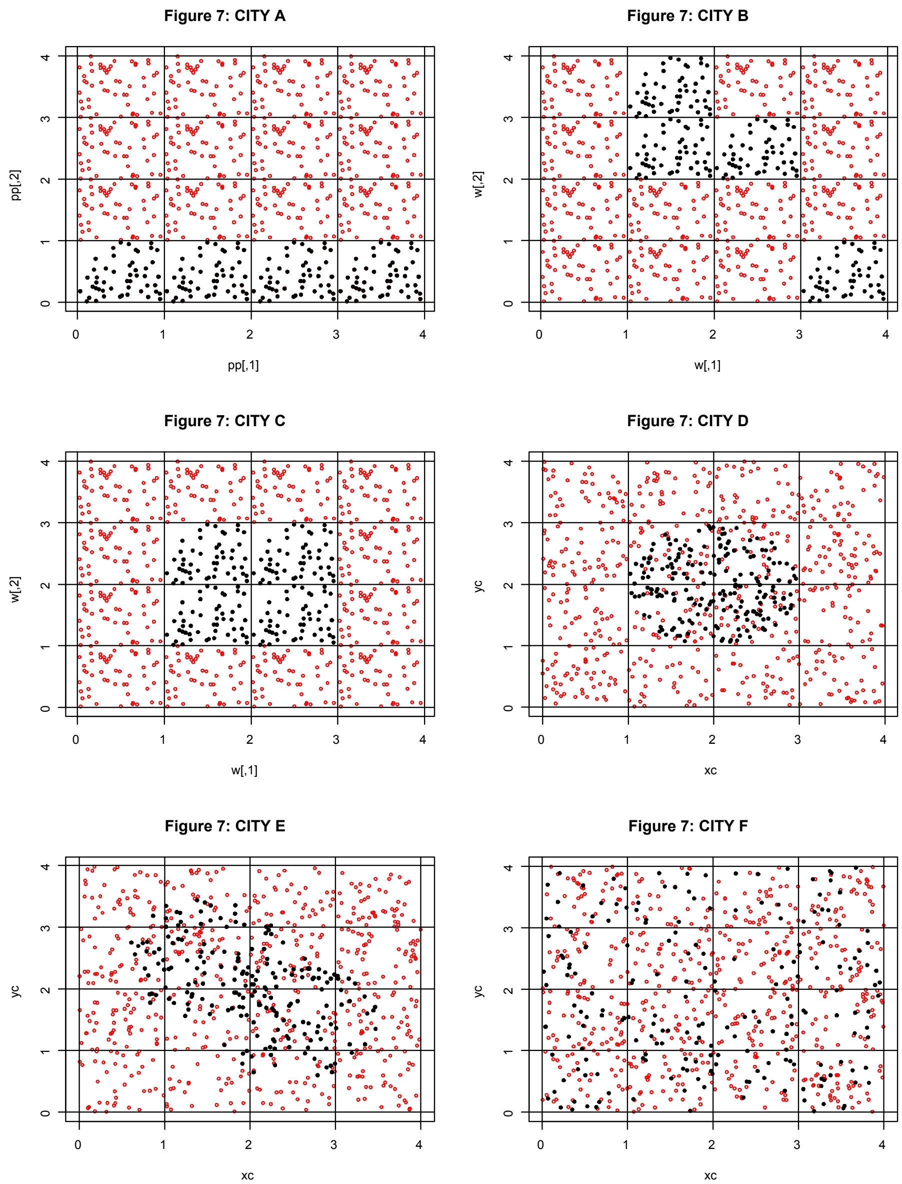
(c) Los Angeles PMSA



(d) Houston MSA

Notes: The distribution is a kernel density estimate of the empirical individual spatial dissimilarity density. The red vertical line is the average spatial dissimilarity, which is the level of spatial dissimilarity for the entire city. The green line is the median, the blue lines are respectively the 10th and 90th percentiles. It is clear that an analysis of the entire distribution is more informative than using only the index (which corresponds to the average).

Figure A1: Artificial Cities



Notes: The squares represent six artificial cities. Black dots represent the black locations and the red dots the nonblacks. Six different spatial distribution are given. City A, B and C have maximum segregation according to the neighborhood-based approach, but this is not the case with the spatial approach. CIty F is the complete integrated case under the spatial approach. City D and E are intermediate cases among these extremes. These datasets are used for the estimations presented in Appendix D.