**Title**
Assessing Inequality using Geographic Income Distributions



**Author**
Dev, Boris

**Publication Date**
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SANTA BARBARA AND

SAN DIEGO STATE UNIVERSITY

# Assessing Inequality using Geographic Income Distributions

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Geography

by

Boris Dev

Committee in Charge:

Professor Sergio Rey, Chair

Professor Arthur Getis

Professor Stuart Sweeney

Professor David Carr

September 2014

The Dissertation of
Boris Dev is approved:

Arthur Getis

Stuart Sweeney

David Carr

Sergio Rey, Committee Chairperson

December 2013

Assessing Inequality using Geographic Income Distributions



by

Boris Dev

To my loving parents, Minakshi and Jyotirmoy Dev

# Acknowledgements

I am grateful to the students, staff and faculty at the Geography Department at San Diego State University.

My advisor Dr. Sergio Rey is a pioneer in developing methods and open source software for the analysis of geographic data. I have been very lucky to work with Serge and his research team. Serge developed and mentored a research team that continues to collaborate with one another on geocomputation projects, despite the fact that we are geographically dispersed. I am especially lucky to work with Charlie Schmidt, Dr. Juan Carlos Duque, and Phil Stephens.

I am grateful to my committee. Dr. Sweeney's challenge for me to integrate the sociology literature into my dissertation improved it. Dr. Getis has been a model scholar for me, especially in terms of clarity and respect when communicating complex ideas. I am grateful to Dr. Carr for encouraging me and pushing me to think about multi-level modeling as a practical application of the metrics formulated in this dissertation.

I am grateful to Dr. Phaedon C. Kyriakidis. Some of the ideas in this dissertation germinated in his lectures on the semivariogram. Dr. Sweeney helped remind me about this.

I am grateful to Jenna for trying so hard not to fall asleep when I was practicing my defense presentation.

I am grateful to my advisors Dr. Stow, Dr. Bosco, Andy O'Clancy, Mike Brown, and and Dr. Eric West. Eric is a great geographer and friend who has advised me for many years.

Finally, I am grateful to all those who gave me a space to write in their homes, Deer Park Institute in Bir, India and Dr. Juan Duque's RISE Institute in Medellin, Columbia.

# VITA OF BORIS DEV

**Education**

| | |
|---|---|
| 2002 | Master of Arts in Geography, San Diego State University. |
| 1992 | Bachelor of Arts, University of Colorado, Boulder. Colorado. |

**Experience**

| | |
|---|---|
| 2013–14 | Developer and co-founder, Geoscore.com. |
| 2013–14 | Software engineering consultant, MapDecisions, Inc.. |
| 2011–13 | Developer, Urban Mapping, Inc.. |
| 2003–08 | Graduate Research Assistant, San Diego State University. |

**Selected Publications**

| | |
|---|---|
| 2010 | B. Dev. "Spatial Econometrics." Entry in *Encyclopedia of Human Geography*. Editor: Barney Warf. Sage Publications, Inc. 2010. |
| 2007 | Yu H., A. Kolovos, G. Christakos, J. Chen, S. Warmerdam and B. Dev. "Interactive spatiotemporal modeling of health systems: the SEKS-GUI framework." *Stochastic Environmental Research and Risk Assessment* 21: 555-572. |

| | |
|---|---|
| 2006 | Rey, S.J. and B. Dev. "Sigma convergence in the presence of spatial effects." *Papers in Regional Science* 85: 217-234. |
| 1997 | Rey, S.J. and B. Dev. "Integrating regional econometric and input-output models in a multiregional context." *Growth and Change* 28: 222-243. |

# Abstract


## Assessing Inequality using Geographic Income Distributions

Boris Dev

Ordinarily, an analysis of income differentials based on standard metrics, such as the variance statistic or the gini coefficient, implicitly weights income differentials among different places the same, regardless of whether some pairs of places are more economically interdependent than others. The problem with the assumption that all pairs of places are uniformly interdependent is that changes in those income differentials considered to be less relevant to the inequality concern being addressed may quantitatively obscure acute changes of more relevant differentials.

This dissertation has three main chapters. The common aim of each chapter is to incorporate geographic information into a metric's formulation in order to make it more relevant to an explicit concern. Each of the chapters of the dissertation share three objectives: develop a spatial view of inequality based on a concern; incorporate the spatial view into a metric's formulation using a spatial weights matrix; evaluate if the results based on spatial assessments diverge from aspatial ones.

An important empirical finding of this research is that a proposed intra-city, inter-race inequality metric registers acute differentials among latino and white

neighborhoods that an additive decomposition metric does not register. A key conceptual finding is the paradox that spatial inequality metrics formulated for different concerns can register the same change in opposite directions.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Interdependence

Geographic areas do not economically associate with one another in the same way. Some associations are competitive. Schools in the United States, for example, may battle with nearby schools for students and funding through influencing the demarcation of school attendance zones (Kozol, 2005; U.S.551, 2007; U.S.717, 1974). Other associations are cooperative. For example, economic growth in a city correlates to job growth in its contiguous rural economy, more so than to growth in the national economy (Partridge and Rickman, 2007). Walter Isard started the academic field of regional science, in the early 1950s, empirically and conceptually investigating the association between this type of economic interdependence among places and economic performance measurements (Isard, 1951). More recent regional science research has investigated the association between interdependence among places and income inequality measurements (Janikas, 2006; Kanbur and Zhang, 1999; Duque, 2004; Rey and Folch, 2011; Rey, 2004a; Rey and Dev, 2006a; Rey and Janikas, 2005; Janikas and Rey, 2005; Rey, 2004b, 2001; Rey and Montouri, 1999).

Ordinarily, an analysis of income differentials based on standard metrics[1], such as the variance statistic or the gini coefficient, implicitly weights income differentials among different places the same, regardless of whether some pairs of places are more economically interdependent than others. The problem with the assumption that all pairs of places are uniformly interdependent is that changes in those income differentials considered to be less relevant to the inequality concern being addressed may quantitatively obscure acute changes of more relevant differentials.

This dissertation proposes different methods for incorporating a spatial view of cooperative and competitive associations among places within a income inequality metric's formulation, with the aim of making the resulting measurements more relevant to different types of inequality concerns. A criteria for judging the relevance of an inequality metric is that a trend upwards in its resulting measurements reasonably reflects a worse situation regarding some explicit public concern.

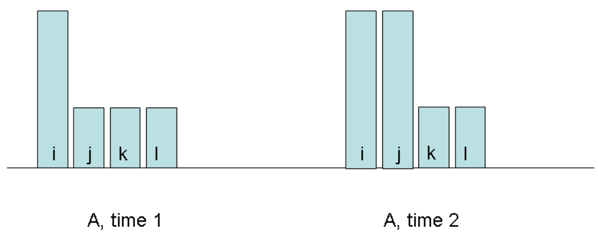
[1]Standard income inequality metrics are reviewed in Cowell (1995); Foster and Sen (1997); Temkin (1993). See Cowell (1995, p.153) for a list of metric formulations based on the summation of income differentials.

## 1.2 Spatial view

Our assessment of whether an income distribution change indicates a better or worse situation is conditional on our spatial view. A spatial view, in the context of this research, is a weighting scheme that gives some income differentials between pairs of places more relevance than others for assessing an inequality situation. Different spatial views are tied to different concerns. The geographic positioning of places and the social position of the people in those places can inform a spatial view.

To illustrate the relationship between a spatial view and an income inequality assessment, consider how it is reasonable to view the change, shown in figure 1.1, of the income distribution from time periods 1 to 2 as a *better* situation. The heights of the bars labeled $i, j, k, l$, denote the average size of the incomes for people living in these hypothetical places. Our assessment of the situation becomes more complicated when we geographically reference each place $i, j, k, l$, as belonging either to the North or the South as in figure 1.2. Looking at figure 1.2, it is plausible to conclude that the distribution has changed for the *worse*, since the people in the poorer area, $k$, of the North may have to suffer more from competing for local resources against the people in the contiguous area $j$, who experienced a rise in their incomes. Though opposite conclusions could be drawn

**Figure 1.1:** Hypothetical distribution change over time: aspatial view

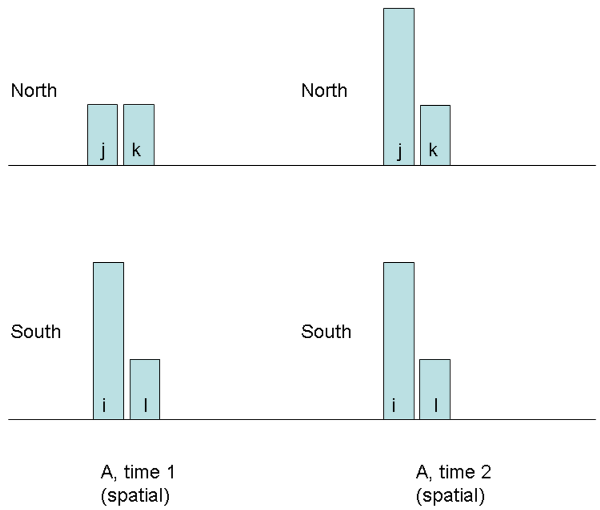



from looking at 1.1 and 1.2, aspatial metrics, those not incorporating a spatial view, would quantify both inequality situations in the diagrams as identical.

## 1.3 Spatial weights

A spatial view of whether geographic proximity represents greater competitive or cooperative relationships can be represented by a spatial weights matrix. Publicly available income data on administrative areal units, unlike less aggregate household or individual level data, contains geographic references for defining a spatial weights matrix. An element of a spatial weights matrix, $w_{ij}$, represents how a pair of areas are connected, typically using 1 or 0, where 1 denotes greater interdependence between the pair of areas than 0. For this research, the rule for

**Figure 1.2:** Hypothetical distribution change over time: spatial view




defining elements of the spatial weights matrix, $\mathbf{W}$, within the formulation of an inequality metric can be thought of as a way to formalize a notion of how some income differentials are more important than others. Once again, a criteria for judging the relevance of an inequality metric is that a trend upwards in its resulting measurements reasonably reflects a worse situation. If $I()$ is an income inequality metric that summarizes a vector $\mathbf{x}$ of income values for a set of administrative areas into a single number, then $I(\mathbf{x}_2) > I(\mathbf{x}_1)$ implies that the income distribution changed for the worse, where 1 and 2 denote the starting and ending time.[2] An aim of this dissertation is to examine how representing spatial view in

[2] Naturally, if comparisons are to be valid across time, then there must have existed reasonably consistent rules for defining the administrative areas of the geographic system being studied. For instance, the rule of the United States census tracts are "Designed to be relatively homogeneous

a metric's formulation can make its resulting measurements more relevant to an explicit aspect of inequality. In other words, $I(\mathbf{x}_2, \mathbf{W}_2) > I(\mathbf{x}_1, \mathbf{W}_1)$ is intended to be more relevant to some explicit aspect of inequality that we are concerned with than $I(\mathbf{x}_2) > I(\mathbf{x}_1)$.

The relevance problem of this research is different from the problem of making valid comparisons between different geographic systems, $I(\mathbf{x}) > I(\mathbf{y})$, where $\mathbf{x}$ and $\mathbf{y}$ are vectors of incomes from different geographic systems. The Modifiable Areal Unit Problem (MAUP) addresses the comparability problem. According to the MAUP, it is problematic to use measurements based on data referenced to a cross-section of areas to make valid comparisons amongst different geographic systems, such as countries, since differences in income inequality measurements between two geographic systems can be arbitrary due to variations in the scale of the chosen administrative unit (i.e., tract, county, state) (Oppenshaw and Taylor, 1991) and regionalization (or zonation) scheme (Duque, 2004; Rey and Folch, 2011).

To understand the meaning of a spatial weights matrix more technically, consider $i$ to be an index for a set of communities (say city neighborhoods), $i = \{1, \cdots, n\}$; And, consider $X$ to be a spatial random process and $X_i$ some economic outcome of the process for area $i$, which can denote income, or some-

units with respect to population characteristics, economic status, and living conditions, census tracts average about 4,000 inhabitants."(Census Bureau, 2013).

thing that income proxies such as the capability of a place to access local goods like education or health. The spatial weights matrix, $\mathbf{W} = [w_{ij}]$, can be thought of as follows:

$$w_{ij} \quad \begin{cases} 1, \text{ if } P(x_i) \neq P(x_i|x_j) \text{ , with } i, j \in I \text{ and } i \neq j \\ 0, \text{ otherwise;} \end{cases}$$

If the element $w_{ij}$ of the matrix is 1 then the probability of outcome in one place $i$ is dependent on the outcome in an another interconnected place $j$.

## 1.4 Overview of chapters

This dissertation has three main chapters. The common aim of each chapter is to incorporate geographic information into a metric's formulation in order to make it more relevant to an explicit concern. Each of the chapters of the dissertation share three objectives:

- Develop a spatial view of inequality based on a concern.
- Incorporate the spatial view into a metric's formulation using a spatial weights matrix.
- Evaluate if the results based on spatial assessments diverge from aspatial ones.

Chapter 2 proposes a metric to identify subsets of high income differentials tied to a concern over acute inter-community competition over local resources. To represent competitive associations, chapter 2 incorporates spatial structure by giving greater weight to income differentials between pairs of census tracts of different racial classification that are co-located within the same metropolitan zone. Borrowing from the field of exploratory spatial data analysis (ESDA) (Anselin and Getis, 2010; Janikas and Rey, 2005), chapter 2 uses spatial permutations (Rey, 2004a) for assessing the empirical significance of incorporating geographic and racial references to define the spatial weights matrix of a metric.

Chapter 3 proposes a convergence metric tied to a concern for institutional unfairness in the distribution of income across states of the United States. Chapter 3 incorporates spatial weights into the metric by giving less weight to income differential changes attributed to economic spillovers from contiguous states of the United States. Extending Rey and Dev (2006b), chapter 3 implements spatial filtering as a method for down weighting income dispersion changes associated to spatially autocorrelated residuals from a income convergence regression. Spatial autocorrelation describes the geographic clustering of like values.[3] Rey and Folch (2011) show that making comparisons using segregation metrics may not be valid

[3]For Getis (2008), the best definition to spatial autocorrelation is "Given a set S containing n geographical units, spatial autocorrelation refers to the relationship between some variable observed in each of the n localities and a measure of geographical proximity defined for all $n(n-1)$ pairs chosen from n (Hubert et al., 1981, p.224)."

since their results are impacted not just by the dispersion parameter in the data generating process of the income values, but also by each city's different geographic dimensions, such as zonation, geographic extent, scale, and, most relevant to this discussion, spatial autocorrelation. Chapter 3 of this dissertation proposes a metric that filters out the impact of spatial autocorrelation on income dispersion change.

Chapter 4 proposes a metric tied to a concern for identifying specific segments of a school attendance zone boundary that might act as barriers to a child's cooperative encounters with children of other income classes. Chapter 4 incorporates spatial structure into the metric by giving greater weight to income differentials among blocks that are contiguous, to proxy potential encounters, and that are located on different sides of an attendance boundary, representing a social barrier attributable to school competition over higher income students. Borrowing from Rey (2004a), Chapter 4 uses spatial permutations for identifying local "hot spots", where a boundary's segment appears to split areas by income size. It is important to note, the aim of chapter 4's metric, like chapter 2's metric, is not to inform a concern for the culprit of the boundary design, but rather to highlight areas, in the spirit of exploratory spatial data analysis (ESDA) (Anselin and Getis, 2010; Janikas and Rey, 2005), in terms of a concern for the effects of institutional boundaries.

## 1.5 Related literature

Like this research, sociologists Morrill (1991); Jargowsky and Kim (2005); Wong (1993); Reardon and O'Sullivan (2004) represent geographic proximity as spatial weights with the aim of making their residential segregation metrics more relevant. Morrill (1991); Wong (1993); Reardon and O'Sullivan (2004)'s specific aim is to attack the checkerboard problem of the most common metric of segregation, the dissimilarity index (Duncan and Duncan, 1955). The dissimilarity index is the percentage of the poor population that would hypothetically need to relocate to a new area in order to make each area have the same composition of poor people. White (1983), who first coined the term, the checkerboard problem, describes it as follows: "Allow the squares on a checkerboard to represent parcels (neighborhoods, tracts, blocks). Once the composition of each parcel (square) is given, any spatial rearrangement of them will still result in the same calculation for the dissimilarity index. A city in which all the nonwhite parcels were concentrated into one single ghetto would have the same level of calculated segregation as a city with dispersed pockets of minority residents." Sociologist's spatial segregations metrics (Morrill, 1991; Jargowsky and Kim, 2005; Wong, 1993; Reardon and O'Sullivan, 2004) give less weight to income and racial composition differentials among areas in geographic proximity. The premise is that differentials

**Table 1.1:** Framework of spatial inequality metrics

| | Types of concerns: | |
|---|---|---|
| Correlates of proximity: | effect | cause |
| cooperation | spatial segregation literature | chapter 3 |
| competition | chapter 2, chapter 4 | Chakravorty (1996) |

among neighboring areas are not as bad as differentials among non-neighboring areas since they indicate potential encounters between black and white, or rich and poor. This downward adjustment is tied to a concern for the effects of residential segregation on limiting the interaction among racial groups (Morrill, 1991; Wong, 1993; Reardon and O'Sullivan, 2004) and income groups of the same race (Jargowsky and Kim, 2005).

Table 1.1 organizes the difference between the research on spatial segregation metrics, discussed above, and this dissertation research along two dimensions. The first dimension, along the table columns, divides metrics by whether they are are tied to a concern of a cause or an effect (outcome) of income and racial imbalances across administrative areas. The second dimension, along the table rows, divides metrics based on whether geographic proximity represents greater competitive or cooperative relationships.

Chapter 3's metric, like spatial segregation metrics, equates proximity to cooperation. Chapter 3, unlike spatial segregation metrics, is tied to a causal concern, the political concern that unfair institutional arrangements might cause higher income differentials among states. In contrast, spatial segregation metrics are tied to a concern for the effects of segregation: the concern that residential imbalances of race (Morrill, 1991; Wong, 1993; Reardon and O'Sullivan, 2004) and income (Jargowsky and Kim, 2005) are instrumental in reducing social encounters, which can lead to the social isolation of poorer minority groups.

Chapters 2's and 4's metric, like spatial segregation metrics, are tied to concerns over the effects of income and race imbalances. Chapters 2 and 4, unlike spatial segregation metrics, equate proximity to competition. In chapter 2, proximity represents competitive associations among ethnic neighborhoods of the same metropolitan zone, and in in chapter 4 proximity represents competitive associations between two schools with contiguous attendance zones

Similar to the sociologist's spatial segregation studies, from Partridge and Rickman (2007)'s perspective proximity between poor rural areas and rich urban areas equates to cooperative associations. Using county cross-sectional observations, they provide evidence that a rural area's proximity to the industries of a rich metropolitan area is associated to more jobs in the poor rural area.

To this author's knowledge few, if any studies, exist to fill the bottom right quadrant of table 1.1 (proximity as competition, concern as unfairness), with the possible exception of Chakravorty (1996). (Chakravorty, 1996, p.1672) proposes a measure of "spatial (or neighbourhood) disparity based on the difference in value between a parcel and the (weighted) average of its neighbours." (Chakravorty, 1996, p.1684) says "It may also be useful to find out how well the new measures provide information on social inequality…" (Chakravorty, 1996, p.1684) hints that his metric could inform concerns of policy unfairness:"The relationship between inequality in space and abstract inequality is particularly intriguing, as it touches upon critical and unexamined policy issues…".

## 1.6 Specific versus general concerns

It is important to distinguish between this research's explicit, narrowly defined concerns regarding income differentials among places versus broader notions of inequality. Cowell (1995); Temkin (1993); Foster and Sen (1997) associate the relevance of inequality metrics to broader notions, such as "how people fare relative to one another", social welfare, social justice, and egalitarianism. Using the measurements of a spatial metric to make general statements on these broader notions of inequality is problematic. Incorporating a spatial view into the formulation of income inequality metrics may increase their relevance to addressing some specific concerns while potentially decreasing their relevance to other specific concerns and broader notions of inequality. If we do not tie an explicit concern to an explicit spatial view of a metric's formulation then the relevance of our inequality assessments will be ambiguous.

To illustrate how different spatial structures are associated to different concerns, consider the following: In what sense would bigger income differentials among places be viewed by an impartial spectator [4] as a good or bad situation? Below are three different inequality concerns that are associated with three different spatial structures. First, if the spectator's concern is political, such as Rawl's notion that unfair institutional arrangements can cause worse income dispersion, then the most relevant income differentials changes to be registered by a metric should be those associated with independent places connected by a shared political system, reflecting the concern in chapter 3. Second, if the spectator's concern is about the effect or outcome of income differentials on the capability of poor places to compete with other places for local resources, such as food, schooling, land or medical care, then the most important income differentials to account for would be between pairs of places in close enough proximity to one another to compete for local resources, reflecting the concern in chapters 2 and 4. Third, if the spectator's concern is about the effect of residential income segregation limiting social encounters between rich people and poor people, reflecting the concern of the sociologists (Morrill, 1991; Wong, 1993; Reardon and O'Sullivan, 2004; Jargowsky and Kim, 2005), then the most important differentials would be among

[4] This use of an "impartial spectator" for a mental experiment is extended from Smith (1776) and Foster and Sen (1997).

contiguous places where there is a lesser chance for rich and poor cooperative social encounters.

The three above income inequality concerns are tied to three different spatial views. A hypothetical change in the spatial pattern of incomes, where the richest people moved closer to the poorest, will have a different meaning in terms of these different concerns. On one hand, this would be irrelevant for the first concern for assessing unfairness, since it is a endogenous, voluntary change. On the other hand, this change would be registered as a decrease (a better situation) in the sociologist's spatial segregation measurements, which are tied to the third concern above relating to social encounters among different groups. In contrast, this change could be registered in the opposite direction, as an increase (a worse situation), by measurements aimed to assess local resource competition, the second concern above, and motivating chapters 2 and 4,

In summary, spatial inequality measurements tied to broad notions, instead of specific concerns, of inequality can lead to ambiguous interpretations. Different specific concerns require different definitions of what is the most relevant spatial structure connecting places. Most importantly, do we equate geographic proximity with competition or cooperation? The answer to the question is conditional on the type of specific concern we prioritize.

## 1.7 Unfairness as a cause versus suffering as an effect

Specific concerns for assessing income differentials among administrative areas can be split in two by direction of causality. One set of concerns relates to an aversion to unfairness in governmental institutions associated with causing higher income differentials among places. The second set of concerns relates to an aversion to suffering as an effect (or outcome) of higher income differentials.

Rawls and Kelly (2001, p.59) explain the concern for unfairness: to "...assess the basic structure according to how it regulates citizens' shares of primary goods... ". Decreasing income differentials among places can indicate progress, or a just economic system, as Rawls (1999, p.78) states "Eventually the resulting material benefits spread throughout the system to the least advantaged." In terms of unfairness as a cause, examples of empirical studies that explicitly tie interregional income inequality measurements to unfairness in government policies are as follows. At the regional level, rising regional income inequality measurements have been associated to trade liberalization policies by Sánchez-Reaza and Rodríguez-Pose (2002) and Rivas (2007). Fan and Casetti (1994) interpret rising state income dispersion as evidence against the Neoclassical economic theory's prediction(Solow, 1956; Barro and Sala-i-Martin, 2004a) that the current political economy based on market principles leads to regional income convergence. At the inter-neighborhood level, persistent trends in Black-White residential segregation measurements are associated to a lack of institutional progress in dealing with historically unfair housing discrimination (Massey and Denton, 1993) , as well as attitudes on race Massey and Denton (1987, p.823). Lieberson and Carter (1982) attribute 85 percent of black segregation in 1970 to involuntary causes. From a qualitative perspective, Kozol (2005) blames politically influenced demarcation of attendance to large income differentials between rich and poor school attendance zones in New York City.

Rey and Dev (1997) and Rey and Folch (2011) point out how a problem of using changes in inequality measurements to make causal inferences is that income values among areas may not be independent of one another. For instance, spatial segregation metrics of Morrill (1991); Wong (1993); Reardon and O'Sullivan (2004); Jargowsky and Kim (2005) are not suited to make inferences on institutional unfairness as a culprit to increased measurements, since these metrics are formulated to be sensitive to changes in the level of spatial autocorrelation. In other words, as Rey and Folch (2011, p.431) point out, "the formulation of these spatial measures assumes that regions with greater clustering of similar neighborhoods are more 'segregated'."

Trend changes in spatial income autocorrelation could reflect changes in the impact of endogenous forces on spatial income distribution change. Changes in the trend of spatial income autocorrelation across time have been recorded by Rey and Montouri (1999) and Rey and Dev (1997). Rey and Folch (2011) point out that changes in spatial autocorrelation of neighborhood incomes can reflect changes in interacting choices of residents leading to residential clustering of racial and income groups (de facto segregation), an issue separate from institutional unfairness (de jure segregation). The consequences of interacting residential choices leading to a racially segregated city is described by the Schelling segregation model (Schelling, 1971; Zhang, 2004; Clark, 1991).The consequences of residential preferences leading to a income class segregated city is described by the Tiebout class of models (Tiebout, 1956; Epple, 2003; Epple and Romano, 2003; Epple and Platt, 1998). When we observe changes in the trend of spatial income autocorrelation across time this could reflect changes in the impact of Tiebout and Schelling type endogenous forces on how people geographically sort themselves across areas. The equifinality problem here is about the the ambiguity of accounting for what portion of a changing trend in a spatial inequality measurement stems from changes in institutional unfairness (exogenous) or from changes in voluntary choices (endogenous).

It is important to note that this equifinality problem is not an issue for metrics aimed at concerns on the effects or outcomes of high income differentials, which are proposed in chapters 2 and 4.

Providing a conceptual framework to empirical studies on the effects of income differentials on suffering, we can extend Sen and Foster (2006)'s notion of 'Smithian interdependencies' from individuals to social groups. Sen and Foster (2006, p.213) explain 'Smithian interdependencies' paraphrasing Adam Smith (Smith, 1776, pp.351-2): "appearing in public without shame" may require more expensive clothes in a richer country than a poorer one, given by the established standards. The same applies to the capability of "taking part in the life of the community." Sen (1997) argues that relative income deprivation is not just unfortunate because some are poorer than others, but rather its importance lies in its potential to be instrumental in the deprivation of people's absolute economic capabilities.

In terms of suffering as an effect, examples of empirical studies that explicitly tie bigger income differentials among places to specific negative outcomes are as follows. In terms of regional income differentials within a country, Kanbur and Zhang (1999) and Xue (1997, p.46) warn that "differences may create serious social and political problems, generate nationalist conflicts and negatively influence China's economic and social stability." Income differentials among neighborhoods of the same city are associated to the deterioration of public infrastructure in the inner city (Massey and Denton, 1993), middle class migration (Wilson, 1987a), and role model migration (Durlauf, 2004). Kozol (2005) associates income differentials between rich and poor school zones of the same city to feelings of social isolation.

Multi-level modeling studies use aspatial income inequality metrics as explanatory variables in regressions to correlate economic inequality of an individual's surroundings to lower health outcomes (Krieger et al., 1997; Diez-Roux et al., 2000; Kennedy et al., 1998; Subramanian et al., 2001, 2004), as well as voting outcomes (Galbraith and Hale, 2008). Evidence exists, however, that a spatial metric, giving more weight to income differentials between different social groups located in the same city, may be an alternative to standard aspatial inequality metrics for explaining outcomes such as crime (Blau and Blau, 1982), the economic outcomes of children (Borjas, 1995), urban medical care (Deaton, 2003) and urban famine (Sen, 1977). Amartya Sen frames the concept of income inequality as an instrumental cause of economic inequality, the broader concept (Sen, 1997). As a possible future direction of research, the spatial income inequality metrics proposed in chapter 2, based on income differentials among urban areas of different ethnicity, and chapter 4, based on differentials among areas on different sides of a school attendance boundary, could be considered as an alternative to aspatial income inequality metrics in multi-level modeling studies, if they are considered closer proxies to capturing to the capability deprivation aspect of surrounding economic inequality within the multi-level model.

Studies, such as Levernier et al. (1998); Partridge et al. (1996); Levernier et al. (2000, 1995); Partridge et al. (1998), also analyze administrative income data (state and county), but they have a different spatial focus and motivating concern than this research. They assess *intra*-area family income differentials, as opposed to this research's focus on *inter*-area differentials. They investigate potential factors influencing family income inequality, such as increasing returns to education, industrial restructuring, international trade, immigration, suburbanization, urbanization, industrial composition, and female labor-force participation. Their motivation is to understand how place-specific attributes may play a causal role in rising income inequality, and not related to this research's concerns for assessing unfairness or identifying subsets of the most relevant income differentials associated to bad outcomes. An interesting empirical finding of these studies (Levernier et al., 1998; Partridge et al., 1996; Levernier et al., 2000, 1995; Partridge et al., 1998) is that increasing convergence at the interregional level co-exists with divergence at the intraregional level.

## 1.8 Summary

We cannot directly observe how people are connected with one another. However, we can approximate the spatial structure of how people are economically connected by cooperative and competitive associations using information on their geographic and social positions. In order to assess inter-area income inequality with public data on administrative areal units, census tracts (in chapter 2), states (in chapter 3), and school attendance zones (in chapter 4), this dissertation proposes new metrics that incorporate spatial weight matrices representing the impact of cooperative and competitive associations on either the causes or effects of spatial income distribution change. A quantitative determination of whether income inequality has gotten better or worse can be thought of as being a function of two pieces of information: one, an explicit concern of why an income differential between two places is bad, and, two, an explicit spatial view of how change in some income differentials among different areas is more relevantly associated to an explicit concern than other differentials. The premise of this research is that the relevance of a change in the income differential between two places is not only related to its size, but also to how the people of the different places are interdependent.

# Chapter 2

# Assessing income inequality across neighborhoods

## 2.1 Introduction

Amartya Sen frames the concept of income inequality as an instrumental cause of economic inequality, the broader concept (Sen, 1997). He argues that relative income deprivation is not just unfortunate because some are poorer than others, but rather its importance lies in its potential to be instrumental in the deprivation of people's absolute economic capabilities. Sen and Foster (2006, p.213) uses the term 'Smithian interdependencies' for a notion he explains by paraphrasing Adam Smith (Smith, 1776, pp.351-2), of "appearing in public without shame" may require more expensive clothes in a richer country than a poorer one, given by the established standards. The same applies to the capability of "taking part in the life of the community." Extending the notion of Smithian interdependencies from an inter-personal to an inter-group level, a problem of standard inequality measurements is that they may obscure acute changes in subsets income differentials associated to the negative outcomes of competition among ethnic neighborhoods over local economic resources such as school funding (Kozol, 1991), role models (Durlauf, 2004), or quality medical care (Deaton, 2003).

The aim of the proposed metric of this chapter is to identify a pattern of widening income differentials between people of different social positions who live within the same geographic zone, or what I term inter-social, inter-zonal inequality. In contrast to additive decomposition approaches, which are useful in evaluating the extent of income inequality in a proportional sense for different parts society (Sen and Foster, 2006), the proposed spatial metrics of this chapter capture the intensity or acuteness of subsets of income differentials. Extending Rey (2004a)'s spatial randomization method, this chapter's approach can identify a significant difference in the average size of a particular subset of income differentials compared to socially or geographically more aggregate sets of differentials. I provide an empirical illustration using the income differentials between over 44,000 census tracts of the United States, referenced by 4 social groups (Black, White, Latino, and Mixed) and 238 metropolitan zones, for the 1980, 1990, and 2000. Intra-zonal differentials are defined using pairs of tracts that are co-located in the same metropolitan area of a total of 238 metropolitan zones.

For assessing the relevance of the spatial (or interdependent) perspective for describing changing income inequality, I propose two necessary conditions. (1) There needs to be a substantive distinction between the effect on economic inequality of an income gap change between two particular social groups in the same zone and other aspatial (or aggregate) changes, ones that ignore the social and geographic references this chapter's approach jointly uses to organize the observations. (2) There needs to be empirical evidence that a change in the average size of this subset of inter-social, intra-zonal differentials is significantly different than changes associated with socially or geographically more aggregate sets of income differentials. Next, I introduce the substantive issue, and then the empirical one.

(1) I argue that a widening of the income gap between two associations of people does not inherently lead to more economic suffering; The affect is conditional on how the gap alters the ability of at least one of the associations to obtain economic resources. The premise here is that the ability of a social group or community to convert income into certain economic capabilities is coupled to their income relative to other groups (or communities) in their same geographic zone. For example, once again, inter-social, intra-zonal income inequalities can cause deprivation from competition between communities over locally bounded economic resources such as school funding (Kozol, 1991), role models (Durlauf, 2004), or quality medical care (Deaton, 2003).

(2) Differentiating between a subset of inter-social, intra-zonal differentials and a socially or geographically more aggregate set of differentials is empirically im-

portant only if the disaggregate subset of differentials changes in a significantly different way than the aggregate one. Extending Rey (2004a), I present a permutation test for identifying a statistically significant difference in size between a intra-zonal, inter-social measure and its socially or geographically more aggregate counterpart. The permutation test shuffles the geographic and social labels of neighborhoods to build empirical distributions of the measures.

As a preview, this chapter's results provide evidence of how using a separate intra-zonal measure and a separate inter-social measure may not necessarily capture what a jointly defined inter-social, intra-zonal measure can: Between 1980 and 1990 the observed inequality *reducing* change associated with differentials between White and Mixed neighborhoods located in the same metropolitan zones was opposite the inequality *enhancing* change of both the socially aggregate intra-zonal measure and the geographically aggregate White-Mixed measure. In addition, this chapter's results show that the large size of the average income differentials associated with Latino and White pairs of neighborhoods located in the same metropolitan zones would be masked by relying on a proportional perspective of an additive decomposition of variance, as well as underestimated by socially and geographically more aggregate equivalent measures.

The chapter proceeds as follows. In section 2.2, I discuss the spatial (or interdependent) perspective of measuring changes in income inequality. In section

2.3, I propose a framework to account for spatial structure in inequality measures. In section 2.4, I discuss an empirical illustration, and, in section 2.5, I end with concluding remarks.

## 2.2 Spatial Weights

Standard income inequality measures based on geographic data make an implicit aggregate assumption that the income gap between two subareas in faraway places is as important as the gap between nearby subareas. To illustrate this, I examine how the variance statistic, $V$, measures changes in income inequality. Consider distributions of income over $n$ subareas (neighborhoods), $i = 1, ..., n$, and let $y_{i,t}$ be the logarithm of per capita income of subarea $i$ at time $t$. Let the average level of income for all subareas be $\bar{y}_t = \frac{1}{n}\sum_{i=1}^{n} y_{i,t}$, so that $V$ is given by:

$$V_t = 1/n \sum_{i=1}^{n} (y_{i,t} - \bar{y}_t)^2 \tag{2.1}$$

Evidence of reduced economic inequality is reflected in equation (2.1) declining over time, such as:

$$V_2 < V_1 \tag{2.2}$$

2 denotes an ending time period, and 1 denotes a starting time. A problem with this description of inequality is that it is possible that an observed flat trend in aggregate income dispersion $V_t$ may hide discordant shifts in important income differentials, or distributional churning.[1] Therefore, the most important income differentials are the ones associated with interdependent parts of society that are particularly sensitive to the link between changes in their relative incomes and economic suffering. For example, substantively, income gap changes between contiguous communities that compete for public school funding in the United States would be more important than changes between family and student neighborhoods.

For this chapter, the problematic implicit assumption of aggregate descriptions is how they weigh all income differentials the same irrespective of the importance of the geographic or social positions of people. I use the term 'income gap' to refer to the squared difference in the logarithm of income between two subareas,

[1] This issue is no longer relevant if differentials between social positions or subareas move in concordance with one another, such as the consequence of a catch-up convergence process (Bernard and Durlauf, 1996). At the regional scale, examples against such a broad based characterization of income gap movements are Quah (1996b)'s results of an evolving bi-polar distribution of European regions and Rey (2002)'s results of spatially conditioned inter-mixing of relative ranks for US states. Though all deal with spatial aspects of income distribution changes, this chapter is different from Quah (1996b) and Rey (2002), because of two reasons. First, this chapter's approach is to use social information jointly with geographic information to represent spatial structure. Second, this chapter's method deal with providing information on economic inequality at small area or neighborhood scale in terms of potential suffering as an *effect* of income inequality, instead of regional economic convergence as a *cause* of lower inequality.

such as:

$$d_{ij} = (y_i - y_j)^2 \tag{2.3}$$

where $i, j$ are pairs of subareas. I believe it is important to differentiate amongst income differentials based on the distance between $i, j$ in terms of affecting the probability that the associated communities compete with one another for economic resources.

Consider how a pair-wise computation of variance, based on income differentials, leads to an equivalent result as equation (2.1). Income variance can be understood as twice the average of all income differentials $d_{ij}$, as follows:

$$2V = \frac{\sum\sum d_{ij}}{n(n-1)} \tag{2.4}$$

The denominator of the above equation, $n(n-1)$, is the total count of all possible differentials between subareas $i, j$. The numerator of the RHS of equation (2.4) is the sum of the squared deviations of income differentials, from equation (2.3). $2V$ is equivalent to the average of the universe of such differentials.

To see how spatial structure is implied by the variance statistic consider it as weighted average of all income differentials.

$$2V = \frac{\sum\sum w_{ij} d_{ij}}{n(n-1)} \tag{2.5}$$

By defining all $w_{ij}$'s as 1 makes explicit the assumption that the spatial structure of the income differentials is irrelevant to capturing the effects of income differentials on economic welfare.[2]

When the intent of using the variance statistic as a global measure of dispersion is not socially relevant, in other words not pertaining to any direct interest for monitoring economic capability deprivation, it is reasonable to weight all pair-wise deviations $d_{ij}$ the same, ignoring variation in $w_{ij}$ (i.e. $w_{11} = w_{12} = ...w_{nn} = 1$) . An example of an intention not related to assessing social welfare is the academic adjudication of neoclassical theory's hypothesis of income convergence across regions of a country (Barro and Sala-i-Martin, 2003), or of Marxist Geography's prediction of divergence (Massey, 1979). However, if an income inequality study is made with the intent to assess economic capability deprivation in the sense

[2]More technically, consider $i, I$ to be an index and set of communities (say city neighborhoods), $i = \{1, \cdots, n\}$; And, consider $X$ to be a spatial random process and $X_i$ the ability of a community $i$ to convert income to local economic resources. This interdependence is conditional on an unobserved, spatial structure by which the ability of a community to obtain economic resources is linked to its income relative to other ones of society. The spatial structure, $W = [w_{ij}]$, can be thought of as follows:

$$w_{ij} \quad \begin{cases} 1, \text{ if } P(x_i) \neq P(x_i|x_j) \text{ , with } i,j \in I \text{ and } i \neq j \\ 0, \text{ otherwise;} \end{cases}$$

of (Sen, 1997), I argue, that it is questionable to weight all pairs the same, especially in light of the problems of potential unfairness or deprivation that are coupled with large income differentials between different social groups competing for economic resources in the same geographic zone.

The advantage of the pair-wise approach to computing variance in equation (2.4) is the simplicity by which we can use weights to represent spatial structure in a formulation of a inequality measure. Squared income differentials between all $n$ subareas in the population can be represented by a matrix $\mathbf{D} = [d_{ij}]$ (from equation (2.3)). $\mathbf{D}$ can be decomposed into two components, spatial and aspatial:

$$\text{2n(n-1)V} = \underbrace{\mathbf{1}^{'}(\mathbf{W} \odot \mathbf{D})\mathbf{1}}_{\text{spatial}} + \underbrace{\mathbf{1}^{'}(\mathbf{W}^{+} \odot \mathbf{D})\mathbf{1}}_{\text{aspatial}} \tag{2.6}$$

$\mathbf{W} = [w_{ij}]$ represents pairs of subareas $i, j$ using binary weights and its matrix complement (switching the ones for zeros) is $\mathbf{W}^{+}$. $\mathbf{1}$ is a $n$ by 1 vector of ones and $\mathbf{1}^{'}$ its transpose.

The general formulation of this chapter's spatial income inequality measures I define as a function $V(\cdot)$ based on the respective spatial weights matrix, as follows:

$$V(\mathbf{W}) = \frac{\mathbf{1}^{'}(\mathbf{W} \odot \mathbf{D})\mathbf{1}}{n_w} \tag{2.7}$$

$n_w$ is the count of the non-zero, non-diagonal elements of $\mathbf{W}$. The result of equation (2.7) is simply the average size of each income gap represented by non-zero elements of $\mathbf{W}$. The result is also the equivalent to the spatial component of equation (2.6). Given that the average size of all differentials for a set of subareas, $I$, is the same as that for some subset of differentials represented by non-zero elements of $\mathbf{W}$ then the values of the aggregate measure $V$ and the spatial measure $V(\mathbf{W})$ would correspond as $1 = V(\mathbf{W})/2V$, as a result of combining equation (2.5) and equation (2.7).

We use observations of a change such as $V_2(\mathbf{W}) > V_1(\mathbf{W})$ to get a more spatially detailed picture of how income inequality is changing than can be viewed with $V_2 > V_1$, as illustrated below in section 4.

All inequality measures, implicitly or explicitly can be thought of as representing spatial structure. If explicitly represented, this spatial structure can be thought of as reflecting notions of the priority of describing some differentials over others. Any particular spatial perspective taken to discern changes in inequality in one part of society could obscure the existence of changes in a related part of society.[3]

[3]For instance, consider a hypothetical change in a situation where the order of the income levels for three groups of world citizens are as follows: 1st-Americans, 2nd-Chinese, 3rd-Tibetans. If only the Chinese income levels move, narrowing their gap with Americans, then the gap between Chinese and Tibetans would simultaneously widen. If Chinese and Tibetans compete with one another for economic resources in the same geographic zone, but not with Americans, then the movement of Chinese incomes could be said to be economic inequality enhancing, even though, hypothetically, the aggregate income dispersion may have remained the same or gone

Standard additive decomposition approaches can also give a disaggregate picture of changing intra-zonal inequality. Consider a set of subareas that can be partitioned into geographic zones. A two part additive decomposition of total dispersion $T$ of incomes of all subareas would appear as follows:

$$T = B + IN \tag{2.8}$$

$B$ summarizes the cumulative size of differentials *between* geographic zones, assuming the subset of subareas in each zone has the same income. $IN$ summarizes the cumulative size of differentials between subareas that are *within* the same geographic zone, or intra-zonal inequality.[4]

There are two differences between this chapter's approach to look at intra-zonal differentials and that of a standard additive decomposition of variance (discussed in more detail in section 3).[5] First, this chapter's approach is based on a mutually exclusive and exhaustive partitioning of income differentials (or pairs of subareas), not the subarea units themselves (Shorrocks, 1982). This is important so we can

down. Therefore, the link between economic suffering and a geographically aggregate inequality measure, which ignores this notion of interdependence, can be considered opaque (Sen and Foster, 2006).

[4] $W$ is normally used to denote a *within* component, but I use the notation $IN$ so as not to confuse this with the way how I denote a spatial weights matrix using $\mathbf{W}$.

[5] Sen and Foster (2006, p.156-157) explains a difficulty regarding interpretations based on additive decomposition: "As a word of caution, though, we might note that sometimes questions that are plausibly asked may not be sensibly answerable. (For example, 'how much of the breakdown of this marriage was the responsibility of the husband and how much of the wife, adding up exactly to a total responsibility of 100%?')... ."

simultaneously account for geographic and social positions to describe emerging inequalities. Second, this chapter's approach based on equation (2.7), accounts for the average size of a subset of differentials, while a standard decomposition approach, equation (2.8), accounts for the proportion of total dispersion, $T$, linked to different partitions of the distribution of incomes.

Ordinarily, an additive decomposition approach to explain geographic changes in inequality focuses on the movement in the proportion of total dispersion attributable to an inter-zonal partition $B_t/T_t$, equation (2.8). An observation of $B_2/T_2 > B_1/T_1$, could be interpreted as an increase in the importance of the *extent* of the inter-zonal pattern for describing inequality in a society. In this chapter's approach, a similar observation of an upward movement, such as $V_2(\mathbf{W}) > V_1(\mathbf{W})$, could be interpreted as an increase in the *acuteness* of an inequality pattern compared to the general pattern in society. Both could be viewed as complementary ways to provide evidence of increasing institutional or economic policy unfairness towards some parts of society in favor of others.

Kanbur and Zhang (1999) use an additive decomposition approach to examine the changing importance of using rural-urban and coastal-inland zonal divisions to describing inequality in China. For their rural-urban results, they partition China into 56 subareas by dividing each of the 26 provinces into one of two zones, a rural zone or an urban zone. They then decompose total inequality into 3 parts: 1) within-rural differentials, 2) within-urban differentials, and 3) between-urban-rural differentials. They report interesting results showing that, while the urban-rural divide has consistently accounted for more inequality, in recent years there is a growing importance of understanding China's income inequality in terms of a coastal-inland division.

Rey (2004a) introduces an inferential method that can assign statistical significance to descriptive results such as Kanbur and Zhang (1999)'s observation of increasing coastal-inland disparity in China. Instead of focusing on how the levels of a decomposition measure such as $B/T$ change over time, Rey (2004a)'s approach is to examine the statistical significance at each time period for a given partitioning framework that the inter-regional inequality measure $B$ is based on. He defines an index $p(B)$ as:

$$p(B) = Pr(B^* > B^R) \tag{2.9}$$

$Pr$ is used to denote a percentile conception of a probability function. This probability function provides pseudo-significance levels by counting the percentage of times that the observed measure $B^*$ is greater than random realizations $B^R$ calculated from randomly rearranging how each subarea is indexed by a geographic zone. The randomization process can be thought of as shuffling the subareas on a map. The randomly computed realizations, $B^R$, are sorted to build an empirical distribution upon which to assess the significance of the size of an observed value of $B$.

The approach adds informational value to complement the decomposition approach: $B/T$ could decrease while still remaining statistically significant and, vice versa, it could also move upwards while remaining statistically insignificant.

What neither Kanbur and Zhang (1999)'s nor Rey (2004a)'s approaches can account for is whether income differentials are significantly changing between rural and urban people living in proximity to one another, within the same provinces (or zones), as I have termed inter-social, intra-zonal inequalities, the perspective of this chapter.

## 2.3 Formulation

In this section, I propose a framework to formulate a class of inter-social, intra-zonal inequality measures. Instead of partitioning subareas of a country into regional subgroups, as is conventional for an additive-decomposable approach, I partition pairs of subareas. The key of this chapter's approach is to represent how all $N \times (N-1)$ pair-wise income differentials are associated with one another, jointly based on the geographic and social position of each subarea. From this representation I formulate a new class of intra-zonal, inter-social measures.

The measures are based on three types of data matrices: $\mathbf{D}$, $\mathbf{G}$, and $\mathbf{S}_{k,l}$. $\mathbf{D} = [d_{ij}]$ denotes a matrix of the squared differences between all the income values (defined in equation 2.3). $\mathbf{D}$ can be decomposed into two components, spatial and aspatial (equation 2.6). We can decompose the spatial component further with two matrices. $\mathbf{G} = [g_{ij}]$ represents pairs of subareas $i, j$, that are located in the same geographic zone. $\mathbf{S}_{k,l} = [s_{ij}]$ represents pairs of subareas $i, j$ that are referenced by the social positions $k, l$.

For this chapter's empirical illustration I use the common zonal location of two neighborhoods within the same metropolitan zone to specify elements of $\mathbf{G}$, as follows.[6]

$$g_{ij} \quad \begin{cases} 1, \text{ if } i \text{ and } j \text{ are in the same zone} \\ 0, \text{ otherwise.} \end{cases}$$

Likewise, different social matrices can be defined, as follows.

$$s_{ij} \quad \begin{cases} 1, \text{ if } i, j \in \{k, l\} \\ 0, \text{ otherwise} \end{cases}$$

where $\{k, l\}$ is a pair-set for the indices of groups of areas defined by social po-

[6] More generally, other potential studies could define such geographic zones based on different concerns (villages located in the same protected forest, schools in the same school districts, tribes in the same village, etc.)

sitions $\{k,l\} = \{\{1,1\}, \cdots, \{n_p, n_p\}\}$, where $n_p$ is the number of social positions considered. In this chapter's empirical illustration the social positions I use to label neighborhoods of the United States are Black, White, Latino, and Mixed.[7]

Before one can define social similarity $\mathbf{S}_{kl}$ some prior notion of relevant social positions is needed. Each society has its own social positions that are expected to play a role in the economic capabilities of its citizens, such as gender, caste, tribe, culture, and inherited occupation (Rawls, 1971).

Spatial structure can then be simply defined jointly using geographic and social information, as follows:

$$\mathbf{W} = \mathbf{G} \odot \mathbf{S}_{k,l} \tag{2.10}$$

In general terms, $\mathbf{W}$ represents one's prior notion of how the intersection of certain attributes affects how some income differentials are more important than others. Specifically, for this chapter's empirical illustration, $\mathbf{W}$, denotes differentials across areas in the same zone associated with the a particular pair-set of social groups $\{k,l\}$.

[7]In this chapter's empirical illustration, I constructed a class of 10 inter-social, intra-zonal measures. In this case $n_p = 4$ so then the total number of possible combinations of racial pair-sets is 10 given by

$$\binom{n_p}{2} = \frac{n_p!}{2!(n_p-2)!}$$

This alludes to the need to have a small number of social positions when using this chapter's approach so that the interpretation of the results can be tractable.

Then the elements of $\mathbf{W}$ can be understood as:

$$w_{ij} \begin{cases} 1, \text{ if } i, j \in \{k, l\}, \textit{ and} \text{ in the same zone} \\ 0, \text{ otherwise} \end{cases}$$

The notational value of equation (2.10) is in how the spatial weights matrix $\mathbf{W}$ simplifies representing spatial structure defined jointly using labels of social and geographic position.

### 2.3.1 Decomposition

Next, I explain how we can represent a mutually exclusive and exhaustive partitioning of income differentials. $\mathbf{D}$, representing the universe of differentials, can be decomposed into two components representing *intra*-zonal and *inter*-zonal differentials, as follows:

$$\mathbf{D} = \mathbf{D} \odot \mathbf{G} + \mathbf{D} \odot \mathbf{G}^{+} \tag{2.11}$$

The *intra*-zonal component $\mathbf{D} \odot \mathbf{G}$ can be further decomposed into a set of matrices $\mathbf{S}_{k,l}$ representing inter-social relations, as follows:

$$\mathbf{D} \odot \mathbf{G} = \mathbf{D} \odot \mathbf{G} \odot (\mathbf{S}_{11} + \mathbf{S}_{12} + ... + \mathbf{S}_{1k} ... + \mathbf{S}_{n_p, n_p}) \tag{2.12}$$

Using this spatial perspective, we can compare subsets of income differentials in terms of their proportion (P) to total inequality and in terms of their average (A) size. The next two measures below contrast the size of income differentials in terms of social positions to understand intra-zonal income inequality.

(P) The proportional approach is given by

$$P_{k,l} = \frac{\mathbf{1}(\mathbf{D} \odot \mathbf{G} \odot \mathbf{S}_{k,l})\mathbf{1}'}{\mathbf{1}(\mathbf{D} \odot \mathbf{G})\mathbf{1}'} \tag{2.13}$$

(A) The average approach is given by

$$A_{k,l} = \frac{V(\mathbf{G} \odot \mathbf{S}_{k,l})}{V(\mathbf{G})} \tag{2.14}$$

The next two measures below contrast the size of income differentials in terms of the shared geographic positions of pairs of neighborhoods within the same zone to understand inter-social income inequality.

(P) The proportional approach is given by

$$P_{k,l} = \frac{\mathbf{1}(\mathbf{D} \odot \mathbf{G} \odot \mathbf{S}_{k,l})\mathbf{1}'}{\mathbf{1}(\mathbf{D} \odot \mathbf{S})\mathbf{1}'} \tag{2.15}$$

(A) The average approach to be given by

$$A_{k,l} = \frac{V(\mathbf{G} \odot \mathbf{S}_{k,l})}{V(\mathbf{S})} \tag{2.16}$$

The difference between the above pairs of equations, (2.15 and 2.16) versus (2.13 and 2.14), is how they account for geographic and social positions. Equations (2.15) and (2.16) account for the relative importance of geographic positions for characterizing inequality between different social groups. Equations (2.13) and (2.14) account for the relative importance of social positions for characterizing inequality amongst people in the same geographic vicinity. This is seen in how the denominator switches between $\mathbf{S}_{k,l}$ and $\mathbf{G}$, respectively.

In the empirical illustration below, I explore how proportional measures, equations (2.13) and (2.15), and the new average measures, equations (2.14) and (2.16) present two different pictures of changing neighborhood inequality in the United States. Additive decomposition approaches are useful in evaluating the extent of success of an inequality reducing policy in a proportional sense for different parts society (Sen and Foster, 2006). In contrast, the average approach here could be useful in monitoring trends regarding the intensity or acuteness of a subset of income gap changes.

Adopting this chapter's notation, Geary's $c$ statistic appears as follows:

$$c = \frac{V(\mathbf{G})}{2V} \tag{2.17}$$

The average measures of equations (2.14) and (2.16) have a form like that of Geary's c statistic (Cliff and Ord, 1973). In a sense, one could think of this chapter's measures as a translation of Geary's $c$ to analyze inequality instead of spatial clustering. There are two differences between this chapter's new inequality measures and standard applications of Geary's $c$ statistic, or more generally, of other cross-product spatial statistics (Getis, 1991). First, in the new inequality measures both geographic and social positions define the spatial weights matrix $\mathbf{W}$. Second, the new inequality function describes pairwise income differentials, not income levels, between different types of neighborhoods.

The variogram function in geostatistics also describes pairwise differentials. In continuous space, the variogram shows at what distance bands high spatial dependence dissipates into randomness (Cressie, 1991). The proposed metric of this chapter uses weights defined in discrete space.

### 2.3.2 Significance of social and geographic labels

The notion that large inter-social, intra-zonal differentials are substantively important, in the sense that they could affect how economic resources are locally distributed, does not necessarily equate to a particular inter-social, intra-zonal spatial perspective being empirically relevant. To assess the empirical relevance of a $\{k, l\}$ inter-social, intra-zonal measure, there are two types of tests we can use. In both tests the objective is to estimate the likelihood that the new measures are larger in size than similar ones based on random selection from more aggregate subsets of income differentials: (i) $\{k, l\}$ inter-social differentials for the entire society, irrespective of geographic position, or (ii) all intra-zonal differentials, irrespective of social positions. The idea is that these more aggregate subsets of income differentials account for a simpler, more conventional view of neighborhood income inequality to be challenged.

I extend Rey (2004a)'s inferential method (equation 2.12) that assess the significance of inter-regional inequality $B$ to assessing different types of inter-social, intra-zonal inequalities defined by different $\mathbf{W}$'s, as follows:

$$p(\mathbf{W}) = Pr(V(\mathbf{W}^*) > V(\mathbf{W}^R)) \tag{2.18}$$

$\mathbf{W}^*$ is the actual observed spatial structure and $\mathbf{W}^R$ its random realization after randomly permutating the columns and rows of the matrix.

As a way to assess the value of the information in $\mathbf{G}$ conditional on $\mathbf{S}$, or vice versa, when defining spatial structure, $\mathbf{W} = \mathbf{G} \odot \mathbf{S}_{k,l}$, I compute pseudo-significance levels. To test the significance of a social pattern within intra-zonal inequality, I use

$$p(\mathbf{S}) = Pr(V(\mathbf{G}^* \odot \mathbf{S}^*_{k,l}) > V(\mathbf{G}^* \odot \mathbf{S}^R_{k,l})) \tag{2.19}$$

To test the significance of a geographic pattern within a type of $\{k, l\}$ inter-social inequality, I use

$$p(\mathbf{G}) = Pr(V(\mathbf{G}^* \odot \mathbf{S}^*_{k,l}) > V(\mathbf{G}^R \odot \mathbf{S}^*_{k,l})) \tag{2.20}$$

The larger the likelihood that an actual change measurement is greater than random realizations, then the more evidence we have to justify the use of the more complex, spatially explicit measures. Specifically, a frequency close to 50 percent means the social or geographic differentiation hold little value for a specific type of inter-social, intra-zonal inequality, while a frequency closer to 100 or 0 percent conversely signifies the opposite, that social or geographic information is valuable in differentiating between types of income inequality patterns.

This randomization is different than Rey (2004a)'s weights matrix permutations. Each random realization, $V(\mathbf{G}^R \odot \mathbf{S}^*_{k,l})$, is not equivalent to randomly rearranging both the rows and columns of the $\mathbf{W}$ matrix; we are not randomizing income values across tracts. We keep the representation of at least one component of how we structure the observational space, $\mathbf{W}$, fixed. For equation 2.20, the social structure, $\mathbf{S}$, is fixed, while we randomize the geographic zonal labeling scheme within each social group. For equation 2.19, the geographic zonation of the space, $\mathbf{G}$, is fixed, while we randomize the social labeling scheme within each zone.

### 2.3.3 Change

This chapter's approach does not have the favorable property of subgroup consistency. On the other hand, the new measures of the approach may reflect changes in aspects of income inequality that subgroup consistent measure may obscure. Sen and Foster (2006, p.157)'s explanation of subgroup consistency is as follows.

> Subgroup consistency requires that if male inequality rises with female inequality unchanged, then overall inequality must likewise register an increase. More formally, if means and population sizes are unchanged in going from $x_1$ to $x_2$, and from $y_1$ to $y_2$, then [sic. notation is modified to fit that used above in this chapter]:
>
> $IN(x_2) > IN(x_1)$ and $IN(y_2) = IN(y_1)$ entail $T(x_2, y_2) > T(x_1, y_1)$

> Note that this property says nothing about the *size* of the overall increase in inequality relative to the change in the subgroup–it is only a *directional* correspondence. Indeed, the increase in subgroup inequality could be precipitous and the increase in aggregate inequality very small, without violating the precept.

It is this ‘precipitous’ but proportionally small change, which Sen and Foster (2006) mention above, that the inferential approach may be used to identify. To illustrate the contrasting cross-sectional view point of the average approach compared to a proportional approach based on an additive decomposition of variance, consider a partitioning of income differentials within a city into two sets of neighborhoods labeled Immigrant and Native. Say the average income gap between Immigrant and Native pairs of neighborhoods is very large, making the $V(\mathbf{W})$ type measures larger than an aggregate variance statistic based on all areas in the city, such as: $1 < V(\mathbf{W})/2V$. Yet, at the same time, if there are very few Immigrant neighborhoods in the city, then these Immigrant-Native differentials may cumulatively make up a fairly small percentage of total income variance in the city. So although an average measure, $V(\mathbf{W})/2V$, can be relatively big compared to other spatial perspectives (or subsets of differentials), a proportional measure, $B/T$, could simultaneously be fairly small compared to other possible neighborhood partitions of the city.

A descriptive measure of percentage change between two time periods, 1 and 2, is:

$$\frac{V_2(\mathbf{W}) - V_1(\mathbf{W})}{V_1(\mathbf{W})}. \tag{2.21}$$

To understand how to identify a inequality pattern showing significant change, let the change in inequality be given by

$$\Delta A = V_2(\mathbf{W}) - V_1(\mathbf{W}) \tag{2.22}$$

Let random realizations be defined as follows.[8]

$$\Delta A^R = V_2(\mathbf{W}^R) - V_1(\mathbf{W}^R). \tag{2.23}$$

An index measuring the significance of a change in the average size of a subset of inter-social, intra-zonal differentials associated with **W** could then be seen as

[8]In a sense, such random realizations are analogous to how a stochastic $\sigma$-convergence process, explained in Quah (1996a), may also change the relative sizes of different subsets of differentials. The major difference is in the spatial randomizations being conditional on geographic or social structure.

follows.[9]

$$p(\Delta A) = Pr(\Delta A > \Delta A^R) \tag{2.24}$$

Values from this equation describe how concordant the movement of a particular subset of income differentials is with income differentials selected randomly.[10]

## 2.4 Empirical Illustration

### 2.4.1 Data

I provide an empirical illustration of this chapter's approach to describe inter-social, intra-zonal inequality patterns using the income differentials between census tract neighborhoods of the United States for the decadal periods 1980, 1990, and 2000. I use data series at the tract unit level of analysis from the Census Summary Tape File 3A. The key variable for this research is neighborhood income and is based on census per capita median household income series at the tract level, and is summarized in table 2.1. Table 2.2 shows the break down of tracts by social groupings for each year.

[9] When using census data to define how neighborhoods are located within the same zones, different **W** matrices will be needed to be specified for each period since the boundaries of administrative zones are changed by government officials for each period.

[10] We are not testing if change in a measure represents substantive dispersion change, as in studies such as Carree and Klomp (1997) and Lichtenberg (1994), but rather if change in

**Table 2.1:** Mean of Household Median Incomes of tracts by social groupings

| Partition | 1980 | 1990 | 2000 |
|---|---|---|---|
| All groups | 18,566 | 33,953 | 47,462 |
| White | 20,936 | 39,027 | 56,370 |
| Black | 11,240 | 19,491 | 28,106 |
| Latino | 11,618 | 21,047 | 30,296 |
| Mixed | 16,052 | 32,380 | 46,882 |

**Table 2.2:** Census tracts counts by social groupings

| Partition | 1980 | 1990 | 2000 |
|---|---|---|---|
| All groups | 44,171 | 44,764 | 47,007 |
| White | 27,289 | 20,783 | 15,400 |
| Black | 37,87 | 3,977 | 4,149 |
| Latino | 903 | 1,453 | 2,469 |
| Mixed | 12,192 | 18,551 | 24,989 |

First, I labeled each tract as belonging to 1 of 4 census racial categories (Black, White, Latino or Mixed), based on the majority of census survey responses from households in each respective tract. If more than 2 of 3 households in a tract identified themselves as Black or Latino (Non-White Hispanic) or if more than 90 percent of the households in a tract responded as White the tract was labeled accordingly. The remainder of the tracts were labeled Mixed. Social space is differentiated by sorting pairs of tracts to define 10 mutually exclusive and exhaustive inter-social income gap subsets: Black-Black, White-White, Latino-

a measure is significantly different than change based on a random grouping of subareas to construct such a measure.

Latino, Mixed-Mixed, Black-White, Black-Latino, Black-Mixed, White-Latino, White-Mixed, Latino-Mixed.

It is important to note the labeling of census tracts by race was done separately for each decade, in order to avoid comparing "apples with oranges". Census tracts grow, split, and join over the decades, and inter-migration happens among tracts. These changes in the geographic system over time make it impossible to use the same spatial weights for all time periods, since racial labels for tracts across 3 decades would be meaningless. Therefore, the spatial weights are different for each time period, though the rules to define them are constant.

Naturally, if comparisons are to be valid across time, then there must have existed reasonably consistent rules for defining the administrative areas of the geographic system being studied. For instance, the rule of the United States census tracts are "Designed to be relatively homogeneous units with respect to population characteristics, economic status, and living conditions, census tracts average about 4,000 inhabitants."(Census Bureau, 2013).

I labeled each neighborhood as belonging to 1 of 238 metropolitan zones. For 1980, 1990, and 2000, I used a consistent set of metropolitan zones constructed based on the census 2000's metropolitan area county building block definitions. The following criteria were used to select the tracts used in the analysis. To reduce the chance of outliers with extreme income values, tracts needed at least

300 households and have a median household income of 2499 dollars in 1980. To try to consistently compare income differences between urban populations, each tract needed to be within a county defined by the census as being part of a metropolitan area consistently for all three decades. Therefore, information for county units was key to determining how metropolitan zones were defined; To be a metropolitan zone in the analysis, metropolitan areas needed to be composed of counties consistently belonging to a urban areas for all three decades. This chapter's analysis includes as geographic zones less metropolitan areas ($n = 238$) than is officially reported existing in the country ($n = 267$, for 1990 and 2000 censuses). 13 New England metropolitan areas are also excluded because they are aggregated from towns and cities instead of county building blocks. 4 metropolitan areas in Puerto Rico are excluded. The remaining 12 did not meet one of the criteria.

As a reminder, in the chapter, I use the term income gap to refer to the squared difference between the logarithm of median household incomes of a pair of neighborhood tracts. An intra-metro gap describes differentials associated with tracts sharing the same metropolitan zone.

**Table 2.3:** Comparing two descriptions of intra-metropolitan neighborhood inequality: Additive decomposition versus chapter's new approach

Inter-social, intra-metro measures divided by same class of intra-metro measure.
In parentheses are percentages an actual measurement is greater than random realizations.

| | Proportions | | | Averages | | |
|---|---|---|---|---|---|---|
| Income gap subset | 1980 | 1990 | 2000 | 1980 | 1990 | 2000 |
| Intra-metro Black-Black | 1 | 2 | 1 | 0.99 (4) | 1.20 (78) | 0.90 (0) |
| Intra-metro Black-Latino | 0 | 1 | 1 | 0.82 (0) | 1.04 (15) | 0.78 (0) |
| Intra-metro Black-Mixed | 8 | 12 | 15 | 1.14 (100) | 1.43 (100) | 1.32 (100) |
| Intra-metro Black-White | 25 | 19 | 12 | 2.06 (100) | 2.29 (100) | 2.10 (100) |
| Intra-metro Latino-Latino | 0 | 0 | 1 | 0.39 (0) | 0.41 (0) | 0.41 (0) |
| Intra-metro Latino-Mixed | 1 | 3 | 7 | 0.96 (34) | 1.10 (100) | 1.17 (100) |
| Intra-metro Latino-White | 7 | 6 | 5 | 2.35 (100) | 2.43 (100) | 2.49 (100) |
| Intra-metro Mixed-Mixed | 8 | 20 | 30 | 0.71 (0) | 0.70 (0) | 0.80 (0) |
| Intra-metro White-Mixed | 34 | 29 | 25 | 1.07 (100) | 0.94 (0) | 0.98 (0) |
| Intra-metro White-White | 16 | 8 | 4 | 0.50 (0) | 0.48 (0) | 0.44 (0) |

Percentages are based on 1000 realizations from randomly shuffling racial labels.

Percentages close to 0 and 100 indicate significance, close to 50 insignificance.

### 2.4.2 Results

The results in the left-hand side columns in table 2.3 show the proportions of total intra-metro dispersion attributable to different inter-social income differentials calculated from equation 2.15. On the right-hand side columns are average inter-social gap results calculated from equation 2.16. In the parentheses are the percentage of times these actual measures were greater than random realizations, based on shuffling social labels, using equation 2.19. The table illustrates how the new approach adds information to the picture of changing neighborhood inequality given by an additive decomposition approach using the same spatial perspectives. The main empirical difference is that the additive decomposition highlights the increasing size of the proportions of inequality attributable to differentials associated with Mixed neighborhoods, increasing from 51 percent in 1980 to 77 percent in 2000. But this decomposition picture hides how the most acute intra-metro income differentials are associated with Latino-White neighborhoods. Looking down the rows of table 2.3 we see that the average Latino-White gap is the highest for all three decades. In contrast, from a proportional perspective it does not appear substantial, and it moves in the opposite direction.

The results in table 2.4 compare the average size of a inter-social, intra-metro gap to the average of the corresponding aspatial inter-social gap, calculated using equation 2.14. By aspatial I refer to those subsets of income differentials chosen

**Table 2.4:** Geographic inequality patterns: Average inter-social, intra-metro gap divided by average inter-social gap, irrespective of geography

In parenthesis is percentage actual measurement is greater than random realization.

| Income gap subset | 1980 | 1990 | 2000 |
|---|---|---|---|
| Intra-metro Black-Black | 1.03 (70) | 0.88 (2) | 0.97 (34) |
| Intra-metro Black-Latino | 1.07 (82) | 0.99 (47) | 1.07 (94) |
| Intra-metro Black-Mixed | 0.93 (2) | 0.77 (0) | 0.90 (0) |
| Intra-metro Black-White | 1.11 (100) | 0.97 (17) | 1.09 (100) |
| Intra-metro Latino-Latino | 0.71 (1) | 0.65 (0) | 0.88 (7) |
| Intra-metro Latino-Mixed | 1.08 (87) | 0.96 (20) | 1.18 (100) |
| Intra-metro Latino-White | 1.66 (100) | 1.63 (100) | 1.88 (100) |
| Intra-metro Mixed-Mixed | 0.92 (1) | 0.74 (0) | 0.95 (1) |
| Intra-metro White-Mixed | 1.23 (100) | 1.05 (100) | 1.22 (100) |
| Intra-metro White-White | 0.90 (0) | 0.75 (0) | 0.77 (0) |

Percentages are based on 1000 realizations from randomly shuffling metro labels.
Percentages close to 0 and 100 indicate significance, close to 50 insignificance.

regardless of whether they belong to the same geographic zone or not. In parentheses is the percentage of times an actual measure was greater than its random realizations, calculated using equation 2.20. Comparing table 2.4 to table 2.3 shows that the social positions of neighborhoods is more important in describing intra-metro inequality than their geographic positions are in describing inter-racial inequality, as might be expected. Nevertheless, for describing inter-social inequality, intra-zonal structure does have significance for most social pairs. This is most evident with how much larger the Latino-White inequality is for neighborhoods located in the same metropolitan areas compared to the universe of Latino-White differentials. This result shows how using just a social perspective, ignoring the geographic one, underestimates the size of the income disparity between Latino and White neighborhoods in the same cities.

The results in table 2.5, calculated based on equation 2.21, and describe the percentage change in the average income differentials over the decadal periods for different subsets of differentials. Looking down the rows illustrates how differently changes have occurred based on the spatial perspective taken, ordered as sets of types of measures from most to least aggregate. The clearest example of discordance is seen in how half the intra-metro measures (bottom rows) for the decadal change between 1990 and 2000 move down and half moved up. Between 1980 and 1990 it appears neighborhood inequality consistently rose from almost all the measures, but was greatest for differentials associated with Black neighborhoods.

As an example of the value of jointly accounting for social and geographic positions, the inequality *reducing* change in the average income gap associated with pairs of White-Mixed neighborhoods located in the same metropolitan neighborhoods is masked by contemporaneous inequality *enhancing* change of both the pure intra-metropolitan and the aspatial White-Mixed measures. Likewise, the results show that the inequality *reducing* change associated with Black neighborhoods between 1990 and 2000 is masked by the intra-metropolitan measure. In addition, the inequality *enhancing* change associated with Latino neighborhoods

**Table 2.5:** Evolving Neighborhood Income Inequality from Different Spatial Perspectives: Percentage change in the average size of a subset of income differentials

| Income gap subset | 1980-1990 | 1990-2000 | 1980-2000 |
|---|---|---|---|
| Universe of Neighborhoods | 34 | -7 | 24 |
| Intra-metro, all social pairs | 10 | 6 | 17 |
| Inter-social, irrespective of geographic position | | | |
| Black-Black | 55 | -28 | 11 |
| Black-Latino | 50 | -27 | 10 |
| Black-Mixed | 66 | -16 | 39 |
| Black-White | 39 | -14 | 19 |
| Latino-Latino | 27 | -22 | 0 |
| Latino-Mixed | 42 | -9 | 29 |
| Latino-White | 16 | -7 | 8 |
| Mixed-Mixed | 34 | -8 | 24 |
| White-Mixed | 14 | -6 | 7 |
| White-White | 29 | -5 | 22 |
| Inter-social, intra-zonal | | | |
| Intra-metro Black-Black | 33 | -21 | 5 |
| Intra-metro Black-Latino | 39 | -21 | 10 |
| Intra-metro Black-Mixed | 38 | -3 | 34 |
| Intra-metro Black-White | 22 | -3 | 18 |
| Intra-metro Latino-Latino | 15 | 6 | 23 |
| Intra-metro Latino-Mixed | 26 | 12 | 40 |
| Intra-metro Latino-White | 14 | 8 | 22 |
| Intra-metro Mixed-Mixed | 8 | 19 | 29 |
| Intra-metro White-Mixed | -3 | 9 | 6 |
| Intra-metro White-White | 7 | -3 | 4 |

Inequality change over time as a percentage of the previous period's value.

**Table 2.6:** Changing geographic patterns of neighborhood inequality

Percentage of times change in measure is greater than change in random realization.

| Income gap subset | 1980-1990 | 1990-2000 | 1980-2000 |
|---|---|---|---|
| Intra-metro Black-Black | 3 | 89 | 25 |
| Intra-metro Black-Latino | 20 | 86 | 52 |
| Intra-metro Black-Mixed | 0 | 100 | 24 |
| Intra-metro Black-White | 0 | 99 | 41 |
| Intra-metro Latino-Latino | 25 | 99 | 89 |
| Intra-metro Latino-Mixed | 9 | 100 | 86 |
| Intra-metro Latino-White | 41 | 99 | 95 |
| Intra-metro Mixed-Mixed | 0 | 100 | 82 |
| Intra-metro White-Mixed | 0 | 100 | 31 |
| Intra-metro White-White | 0 | 73 | 0 |

Percentages are based on 1000 realizations from randomly shuffling metro labels.

Percentages close to 0 and 100 indicate significance, close to 50 insignificance.

during the same period is masked by the corresponding aspatial, inter-racial measures (except the Black-Latino pairing).

Based on equation 2.24, tables 2.6 and 2.7 list pseudo-significance levels for the inter-social, intra-zonal change measures using random shuffling of geographic and social labels, respectively. Table 2.6 lists the percentage of times change in the actually observed inter-social, intra-zonal measures are greater than randomly simulated intra-zonal realizations. These values give us a picture of the relevance of intra-zonal patterns in changing inequality. For both decadal periods there exist significant differences in change, but in opposite manners. Between 1980 and 1990 intra-metro inequality changed to a lesser degree than the rest of the country for all pairs, and significantly so for about half the pairs. Between 1990-2000 the

**Table 2.7:** Changing social patterns of neighborhood inequality

Percentage of times change in measure is greater than change in random realization.

| Income gap subset | 1980-1990 | 1990-2000 | 1980-2000 |
|---|---|---|---|
| Intra-metro Black-Black | 97 | 0 | 0 |
| Intra-metro Black-Latino | 100 | 0 | 5 |
| Intra-metro Black-Mixed | 100 | 0 | 100 |
| Intra-metro Black-White | 69 | 38 | 58 |
| Intra-metro Latino-Latino | 82 | 16 | 58 |
| Intra-metro Latino-Mixed | 100 | 21 | 100 |
| Intra-metro Latino-White | 82 | 30 | 68 |
| Intra-metro Mixed-Mixed | 96 | 100 | 100 |
| Intra-metro White-Mixed | 00 | 100 | 0 |
| Intra-metro White-White | 00 | 75 | 0 |

Percentages are based on 1000 realizations from randomly shuffling racial labels.
Percentages close to 0 and 100 indicate significance, close to 50 insignificance.

pattern of change was the opposite, intra-metro neighborhood inequality rose to a higher degree than the average for aspatial inter-social differentials.

In table 2.7 we see that although the intra-metro, inter-social inequality movements appear concordant in table 2.5, the social labels I use still hold informational value in explaining variations in change.

Below I summarize some key results.

- Between 1980 and 1990, as an example of the value of jointly (or simultaneously) accounting for social and geographic positions, the inequality *reducing* change in the average income gap associated with pairs of intra-metro White-Mixed neighborhoods is masked by inequality *enhancing* change of both the socially aggregate intra-metropolitan measure and the corresponding aspatial White-Mixed measure. Also between 1990 and 2000, Latino-White income differentials within the same metro zones increased 8 percent, but simultaneously went down -7 percent for the subset of all Latino-White pairings, irrespective of metro zone.

- Between 1980 and 1990, Black neighborhoods were associated with the largest inequality enhancing changes. This increase was significantly greater for aspatial differentials than for intra-metro differentials.

- In reversal, between 1990 and 2000, Black neighborhoods were associated with the largest inequality reducing changes.

- As an example of discordance, changes between 1990 and 2000 in intra-metro inequality were discordant across social pairs; 5 measures moved up and 5 down.

- Between 1990 and 2000, Mixed neighborhoods were associated with the largest inequality enhancing changes.

- Significant for all decades, the strongest inequality pattern is associated with intra-metro White-Latino neighborhoods. An additive decomposition of variance masks this result. The results also show that spatial or intra-metro White-Latino neighborhood inequality is significantly greater than aspatial White-Latino inequality.

### Sociology literature

There exist in the sociological literature two important and competing explanations of the poor economic performance of inner city Black neighborhoods during the 1980's that are linked to two different mechanisms of social isolation of poor central city Black areas.[11] For Massey and Denton (1993), segregation is primarily an effect caused by unfair institutional policies in society towards Blacks. For Wilson (1987b), segregation occurs not so much from exogenous forces, but as part of a larger complex social and economic process where the important social split to look at is not just Black and White, but Black Middle class and Black Poor. Wilson (1987b) stresses social isolation of poorer urban Blacks is based in large part by the out-migration of middle class Blacks. Though the results here do not provide evidence favoring one theory over the other, they do describe a

[11] For neighborhood units of analysis, Blau (1977, p.168-69)'s work emphasizes the importance of looking at the intersection of social and geographic space as follows.

> "Despite the fact that most social associations occur within communities, group and status barriers to social associations within communities inhibit them less that such barriers among communities. The apparent paradox results from the influences of the spatial distribution on social associations. Group and status barriers to social associations within communities are attenuated by the counteracting effect of physical propinquity, whereas group and status differences among communities constitute barriers to social associations that are reinforced by spatial segregation".

Blau (1977)'s understanding of space and inequality helps understand the relevance of the inter-social, intra-metro spatial perspective.

situation of increased neighborhood income inequality in the 1980's that is empirically consistent with both Wilson (1987b) and Massey and Denton (1993) of increased Black-Black and Black-White neighborhood income differentials.

The result of increased inequality associated with intra-metro Black neighborhoods is also consistent with Jarkowsky (1997)'s results.[12] Jarkowsky (1997) interprets his results as support for Wilson (1987a)'s hypothesis of the sociological importance of the out-migration of Black middle class families to suburbs from central city Black neighborhoods in the 70's and 80's as a factor to their increased poverty. This chapter's results also show that the rise in intra-metro Black inequality was significantly less than the rise in aspatial Black inequality. This result is consistent with how Jarkowsky (1997) descriptively emphasized that the rise in poverty amongst urban Black neighborhoods in the 80's was not broad based but concentrated in particular cities, such as old industry cities in the Midwest. What is left uncertain is how we explain the partial reversal: Between 1990 and 2000 the largest inequality reducing changes are associated the intra-metro and

[12]Jarkowsky (1997)'s results are based on what he terms a Neighborhood Sorting Index. It is calculated for each city separately. It uses the ratio of two variance statistics of income in a city: one of per capita median incomes across census tracts, the other of the income of households. The Index is intended to capture the degree of economic segregation of households by census tracts within a city. Though his Index accounts for intra-metro income differentials between the same races it does not account for differentials across races. Another difference between this chapter's results and Jarkowsky (1997)'s is that this chapter's approach is geographically global in the sense that the measures account for subareas across a broad range of geographic zones and his measures are based on subareas in an individual city. Further, this chapter's approach has a probabilistic basis to test for significance.

aspatial Black neighborhood differentials. This chapter's approach complements the results of Jarkowsky (1997) by providing results in the context of other inter-racial changes, and by having an inferential basis to compare intra-metro versus aspatial changes.

## 2.5 Concluding Remarks

In summary, this chapter presents an empirical approach to assess economic inequality that accounts for the average size of the income differentials between different socially and geographically defined subsets of neighborhoods. Substantively, I argue that it is important to assess changes in inequality between different social groups that are located in the same geographic zones, because changes in such differentials may have distinct consequences on a community's capability to obtain certain locally bound economic resources, such as health care or education.

The inferential approach for testing the informational value of incorporating spatial information (geographic and social references) is an extension of Rey (2004a)'s spatial randomization method. Empirically, the new approach presented here can identify a significant pattern of change in the relative incomes between two social groups occurring contemporaneously across many geographic zones of a society. The empirical illustration is based on income differentials between over 44,000 neighborhoods of the United States, referenced by 4 social groups (Black, White, Latino, and Mixed) and 238 metropolitan zones, for the 1980, 1990, and 2000 decadal census periods. The results illustrate how assessments of neighborhood income inequality in the United States are sensitive to the spatial perspective we use to formulate the inequality measures in terms of inter-racial and intra-metropolitan groupings. The key advantage of the new exploratory approach is that it may highlight in what parts of society new economic inequalities may be emerging, which might otherwise be obscured by more aggregate approaches.

A potential avenue for future research is to define spatial weights in continuous space rather than using discrete categories. For example, geographically, subsets of health or crime differentials could be defined using distance from public facilities. Socially, subsets of differentials can be defined in continuous space in terms of observations on communication networks between places.

A disadvantage of this type of quantitative analysis is that the most important types of observations relating to economic inequality may not have been recorded in published statistical datasets that academic researchers are constrained to. For instance, in Mexico and France there cannot be information distributed on racially labeled social positions, since the concept officially cannot be used for differentiating between its citizens. Another disadvantage is that in order to represent intersectional notions of inequality is that we are limited to using only intersectional notions that can be tractably translated into a computational function and can be based on available data. I use fairly crude functions based on census racial and metropolitan information to label each neighborhood's social and geographic position. The binning of neighborhoods into social positions based on racial attributes is based on conventional ideas of historical context rather than scientific reasoning. Due to these disadvantages, this chapter's approach should be viewed as exploratory, hypothesis seeking, not a confirmatory, hypothesis proving (Rey and Janikas, 2005; Tukey, 1977).

The significance of societal racial categories for describing income inequality within cities of the United States is obviously not new. The types of quantitative measurements presented in this chapter, however, might be useful in informing discussions assessing economic inequality in other societies of the world where the relevant social and geographic positions are less clear for analysts (though perhaps obvious for lay people). If for the United States changing patterns of inter-racial, intra-metro income inequality are masked by aggregate measures, then this raises the question as to how similar inter-social, intra-zonal patterns in other societies of the world are being obscured.

For instance, measuring world citizen income inequality, some academics have recently argued that the painful consequences of recent economic changes experienced by people around the globe (which some believe are linked to new international trade arrangements termed 'globalization') are exaggerated. They use as evidence a downward change in the level of their aggregate dispersion metric. The concern here is that such interpretations, based on the direction of change in the level of income dispersion across individuals of a large geographic zone, such as a country (Quah, 2002) or the world (Sala-i-Martin, 2006), mask disaggregate patterns of inequality.

In a magazine article (Barro, 2002) titled “The U.N. is dead wrong on poverty and inequality”, the economist Robert Barro argues that the assessment by the U.N. in the 1999 Human Development Report contradicts the facts. He quotes the U.N. report: “...The past decade has shown increasing concentration of income, resources, and wealth.” He explains: “For world inequality, we can think of the changes in two parts. The first is within countries, the second is across countries.” Citing Sala-i-Martin (2006) as evidence, he says that over the past 30 years the intra-country inequality rise has been ‘small’ compared to the inter-country inequality reducing changes. This chapter’s results for the United States suggest a need, especially considering the context of violence and inequality, to picture changing world income inequality not in terms of socially homogeneous countries, but in terms of socially heterogeneous societies.

The most problematic consequences of economic changes, such as globalization, may not only be related to an increase or decrease in overall aggregate levels of inequality across countries or individuals, but to new inequality patterns between social groups living side by side. For example, prominent in the sociological literature, Wilson (1987a) documents some of the mechanisms generating unequal consequences of industrial restructuring in the economy of the United States in the 70's and 80's, focusing on the harsh absolute economic disadvantages that fell upon urban Black neighborhoods of Chicago caused by changes in the relative incomes of those communities. Aggregate measures of world citizen inequality may be obscuring ongoing similar consequences that might be currently occurring in other societies as an effect of analogous economic restructuring brought by globalization.

As a conjecture, such obscurations may be especially pernicious when researchers base their pronouncements of the direction of change in inequality on empirical analysis of societies that are foreign to them; for such societies, it is more difficult for researchers to appreciate the importance of accounting for variation in historically significant social positions that might be masked by conventional geographic aggregation of observations.

# Chapter 3

# Assessing income inequality trends across regions of a country

## 3.1 Introduction

Assessments of changes in the distribution of income across regions of a political system can provide information for appraising those institutions that regulate the conflicting economic interests of different geopolitical units of a political system. Or, in other words, from a Rawlsian (Rawls and Kelly, 2001) perspective, by looking at whether income dispersion moves up or down we are trying to understand the fairness of one aspect of the political economic system.[1] For this chapter's empirical illustration the units of analysis are states of the United States. This chapter examines how ignoring spatial aspects of the data may obscure our judgments of fairness in terms of regional income inequality.

[1]Rawls (1999, p.229) says "From the beginning I have stressed that justice as fairness applies to the basic structure of society. It is a conception for ranking social forms viewed as closed systems. Some decisions concerning these background arrangements is fundamental and cannot be avoided. In fact, the cumulative effect of social and economic legislation is to specify the basic structure".

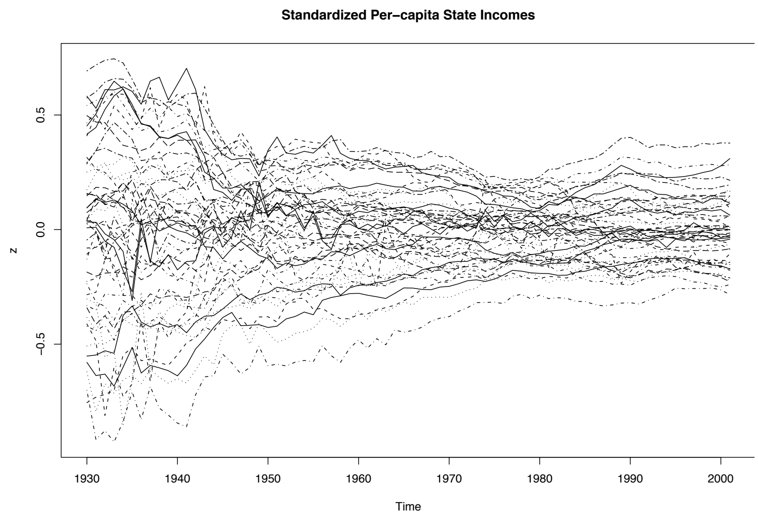
To examine how spatial aspects of the data complicate matters, I incorporate a spatial weights matrix into a convergence data generating process. The idea here is to model how the distribution of state incomes can be changing both from an aspatial catch-up convergence component and a random spatial error component. The results of this spatial error model are then used to filter out the spatial component from a time-series of income variance. The results of this chapter show that the observed change in the dispersion trend of state incomes, from slow convergence to divergence, in the 1980's may not have necessarily been caused by a large change in the underlying convergence process, but rather could plausibly have been caused from a geographically patterned economic disturbance.

In addition to this introductory section, this chapter has four main sections. The first, section 3.2, provides background on the empirical assessment of regional income dispersion. Then in section 3.3, I discuss the normative relevance of assessing income convergence amongst regions of a country. Section 3.4 describes a method for incorporating a spatial weights matrix into assessments and section 3.5 presents an empirical result.

## 3.2 Background

$Y_{it}$ is the income per-capita of a region $i$ at time $t$. We transform this series as such $y_{it} = ln(Y_{it}) - mean(ln(Y_t))$. Figure 3.1 shows a plot of this series for the

**Figure 3.1:** Income per-capita convergence for 48 U.S. states: 1929 to 2000




48 contiguous states of the United States. Visual inspection of the plot reveals an overall downward decline in dispersion. If $\sigma_t^2$ is the variance of the cross-sectional productivity observations $y_{it}$ then we say that there is positive evidence of a $\sigma$-convergenceprocess existing when we measure $\sigma_t^2 < \sigma_{t-k}^2$, where $k$ is some arbitrary time span of years. This $\sigma$-convergence measure is an aggregate way to quantify changes in dispersion seen in figure 3.1. After World War II, matching the pattern of $\sigma$-convergence for US states, strong $\sigma$-convergence trends have been found for the administrative regions of Japan and Europe (Barro and Sala-i-Martin, 2004b).[2]

[2]Though this study uses variance, others statistics are used for $\sigma$-convergence such as the Gini coefficient, the standard deviation, and the coefficient of variation. All are aggregate ways to characterize the inequality dynamics of a set of regions, and each has advantages and disadvantages (Sen and Foster, 2006, see Chap. 1). Yamamoto (2008)'s plot of $\sigma$-convergence for the 50 US states using each statistic shows little variation in how they change over time,

Ordinarily, researchers plot some income dispersion measure over time spans of five to ten years to visually identify trends in $\sigma$-convergence. These plots suggest to viewers how recent trends may be expected to evolve into the future, providing information on the structure of an economy. One problem this research addresses is how to disentangle random income fluctuations from a real economic structural change, a secular trend. Although $\sigma$-convergence is mainly used as a descriptive measure, researchers may be over confident, I argue, in how they interpret their observations of a sudden change in a long run dispersion trend.

As an example of how researchers characterize change in a trend, Drennan and Lobo (1999, p.358), describing a plot of the coefficient of variation for income across metropolitan areas of the US from the mid-1970's until 1989, say they note a "decided upward trend...which suggests divergence".[3] Kanbur and Zhang (1999); Chen and Fleisher (1996); Jian et al. (1996, p.692,147,5) do this for provinces in China; Azzoni (2001, p.142,143) does this for states in Brazil; Drennan and Lobo (1999, p.358) for metropolitan areas in the USA; Cuadrado-Roura et al. (1999, p.40) for the NUTS Level II regions of Spain; Gezici and Hewings (2004, p.124) for provinces of Turkey, Coughlin and Mandelbaum (1988, p.25) for states of the

in their overall trends, which is the main focus of this study. The advantage of the variance measure of $\sigma$-convergence is its direct link to the concept of $\beta$-convergence.

[3]The authors follow-up with another article (Drennan et al., 2005) where they use a unit root time-series test. This test will not be applied for the proposed research mainly the decadal census data that will be analyzed does not have enough observation, degrees of freedom, required for the time-series test.

USA; Tortosa-Ausina et al. (2005) for provinces in Spain, and Nissan and Carter (1999) look at the gap between non-metropolitan and metropolitan areas in the US.

Historical trends in $\sigma$-convergence are also interpreted in terms of policy changes. Sánchez-Reaza and Rodríguez-Pose (2002) associate the open trade policies in Mexico of GATT, during the 1980's, and NAFTA, during the 1990's, to increased regional disparity trends. Prior to theoretically specified regression analysis, Rivas (2007) uses movements in the Gini coefficient to support the evidence of increased regional inequality during the 1990's due to trade liberalization in Mexico.

Fan and Casetti (1997) use the aggregate increase in the income dispersion across states of the United States in the 1980's as evidence to make the strong pronouncement that the ideas behind the Neoclassical economic convergence hypothesis are obsolete. Their pronouncement is strong because it can be interpreted in a normative manner: that the expected fairness associated with the structure of the political economic system towards initially poor states of the United States has changed and should be put into question. I will discuss this idea in the next section.

## 3.3 Relevance

In this section, I discuss how the regional convergence hypothesis can be linked to the idea of normatively assessing unfairness. In particular, the focus here is on how empirical matters become complicated because of how spatial patterns can characterize how an income distribution evolves.

Some Marxist geographers have theorized that the profit interests of the owners of capital are in line with preventing the convergence of poorer regions to richer ones, (Massey, 1979). More simply, Marxist believe that the basic structure is unfair towards regions that are poor at the start.[4] Others emphasize how spatial agglomeration forces (Krugman and Venables, 1995) may concentrate the most productive high income earning activities into a few geographic market areas, at the expense of peripheral regions leading to regional income divergence. Most influential in the economic literature is Neoclassical theory's prediction of regional income convergence (Barro and Sala-i-Martin, 2004b).

As a conceptual aid, in order to represent changing patterns in the income gaps amongst people living in pairs of geographic areas, $i$ and $j$, I borrow Bernard and Durlauf (1996, p.165)'s proposed two definitions of hypothetical ways the incomes of people in different places can be thought of converging. They state that con-

[4]This discussion presumes we are dealing with a country characterized by competitive markets.

vergence between members of a set of $I$ economies may be defined analogously by requiring that every pair within the set exhibits the defined types of convergence. Below, $\Upsilon$ denotes all the information at time $t$. Their first definition (in reverse order from their paper) says countries $i$ and $j$ converge if the long-term forecasts of (log) per capita income, $y$, for both countries are equal at a fixed time $t$.

*Definition 1: Convergence as closing the income disparity gap.*

$$\lim_{T \to \infty} E(y_{i,t+T} - y_{j,t+T} | \Upsilon) = 0 \tag{3.1}$$

The $\sigma$-convergence measure is used as an aggregate, exploratory way to quantify how a geographic income distribution changes to see how it matches Definition 1. The mechanism by which convergence according to Definition 1 will occur is by poorer regions of the same country, which presumes they all have the same steady-state productivity potentials, growing faster than their richer counterparts. In other words, they will converge by catching up as defined as follows.

*Definition 2: Convergence as catching up. If $y_{i,t} > y_{j,t}$ then,*

$$(y_{i,t+T} - y_{j,t+T} | \Upsilon) < y_{i,t} - y_{j,t} \tag{3.2}$$

Presenting a catch-up mechanism, Definition 2 is based on the assumption of diminishing returns to capital , which lowers the relative growth of a richer region who is close to its steady-state productivity as compared to a capital poor region (Solow, 1956).

The prediction associated with Definition 2 can be viewed as a test of the *validity* of Neoclassical theory's characterization that poor economies within a country have a natural tendency to catch-up to richer economies. Across a set of isolated economies, the mechanism for the catch-up process is based on the assumption of diminishing returns to capital of neoclassical economic theory.[5] Across the set of economies the prediction assumes geographically invariant population growth rates, investment rates, and consumption preferences, which is often assumed to be more probable for places within a country than across different countries. Also for places within a country more than across countries the prediction is strengthened by greater interregional flows of capital and labor due to commerce laws, as well as greater flows of knowledge due to language similarity and lower communication costs. The theory says capital investments will gravitate over time to poorer areas that lack capital because of the market incentive caused by higher returns in these areas.

[5]Originally, this theory was mathematically formalized by Solow (1956) for variations over time for one economy.

The empirical evidence associated with Definition 1 can be viewed as a test of the *relevance* of Neoclassical theory's characterization of the economy. Even when Definition 2 is met, given that random shocks can also alter the distribution of regional income levels, there is no guarantee of absolute convergence according to Definition 1.[6]

From the economist's perspective the proclaimed reason for assessing regional income distributions is not for appraising institutional fairness, but rather for adjudicating the convergence hypothesis. The economist's convergence hypothesis, however, might also be thought of as an empirical tool for the assessing the temporal aspect of Rawls (1999, p.53)'s Second or Difference Principle: "Second: social and economic inequalities are to be arranged so that they are both reasonably expected to be to everyone's advantage...". A catch-up convergence process matches Rawls (1999, p.78)'s reasoning of a situation of just inequalities in that "Eventually the resulting material benefits spread throughout the system to the least advantaged." In Rawls (1999)'s context, he is referring to social classes, but here I am referring to geographic areas. If regions that are initially poor are expected to catch-up to richer regions, and gaps between all regions shrink, then

[6] In a sense, Definition 1 relates to time and Definition 2 relates to space. Definition 1 can be thought of as a way to characterize the consequences of a convergence process that is consistently stable over time, where unexplained disturbances or shocks cancel themselves out quickly enough to infer that decreasing income disparities will naturally evolve into the future. And Definition 2 can be thought of as a way to characterize convergence being stable over space with no polarization or groupings in the process occurring due to unsymmetrical spatial structure between economic relations between geographic areas.

one might say this is evidence that the basic structure of the political economy is fair in terms of regional income disparity. The normative relevance of a justification of institutional structures based on the ideas of Neoclassical theory is that they equate to a form of procedural justice for regional economies; In other words, free market policies will automatically lead to the egalitarian goal of income convergence across regions.

In understanding the relevance of empirical methods to measure economic inequality, the concern of theoretical adjudication is intertwined with Rawls and Kelly (2001)'s concern regarding injustice in term of unfair institutional arrangements. For instance, if people are confident in the empirical evidence supporting theoretical predictions of Definition 1, of absolute or $\sigma$-convergence, then this weakens social justice arguments in favor of the geographical targeting of development policies to raise human welfare. Likewise, while monitoring regional income inequalities, an observation of a sudden change towards more income inequality that is spatially patterned, not broad based, may not necessarily indicate underlying social injustices stemming from regional economic policies, but can still be used to forewarn us of potential negative human welfare consequences that may follow.

An ambiguous social justice question is raised by the fact that Definition 2 of catch-up convergence is not a sufficient condition to reach Definition 1 of absolute convergence. The fact that these definitions are not equivalent has raised objections to using Definition 2 as the primary empirical evidence of the Neoclassical convergence hypothesis (Quah, 1996a), believing that Definition 1 represents a more relevant concept of convergence (Friedman, 1992).

For Sala-i-Martin (1996), empirical evidence of Definition 2 without Definition 1 could still be interpreted normatively as a positive outcome. Sala-i-Martin (1996) defends the relevance of $\beta$-convergence assessments, of Definition 2, on normative grounds that the idea of catch-up convergence is linked to fairness. He uses the analogy of a hypothetical situation where there are basketball teams amongst whose overall performance dispersion may never lessen, absolute or $sigma$-convergence of Definition 1, but at the same time, due to a fair draft regulating structure, weaker teams are expected show catch-up or $beta$-convergence by reducing their performance gap with initially stronger teams over each time span.[7]

Complicating the analysis of empirical evidence used to judge competing theories are methodological problems inherent with the use of geographic units of observation. What Sala-i-Martin (1996) ignores is that basketball teams can be assumed to be independent of one another while this assumption is harder to make for regional economies. At the regional scale, examples against an aspatial, broad

[7]Sala-i-Martin (1996) explains this in a footnote.

based characterization of income gap movements are Quah (1996b)'s results of an evolving bi-polar distribution of European regions and Rey (2002)'s results of spatially conditioned inter-mixing of relative ranks for US states. These results have important normative relevance. They allude to how evidence of a catch-up convergence process among states or regions may obscure the possibility that the income gaps between larger geographic clusters of interdependent regions may not be converging even when there is evidence of $\beta$-convergence amongst what appear to be independent areas or regions.[8] The catch-up process of Definition 2 may not be spatially invariant. Some of the mechanisms of economic convergence are flows of investments, of people, and of innovative technologies between rich and poor countries (Lucas, 2002). For regions of a country, these types of flows may be regulated by geographic distance.

Spatial dependence occurs when observations are not independent of one another, as usually assumed in standard statistical analysis (Anselin, 2003). For instance, changes over time in the income levels of nearby communities may be closely interrelated across space, so that over small time periods random geographically clustered economic shocks may obfuscate the monitoring of changes in inequality measures associated with substantive changes in the economic climate

[8]For example, there could be three large geographic clusters of regions ranked as low, middle, and high. Within each there can be catching-up convergence, but not between each. Nevertheless, an aggregate, aspatial assessment may characterize the situation as meeting Definition 2, even though the situation does not seem to imply a fair economic structure for the initially poor clusters of regions.

of growth. In this chapter, I propose a method to deal with spatial autocorrelation of disturbances when testing for $\sigma$-convergence, Definition 1.[9]

## 3.4 Spatial Weights

For this research, in an effort to use an empirical measure of change in income inequality as a signal of change in economic inequality, I am concerned with differentiating between the role of random economic disturbances, or noise, and the role of an underlying $\beta$-convergence process. The key departure of this research from the literature is that I do not assume that economic disturbances are independent across subareas or regions of a country. In other words, a spatial process is believed to be generating the disturbances where by each region's expected income level is affected by the random economic shocks of its neighbors. How such chance events that could be geographically clustered might include natural disasters or periodic regional demand recessions.

[9] Another measurement problem in human geography, referred to as the Modifiable Areal Unit Problem (MAUP), is that generalizations on the distributional characteristics from observations at one spatial scale and zonal scheme cannot be carried over to other spatial scales and zonal schemes (Oppenshaw and Taylor, 1991). In reference to Definitions 1 and 2 above the challenge is to organize the set of $I$ economies in order that the definitions are most relevant to the concern at hand. For instance, a positive pronouncement of greater inequality amongst states of the US may potentially hide a contemporaneous negative change of lesser inequality amongst units based on another grouping such as rural versus urban areas or suburbs versus inner city neighborhoods. Addressing issues of the MAUP, Rey (2004b) and Yamamoto (2008) both plot inequality trends using different aggregation schemes. They also look at the portion of global inequality that is accountable by intra-regional versus inter-regional inequality, alluding to the way different income convergence mechanisms play themselves out at different spatial scales. What is evident from these studies is that each aggregation scheme presents us with a different cross-sectional view of inequality.

Rey and Montouri (1999) also discuss this in terms of inefficient regression estimates. They find that the idea of structural change in the speed of catch-up convergence for states of the US, based on different regression estimates of the strength of a $\beta$-convergence process for the first half and second half epochs from 1929 till 2005, may just be the effect of spatial error dependence. In contrast to Rey and Montouri (1999), this research is not looking at spatial effects in terms of how we estimate the over-all speed of a catch-up or $\beta$-convergence process, but rather it looks at spatial effects in terms of how they may obscure judgments in terms of how we impute a change in the structure of the economy due to a change at one time span in the trend of aggregate dispersion.

Rey and Dev (2006a) also look at the role of spatial effects on $\sigma$-convergence, but as a component of total variance for different slices of time, rather than the change in variance between two time periods, the aim here. As another attempt to control for common regional disturbances on convergence descriptions, Barro and Sala-i-Martin (2004b) calculate what they term a shock variable based on how each state's disturbance is a function of its industrial structure (sector share) and specify it within $\beta$-regressions over 10 year time spans. In time periods and for geographic areas where industry sector data is not available the filtering method proposed below could be an alternative approach to control for geographically common shocks.

### 3.4.1 Data Generating Process

Consider a process that characterizes the dynamics of economic productivities for a set of regions and is approximated by:

$$y_{it} = by_{i,t-1} + \nu_{it}, t = 1, \cdots, T, i = 1, \cdots, N \tag{3.3}$$

$b$ we assume to be time and space invariant, and is estimated in this study using OLS by pooling all the observations. $\nu_{it}$ represents the disturbances.[10] Say $\sigma_1^2$ is the variance of productivities at an initial time period, and $\sigma_\nu^2$ the variance of the disturbances, then if $b^2 < 1 - \sigma_\nu^2/\sigma_1^2$, the productivities will converge over time, in the sense of Definition 1, the gaps between all economies will shrink. Ordinarily, in equation 3.3, $\nu$, are assumed to be identical and independently distributed (i.i.d) if this was assumed to be an aspatial process.

#### Spatial error

Disturbances can be modeled in vector notation as spatial autoregressive process such as:

$$\nu_t = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{e}_t, \tag{3.4}$$

[10] Kennedy (2003) describes that one of the ways the existence of the disturbance term is justified is for representing the omission of the influence of innumerable chance events.

where $\nu_t$ is from equation 3.3 for each time period $t$, and $\mathbf{e}_t$ is a vector of unobserved disturbances, which are i.i.d. For notational simplicity below, we denote the right-hand side of the above equation as $\boldsymbol{\Omega}$ which denotes the spatial multiplier matrix that approximates the spatial structure by which shocks to the economies may ripple back and forth across geographic borders as the following:

$$\boldsymbol{\Omega} = (\mathbf{I} - \rho\mathbf{W})^{-1}. \tag{3.5}$$

$\mathbf{W}$ denotes a binary contiguity matrix that has been row standardized.[11]

We can estimate the scalar $\rho$, which represents the strength of the dependence between neighboring regions, with the following regression:

$$\nu_t = \rho W \nu_t + e_t. \tag{3.6}$$

In order to remove this cumulative effect over a time span $k$ of all shocks in the $\sigma$-convergence measure we formulate a filtered series, $y^*_{i,t}$, combining equation

[11] The spatial weights matrix implemented here represents the geographic configuration of the 48 contiguous states of the United States. When off-diagonal elements of $W$ before it is row standardized are $w_{ij} = 1$ it refers to region $i$ and $j$ sharing a common administrative border. Zeros are placed along the diagonal of $W$ and also in all cells where no linkage exists in the network amongst these pairs of regions. This matrix is meant to be a rough approximation of a complex system of economic flows and forces which may influence how shocks in neighboring regions are associated with one another

3.3 and 3.4, using the following identity:

$$y_{i,t}^{*} - y_{i,t} = \sum_{i=1}^{N} \sum_{k=0}^{T} b^{k} \omega_{ij} e_{i,t-k}. \tag{3.7}$$

The right-hand side of equation 3.7 represents the cumulative effect of each past period $k$'s common shock from an initial period $t-T$ up through the period just before last year of the time span, $t-1$, under consideration. $\omega_{ij}$ denotes the elements of the spatial multiplier matrix $\mathbf{\Omega}$ in equation 3.5. $T$ denotes the total number of years in the time span.

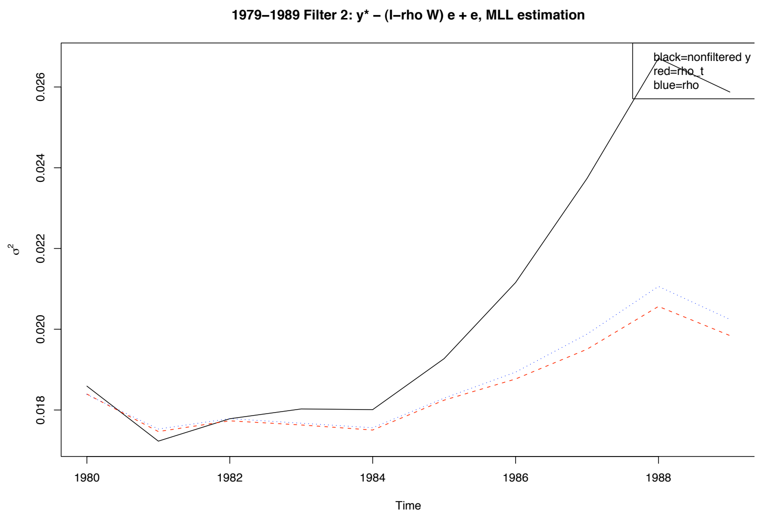
The final filtered series, $y^{**}$, is calculated by subtracting out the spatial component from the original series, but including the aspatial residual component, $e_{it}$ of 3.6, as follows.

$$y_{i,t}^{**} = y_{i,t} - \sum_{i=1}^{i=N} \sum_{k=0}^{T} b^{k} \omega_{ij} e_{i,t-k} + e_{i,t-k}. \tag{3.8}$$

## 3.5 Empirical Results

Using a spatial filtering approach the result shows that random disturbances of a spatial nature might obscure how researchers identify changes in the basic structure of the economy. The top line of figure 3.2 represents the original $\sigma$-convergence measure, the unfiltered variance of state per-capita income over time, what would typically be observed, and the bottom lines are filtered series,

**Figure 3.2:** Filtered versus unfiltered $\sigma$-convergence: 1980 to 1990, $\mathbf{y}_t$ versus $\mathbf{y}_t^{**}$




removing the cumulative effect of spatial autocorrelation of random disturbances on the trend. The difference between the two bottom lines is that the higher (red) one and is based OLS estimation of the spatial parameter and the lower (blue) line is based on Maximum Likelihood estimation.

The empirical results here conflict with Fan and Casetti (1997)'s statement that the idea of convergence to describe the regional economic system of the United States has become obsolete. Assessing state income inequality as unfairness is made more complex when we take a spatial view of how income gaps change. There can simultaneously be an underlying convergence process that characterizes the fairness of the overall institutional structure along with expected periods of divergence due to random shocks over time that are not related to changes in the convergence mechanism, but whose significance in terms of unfairness appears magnified by spatial patterning. Nevertheless, even though this type of increased dispersion may not be related to a change in the fairness of the basic economic structure, it could still be linked to expected suffering on the part of some regions negatively affected by the change.

## 3.6 Summary

In summary, when we try to describe changes with time-series plots in regional inequality when the regions are interdependent, care should be taken in how we interpret the results. The issue will be greater for geographic areas characterized by high levels of spatial autocorrelation. For example, using the Moran's I statistic, significant autocorrelation has been found across 48 contiguous states of the USA for each year of the time span between 1930 and 1995 (Rey and Montouri, 1999), across 136 NUTS2 level regions of the European Community for each year between 1980 and 1995 (Le Gallo and Ertur, 2002), across counties nested within the majority [12] of states of the USA between 1969 and 2000 (Janikas and Rey, 2005), and across states of Brazil from 1939 to 1996 (Mossi et al., 2003). The filtering approach can be thought of as an exploratory tool in guiding norma-

[12] Janikas and Rey (2005) indicate 35 out of 48 states had significant clustering.

tive interpretations of sudden changes in aggregate income dispersion trends of geographic areas.

# Chapter 4

# Assessing income segregation associated with the design of a school attendance zone

## 4.1 Introduction

This chapter proposes a metric tied to a concern for identifying specific segments of a school attendance zone boundary that might act as barriers to a child's cooperative encounters with children of other income classes. The proposed metric incorporates spatial structure by giving greater weight to income differentials among blocks that are contiguous, to proxy the blocking of potential encounters by an attendance boundary. Borrowing from Rey (2004a), this chapter uses spatial permutations for identifying local hot spots, where a boundary's segment appears to split areas by income size (hot spots in terms of statistically significant high income differentials and not clustering as in Getis (2008)). It is important to note, the aim of the metric proposed here, like chapter 2's metric, is not to inform a concern for the culprit of the boundary design, but rather to highlight areas, in the spirit of exploratory spatial data analysis (ESDA) (Anselin and Getis, 2010; Janikas and Rey, 2005), in terms of a concern for the effects of institutional boundaries.

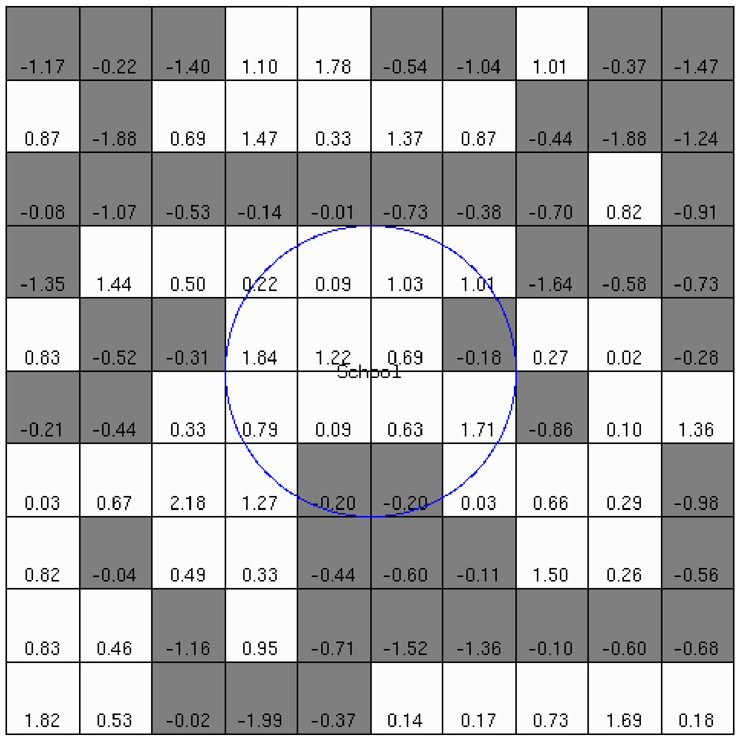
Concerns of unfairness as a cause are different from concerns of suffering as an outcome or effect. To examine how the assessments of the institutional fairness of attendance zone assignments may be clouded by voluntary residential sorting, I divide the data generating process of geographic income distributions into two components. One is spatial and not related to unfairness, the other is aspatial and related to unfairness. For the former I approximate the residential voluntary self-sorting of people by income with a spatial dependent data generating process. The idea behind this was that observations of residential income clustering appear in the same way as simulated spatial autocorrelation patterns. An aspatial income generating process was formulated to approximate the fair or just drawing of attendance zones. The results show how there is a problem in that both spatial and aspatial components both influence significance testing of inequality measurements thereby obscuring judgments of whether segregation is due to random residential sorting or unfair policies.

The issue of voluntary residential sorting does not impact measurements if we explicitly make our concern one of outcomes. To illustrate the effects or outcome of school boundaries, consider how Kozol (2005) associates large income gaps between school zones of New York City as a cause of the unbalanced distribution of educational resources. Kozol (2005), in explaining the factors leading to the desperate conditions of some school children in New York City, associates the large median income gaps between school neighborhoods as a causal reason for why it is harder for the poorer zones to offer competitive salaries and facilities to attract the best teachers. Kozol (2005, p.46) gives as an example the median teacher salary gap between schools in the poorer Bronx and the richer Queens as being \$46,000 queens versus Bronx \$64,000 for the New York City 2002-3 school year. He emphasizes that this excludes additional resources from private funding that is often a part of the school salary budget that is not shown in public data records. Another negative outcome of school boundaries is social isolation. To illustrate, Kozol (2005, p.16) quotes a church pastor talking about children he works with from the poorer schools of New York City: "They don't have any friends who are white children. When I take them with me sometimes to Manhattan to go shopping at a store for something special that they want or to a movie on one of their birthdays, and they find themselves surrounded by a lot of white kids, many of the younger ones get very scared. It's an utterly different world for them. In racial terms, they're almost totally cut off." Attendance zone boundaries may act as a barrier to potential inter-racial social interaction. Illustrating this in concrete terms, in Kozol (2005, p.28) there is a quote: "It's like we're being hidden," said a fifteen-year-old girl named Isabel I met some years ago in Harlem. "It's as if you have been put into a garage where, if they don't have room for something but aren't sure if they should throw it out, they put it there where they don't need to think of it again." Concern exists regarding the coupled relationship between economic deprivation and social isolation (or exclusion) of some minority groups in cities of the United States, especially through the school system (Wilson, 1987a; Kozol, 1991).

The local spatial view taken in this chapter is that nearby pairs of neighborhoods are more important to children's potential encounters than far away neighborhoods. From this local spatial perspective of a child's perceptions, a local measure of sorting due to particular segments of an institutional boundary can be formulated. The empirical significance of this measure is that we can assess inequality in terms of local suffering caused by local boundary segments instead of focusing on assessing the institutional fairness of an entire school attendance zone design. The new local measure can be seen as an exploratory approach that takes a local perspective, or block by block perspective, to assess the role of segments of an attendance zone boundary on inequality.

This chapter has three main sections. In section 4.2, I explain an approach for incorporating a spatial view of inequality into segregation assessments. In section 4.3, I propose a new local measure. Then, in section 4.4, I explain how residential

**Figure 4.1:** Hypothetical geographic space with a school zone, income realizations, and block-like subareas




sorting, in the form of spatial income autocorrelation, can affect our judgments on the fairness of a school attendance zone design.

## 4.2 Spatial weights

Figure 4.1 illustrates the geographical relationships considered in this chapter for describing how a zonal boundary affects inequality. The lattice shape replicates a set of block-like subareas. Those blocks intersected by the circle and within the circle denote blocks assigned to the school in the middle of the lattice. The values within each block represent income values. I assume that children from blocks assigned to the same school zone will form a social group since they are more likely to interact with one another than otherwise. A complementing assumption is that if there is no school zone children from different blocks would interact with one another freely. Our inequality concern is whether the presence of this school zone will structure children's interactions in a way correlated to the income level of their blocks.

The approach here is to formulate a spatial inequality measure that accounts for whether the average income gaps between blocks inside and outside the circle, between the in-zone group and out-zone group, is more or less than the average gap between all pairs of blocks; An advantage of the spatial inequality measure to be formulated is that it can be used as a test statistic in spatial permutation tests, which shuffle the assignment of subareas to a school, in order to gain an idea of the statistical significance of a zonal design on inequality.[1]

To begin, we can consider that the total variance based on all the subareas can be decomposed into two parts: spatial and aspatial. In this case we use the

[1] Walzer (1983, p.215) says: "Randomness is the most obvious associative principle. If we were to bring children together without regard to the occupations and wealth of their parents, without regard to the political or religious commitments of their parents,...we might produce perfectly autonomous educational communities." Walzer (1983, p.215) just uses this idea of randomness as an associative principle as a frame of reference for understanding the kind of association advocated by leftist groups, but he does not support it. (He finds it extreme.) I quote this not because of political reasons, but only because Walzer (1983)'s idea of randomness as an associative principle conveniently matches what is being captured in the school income inequality assessment approach being introduced here.

term spatial income gaps to refer to how the drawing of the geographic boundary effects income inequality between the two resulting school groups of children, those within and those outside of the school zone or circle in figure 4.1.

The approach is based on the variance statistic, $V$, which can be used to measure changes in income inequality. Consider distributions of a random variable, $Y$, say per capita income, over $n$ subareas (blocks), $i = 1, ..., n$, and let $y_{i,t}$ be per capita income of subarea $i$ at time $t$. Let the average level of income for all subareas be $\bar{y}_t = \frac{1}{n}\sum_{i=1}^{n} y_{i,t}$, so that $V$ is given by:

$$V_t = 1/n \sum_{i=1}^{n} (y_{i,t} - \bar{y}_t)^2 \tag{4.1}$$

Evidence of reduced economic inequality is reflected in equation 4.1 declining, such as:

$$V_2 < V_1 \tag{4.2}$$

Ordinarily, the subscripts 2 and 1 above denote ending and starting time periods. Here, instead of comparing inequality changes based on different income distributions of $Y$ at different time periods, as in chapters 2 and 3, in this chapter I am concerned with how average income gaps, between areas inside a school zone and outside a school zone change between other hypothetical school attendance assignments. Other hypothetical assignments can be represented as different realizations of the spatial structure of social interactions, $W$, imposed by a school zone design.[2] So for this chapter, 2 and 1 in equation 4.2, denote different sets of neighborhood school assignments.

As discussed in chapter 2, I use the term 'income gap' to refer to the squared difference in the income between two subareas, such as:

$$d_{ij} = (y_i - y_j)^2 \tag{4.3}$$

where $i, j$ are pairs of subareas. I believe it is important to differentiate amongst income gaps based on the social distance between $i, j$ (here, defined by the different school zone assignments of each neighborhood) in terms of affecting the probability that the associated communities compete with one another for economic resources.

Consider how a pair-wise computation of variance, based on income gaps, leads to an equivalent result as equation 4.1. Income variance can be understood as twice the average of all income gaps $d_{ij}$, as follows:

$$2V = \frac{\sum \sum d_{ij}}{n(n-1)} \tag{4.4}$$

[2]The notion of how different zonation schemes can lead to different results is also known as the Modifiable Areal Unit Problem (MAUP) (Oppenshaw and Taylor, 1991). Here, this situation is not a problem but a useful informational property that allows us to assess the significance of the inequality measure with the use of spatial permutation tests.

The denominator of the above equation, $n(n-1)$, is the total count of all possible gaps between subareas $i, j$. The numerator of the RHS of equation 4.4 is the sum of the squared deviations of income gaps, from equation 4.3. $2V$ is equivalent to the average of the universe of such gaps.

To see how spatial structure is implied by the variance statistic, it can be considered as the weighted average of all income gaps.

$$2V = \frac{\sum\sum w_{ij}d_{ij}}{n(n-1)} \tag{4.5}$$

By defining all $w_{ij}$'s the same, as 1, reflects a situation of where all gaps are considered the same regardless of heterogeneity of areal attributes such as school assignment, race, or geographic position. The advantage of the pair-wise approach to computing variance in equation 4.5 is the simplicity by which we can use weights to represent spatial structure in a formulation of a inequality measure.

Squared income gaps between all $n$ subareas in the population can be represented by a matrix $\mathbf{D} = [d_{ij}]$ (from equation 4.3). $\mathbf{D}$ can be decomposed into two components, spatial and aspatial:

$$2\text{n(n-1)V} = \underbrace{\mathbf{1}^{'}(\mathbf{W} \odot \mathbf{D})\mathbf{1}}_{\text{spatial}} + \underbrace{\mathbf{1}^{'}(\mathbf{W}^{+} \odot \mathbf{D})\mathbf{1}}_{\text{aspatial}} \tag{4.6}$$

$\mathbf{W} = [w_{ij}]$ represents the association between pairs of subareas $i, j$ using binary weights and its matrix complement (switching the ones for zeros) is $\mathbf{W}^{+}$. $\mathbf{1}$ is a $n$ by 1 vector of ones and $\mathbf{1}^{'}$ its transpose.

I define as a function $V(\cdot)$ based on the respective spatial weights matrix, as follows:

$$V(\mathbf{W}) = \frac{\mathbf{1}^{'}(\mathbf{W} \odot \mathbf{D})\mathbf{1}}{n_w} \tag{4.7}$$

$n_w$ is the count of the non-zero, non-diagonal elements of $\mathbf{W}$. The result of equation 4.7 is simply the average size of each income gap represented by non-zero elements of $\mathbf{W}$. The result is also the equivalent to the spatial component of equation 4.6. If the average size of all gaps for a set of subareas, $I$, is the same as that for some subset of gaps represented by non-zero elements of $\mathbf{W}$, then the value of the aggregate measure $V$ and the spatial measure $V(\mathbf{W})$ would be equal, $1 = V(\mathbf{W})/2V$, as a result of combining equation 4.5 and equation 4.7. The general formulation of the spatial income inequality measure is based on this ratio of measures, the spatially significant gaps to all gaps. Next, I will discuss how spatial structure $W$ can be represented.

### 4.2.1 Formulation

**D** can be decomposed into two components, spatial and aspatial (equation 4.6). The challenge is how to represent the spatial structure, $W$, by which to assess how a zone's design alters social interactions of a place in an unbalanced or unequal manner. The key of the approach is to represent how all $N \times (N-1)$ pairwise income gaps between areas are associated with one another, jointly based on their geographic position, social position (or racial attribution), and school zone assignment. To begin, we can understand this notion of spatial structure in terms of three components that may be relevant in how a school zone affects relationships amongst children, as follows.

- Neighborhood social interaction
- Barriers to interaction from school zones
- Racial dissimilarity

Below I represent these components as matrices.

**Neighborhood social interaction**

$\mathbf{G} = [g_{ij}]$ represents pairs of subareas $i, j$, that are located in the same vicinity that are influenced by the social division effect from a school zone design, and its elements are defined as follows. Let $N(i)$ denote the set of subareas $j$ that are considered neighbors of $i$.

$$g_{ij} \quad \begin{cases} 1, \text{ if } j \in N_i \\ 0, \text{ otherwise;} \end{cases}$$

The notion of neighborhood $N(i)$ means that there is more potential for encounters between children in $i$ and $N(i)$ to affect the socio-economic outcomes of children in $i$ than encounters between children in $i$ and people in other subareas, assuming no school zone effect.[3] In an intuitive sense, this would mean that a subarea in Montana should not be considered in the same neighborhood as a subarea in New York City in an assessment of a school zone. This type of geographic interpretation, however, is legalistically important in more subtle situations such as in the Detroit's school desegregation case of *Milliken v. Bradley, 418* U.S.717

[3]More technically, we can understand a neighborhood with following notation.

| | |
|---|---|
| $i, I$ | index and set of areas (city blocks), $i = \{1, \cdots, n\}$; |
| $X$ | spatial random process; ($X_i$ social outcome) |
| $w_{ij}$ | $\begin{cases} 1, \text{ if } P(x_i) \neq P(x_i \mid x_j) \text{ , with } i, j \in I \text{ and } i \neq j \\ 0, \text{ otherwise;} \end{cases}$ |
| $N_i$ | $\{j \mid w_{ij} = 1\}$; |

(1974) where there was a debate on whether subareas of the suburban ring should be counted when assessing school racial imbalances in the central city (Delaney, 2001). As will be shown below, a set of local measures can be formulated with different definitions for each subarea's local space of potential social interaction, $G_i$, that has the advantage of capturing a local resident's perspective of how local segments of a boundary appear to her to sort by income.

### Barriers to interaction from school zones

$\mathbf{C}_{k,l} = [c_{ij}]$ represents pairs of subareas $i, j$ not in the same school zone, and its elements are defined as follows. $\Im(\cdot) f : \mathbf{i} \rightarrow Z.$ is a function that tells what zone $z$ subarea $i$ belongs to.

$$c_{ij} \quad \begin{cases} 1, \text{ if } \Im(i) \neq \Im(j) \\ 0, \text{ otherwise;} \end{cases}$$

The weights $c_{ij}$ can be used to account for how a zonal boundary divides the people of subarea $i$ from subarea $j$. By defining all $c_{ij}$'s the same, as 1, reflects a situation of where there are no school zone boundaries to structure the encounter or interactions of children from different neighborhoods.

### Racial Dissimilarity

$\mathbf{R} = [r_{ij}]$ represents social or racial dissimilarity between pairs of subareas $i, j$ and its elements as defined as follows.

$\wp(\cdot) f : [\mathbf{a_1}, \mathbf{a_2}, \cdots \mathbf{a_z}] \rightarrow P$. is a function that classifies a given subarea $i$ into one and only one social position, $P$ based on areal attributes $[\mathbf{a_1}, \mathbf{a_2}, \cdots \mathbf{a_z}]$.

$$r_{ij} \quad \begin{cases} 1, \text{ if } \wp(i) \neq \wp(j) \\ 0, \text{ otherwise;} \end{cases}$$

Before one can define racial similarity some prior notion of relevant racial or social positions is needed.[4]

Spatial structure of how income gaps are structured by a school zone can be thought of as follows.

$$\mathbf{W} = \mathbf{G} \odot \mathbf{C} \odot \mathbf{R} \tag{4.8}$$

The above racial, school zone and neighborhood spatial components can be combined to form a single spatial weights matrix, $W$, whose elements can be un-

[4]Each society has its own social positions that are expected to play a role in the economic capabilities of its citizens, such as gender, caste, tribe, culture, and inherited occupation (Rawls, 1971).

derstood as follows.

$$w_{ij} \quad \begin{cases} 1, \text{ if } \wp(i) \neq \wp(j) \wedge \Im(i) \neq \Im(j) \wedge j \in N_i \\ \\ 0, \text{ otherwise;} \end{cases}$$

Using this spatial perspective, an advantage of a matrix representation of the pair-wise relationships is the simplicity by which we can compare subsets of income gaps in terms of their proportion to total inequality, and, the main focus here, in terms of their average size.

I borrow Rey (2004a)'s spatial permutation method (explained in chapter 2 with equations 2.12 and 2.18) that he uses to assess the significance of inter-regional inequality within a country and extend it to assessing the inequality associated with the spatial structure imposed on children by the design of a school zone. This spatial structure is defined by different $\mathbf{W}$'s, as follows:

$$p(\mathbf{W}) = Pr(V(\mathbf{W}^*) > V(\mathbf{W}^R)) \tag{4.9}$$

$\mathbf{W}^*$ is the actual observed spatial structure and $\mathbf{W}^R$ its random realization after randomly permutating the columns and rows of the matrix. For the global inequality measures to form an empirical distribution of random realizations this is equivalent to randomly shuffling the school assignments of subareas to a school.

The smaller the likelihood that an actual realization's measurement is greater than random realizations means the more significant the evidence is towards associating a school zone design as having increased inequality for the geographic zone being considered.[5]

A proportional measure of inequality linked to racial differences in the composition of neighborhoods is given by:

$$P = \frac{\mathbf{1}(\mathbf{D} \odot \mathbf{G} \odot \mathbf{C} \odot \mathbf{R})\mathbf{1}'}{\mathbf{1}(\mathbf{D} \odot \mathbf{G} \odot \mathbf{C})\mathbf{1}'} \tag{4.10}$$

The difference in the numerator and the denominator is that the racial dissimilarity matrix $\mathbf{R}$ is in the numerator. This measure provides a description of the proportion of total income dispersion between in-zone and out-zone subareas attributable to racial dissimilarity between those subareas.

This measure can be thought of as accounting for an intra-zonal, inter-social perspective of inequality which can be extended more generally to encompass a notion of intersectionality, highlighting how the most acute inequalities may exist where particular societal positions intersect (Crenshaw, 1991), in particular, race and income.

[5]In a sense, equation 4.9 captures the idea of Walzer (1983, p.215) of randomness as an associative principle. See footnote 5 of this chapter.

For the proportional measure we could compute pseudo-significance levels as a way to assess the value of the information in **C** or **R** conditional on **G** when defining spatial structure. For instance, to test the significance of a school zone pattern of attendance assignments on income inequality compared to random assignments, we use

$$p(\mathbf{S}) = Pr(V(\mathbf{G}^* \odot \mathbf{C}^*) > V(\mathbf{G}^* \odot \mathbf{C}^R)) \tag{4.11}$$

The proportional measure is given to illustrate the advantage of using a decomposable spatial approach, but the focus of this chapter is on the average income gap inequality measure.

An average approach is to measure global income inequality associated with a school zone is given by

$$A = \frac{V(\mathbf{G} \odot \mathbf{C})}{V(\mathbf{G})} \tag{4.12}$$

This measure is the ratio of the average squared income gap from 2 sets of pairwise relationships. The numerator accounts for the summation over the subset of gaps between just subareas in the school zone and subareas out of the school zone. In other words, we are referring to gaps between block subareas inside and outside the circle in figure 4.1. In a sense, the denominator is a device to represent what the social associations would hypothetically be without the enforcement of school attendance zones.[6] The denominator accounts for the summation over all the set of gaps between subareas of what can be considered the essential geographic area of potential social encounters. In other words, we are referring to all pair-wise relationships between block subareas in figure 4.1, regardless of the school zone circle. The measure can be thought of as a useful as a 'rule of thumb' to analyze the effect of school assignments on inequality.

The index of dissimilarity is considered the 'work horse' measurement in segregation studies dealing with racial imbalances (Clotfelter, 2004). I will compare a new local and the above global measure with it in simulations below to understand how spatial autocorrelation effects conventional statistical assessments. The dissimilarity index is defined as follow.

$$D = .5 \sum |B_j/B - W_j/W| \tag{4.13}$$

,where $B$ and $W$ are total number of nonwhite and white subareas in a geographic system, and $B_j$ and $W_j$ are the nonwhite and white number of assignments of subareas within the school zone.[7]

[6]See footnote 5 of this chapter.

[7]For this measure we need categorical variables so the income realizations below, in section 4.4.1, Monte Carlo simulations have been transformed into a binary values as a function of whether they are above or below zero.

## 4.3 Local index of sorting by a boundary

In this section, I introduce a local inequality measure because of empirical and substantive problems with global measures. The empirical problem is that residential self-sorting, or residential *de facto* segregation, can obscure judgments of school segregation by the design of school attendance zones (explored in the next section). The substantive problem is that global measures, when they show no evidence of institutional unfairness, may mask local suffering on the part of some neighborhoods in terms of social isolation.

Legal solutions to change the boundaries are made complex due to the uncertainty around the intentionality of the segregated attendance zone boundaries in view of larger patterns of voluntary residential movements across attendance zones, as well as the hardships of increased transportation costs that desegregation policy changes may cause. The characterization of the mechanisms underlying school segregation have been legally polarized between the terms *de jure* and *de facto* segregation. I believe, assuming the perspective of a hypothetical local resident that seeks increased interracial contact, that there can exist a subtle middle ground, where one cannot distinguish between *de facto* and *de jure* segregation since both can contemporaneously exist. I propose an empirically based index to assess the significance of this situation for local areas of a city that is based on a test for the significance of a local school segregated pattern of school attendance.

The problem with global measures is that they ignore the income dissimilarities that exist just at the edge of a school zone's boundary. The new local approach is based on the idea that we can capture a local resident's spatial view of inequality within a measure. One's judgments regarding the geographic distribution of a social attribute may partly be a function of what one can concretely observe with one's eyes. A child resident of a neighborhood has a geographic horizon from which she can infer patterns of racial segregation in the world. What appears random to her may appear clustered to the researcher, and vice versa, what appears clustered to her may appear random. The intention here is to formulate a measure to capture this perspective regarding the affect of how a boundary representing areal assignments to some institution, such as a school attendance zone, may be perceived as sorting children to increase inequality, capturing a sentiment of potential social isolation of how a boundary divides social interactions of people in the same geographic vicinity.

The global measure of equation 4.12 is based on the proposition that if there was no school zone children would be free to interact with one another within the geographic field of study, irrespective of their relative location to one another. This may be socially realistic. In other words, this proposition is equivalent to weight all pair-wise deviations $d_{ij}$ the same, ignoring variation in $g_{ij}$'s within the study area (i.e. $g_{11} = g_{12} = ...g_{nn} = 1$). It may be questionable to weight all pairs the same, especially if the objective is to capture the perspective of a local resident's feeling of being socially sorted, or segregated, rather than assessing institutional unfairness in overall attendance zone design.

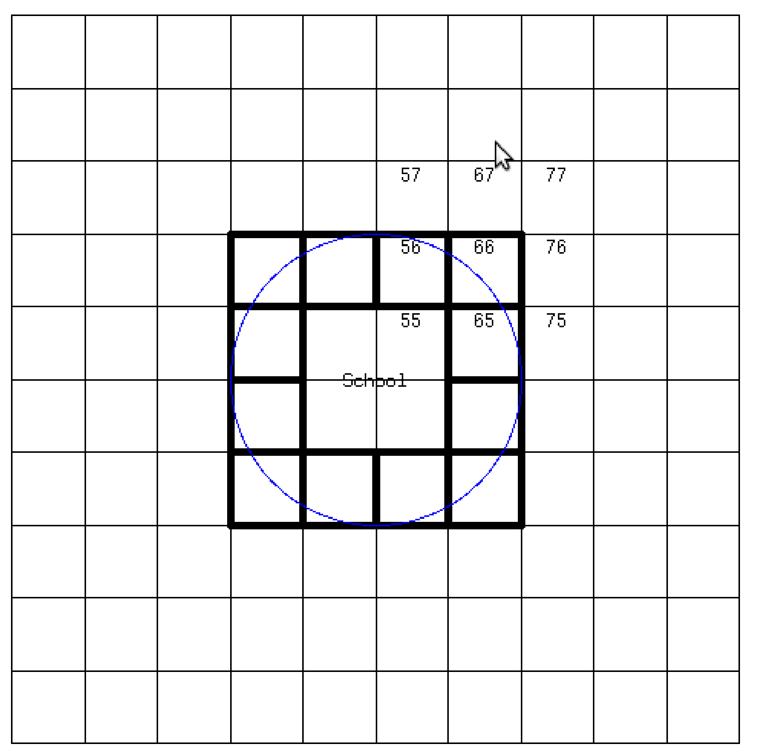
A local average approach is given by

$$A_i = \frac{V(\mathbf{G}_i \odot \mathbf{C})}{V(\mathbf{G}_i)} \tag{4.14}$$

This measure is calculated just for subareas along the boundary of a school zone to highlight particular segments that effect inequality or unbalanced assignments. The key difference between the global and local average measure is how we geographically shrink the notion of a neighborhood, $N(i)$ (from section 4.2.1) to define equation 4.14.

One notion of a neighborhood is the geographic area that we consider the potential social encounters amongst children in $i$ to be spatially bounded by. (In concrete terms, one way we could imagine a local conception of a neighborhood is the area within which a child can bike or walk to meet her friends.) Figure 4.2 illustrates the spatial relationships by which the local measure is calculated. The block subareas that have a thick edge denote ones that are along the boundary

**Figure 4.2:** Boundary subareas a school zone and local neighborhood of subareas



of the school zone, the subareas that are used to compute the local measure. All numbered blocks in figure 4.2 are considered to be part of the neighborhood set $N(i)$, where $i$ is block 66 ($i = 66$). (In figure 4.2, I use the digits at the top of the block subareas to denote the value of their indices $i$. These match their coordinate positions on the lattice where an index of $i = 00$ would denote a subarea block located in the bottom left corner of the grid.) For this local neighborhood, the in-zone subset of blocks is (55, 56, 65, 66) and the out-zone set of blocks is (57, 67, 75, 76, 77, 76). The local measure of equation 4.14 computed for $i = 66$ can be understood as a fraction of the summation of squared income gaps between the in-zone blocks (55, 56, 65, 66) and out-zone blocks (57, 67, 75, 76, 77, 76) over the summation of all the pair-wise gaps for $i$'s local neighborhood set (55, 56, 57, 65, 66, 67, 75, 76, 77).

Statistical inference for the local inequality measure is problematic because of the small number of combinations due to the small sized neighborhoods defined with $N(i)$. Instead of random shuffling of assignments, as done with the global inequality measure, an empirical distribution can be made by exhaustively calculating different realizations of the measure for all possible combinations of in-zone subareas that could be assigned from each subareas set of neighborhoods. The advantage with this method is that we can at least know the number of possible unique combinations the reference distribution is built upon.

The number of possible unique combinations of objects when the order of objects doesn't matter and objects are replaceable (cannot be repeated within a combination realization) is as follows:

$$\begin{pmatrix} n \\ \\ r \end{pmatrix} = \frac{n!}{r!(n-r)!}$$

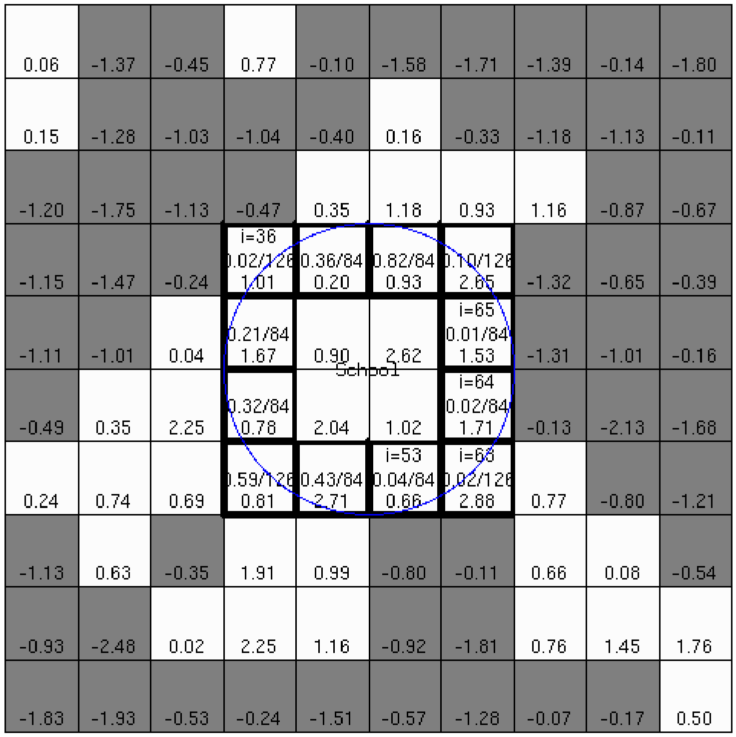
Empirical distributions of the local inequality measure for the block subareas on the corners of the boundary area, such as block $i = 66$ in figure 4.2, are based on taking different combinations of 4 in-zone subarea objects from a total of 9 subareas ($n = 9, r = 4$) to calculate the measure. This results is 126 combinations of in-zone sets, the remaining non-corner blocks would have 84 combinations ($n = 9, r = 3$). If the actual local measure for a neighborhood $N(i)$ is more extreme than a significant part of the distribution of realizations based on all possible combinations of school assignments in the neighborhood, then this is evidence that a resident of that neighborhood would feel that the school boundary unequally divides her neighborhood's social relationships.

Figures 4.3 and 4.4 show simulated results of the local measure. In figure 4.3, the bottom of each block reports the value of $y_i$, income realization in each subarea $i$. The blocks along the zone boundary (with darker edges) contain more information. The subarea index $i$ is reported in the top of the block only when that subarea is associated with a local index that has a percentage value (pseudo-pvalue) of .05 or less, indicating that its surrounding neighborhood has been segregated by the boundary segment: 36, 53, 63, 64, 65. For example, in the top left block $i = 36$ the middle pair of numbers report the results from the local inequality index test. The first number in the pair, .02, denotes the percentage of times the spatial permutations are greater than the actual measure. The second number 126 denotes the number of possible spatial combinations that were used to build the empirical distribution.

## 4.4 Inequality measures in the presence of spatial autocorrelation

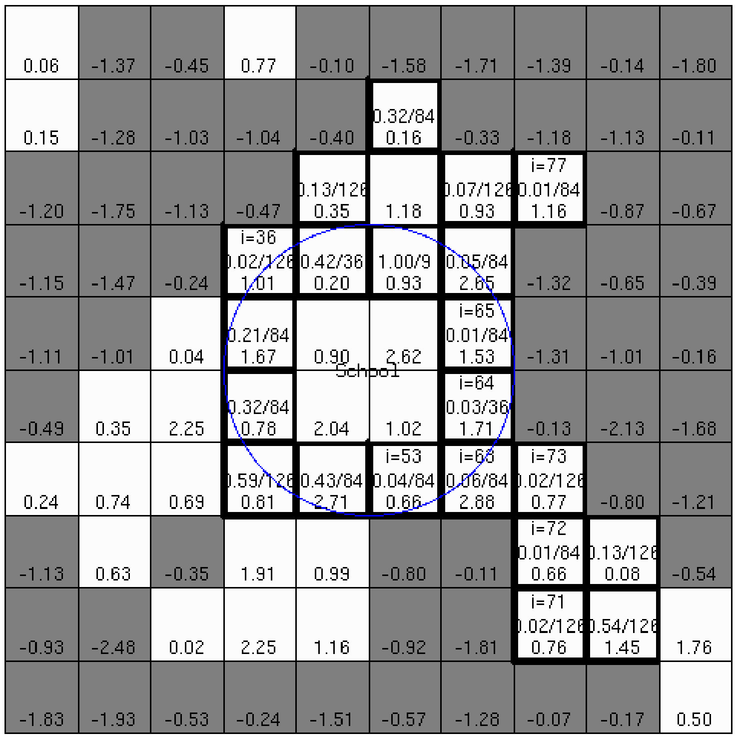
In this section, I discuss the empirical problem of the presence of spatial autocorrelation on assessing school zone income inequality. In the presence of residential segregation there may be a problem in identifying unfair segregation with global measures. If the underlying data generating process of the geographic income distribution between neighborhoods within and outside of a school zone is based on a combination of spatial dependence of incomes and institutional design, then it may not be possible to determine how much of total inequality between in-zone and out-zone assigned children is the responsibility of the design of the

**Figure 4.3:** Hypothetical geographic space with a school zone, income realizations, and block-like subareas




school zone and how much from clustered residential patterns. The spatially patterned income values in figures 4.3 and 4.4 suggest how a randomly designed school zone might be wrongly inferred as being intentionally segregated in the presence of residential income sorting (spatial autocorrelation of incomes). The clustered patterns of high (light) and low (dark) values in figures 4.3 and 4.4 also illustrate how spatial autocorrelation of the income realizations can replicate residential income segregation.

**Figure 4.4:** Hypothetical segregated school zone in presence of spatial autocorrelation




The dissimilarity index for describing school zone inequality in figure 4.3 is .37. The inequality measure $A$ is 3.59. The realizations $y_i$ in figure 4.3 are based on a spatial autoregressive data generating process ($\rho = .9$). The school zone income inequality we observe in figure 4.3 is actually not due to an unfair or segregated zonal design, but essentially a random design of school assignments that appears as though it is intentionally segregated.

### 4.4.1 Simulations

Although segregation measures are mainly used in a descriptive fashion, their statistical interpretations may shed light on how their results match our normative concerns. The size performance of spatial permutation tests for three types of in inequality measures are investigated here: the global income inequality index $A$ in equation 4.12, the local inequality index $A_i$ in equation 4.14, and an index of dissimilarity, $D$ in equation 4.13 below.

Simulated realizations are from a spatial DGP (explained with equation 3.4 in chapter 3 of this dissertation) with different degrees of dependence $\rho$:

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{e}_t, \tag{4.15}$$

$\mathbf{y}$ is a vector of income values for the subareas. $\mathbf{W}$ denotes a binary contiguity matrix that has been row standardized.

The objective of the Monte Carlo simulations was to see what percentage of times the three tests falsely reject a true null hypothesis of random assignments. The 'true' pvalues upon which the performance of the permutation tests are assessed against are based on forming empirical distributions based on 1000 realizations of $\mathbf{y}$ from equation 4.15.

**Table 4.1:** Monte Carlo results of false positives of zonal inequality under a spatial data generating process of $\sigma_e = 1$, $N = 100$ ($pvalue = 0.05$)

| | Fraction of tests rejected for the null hypothesis | | |
|---|---|---|---|
| $\rho$ | Dissimilarity Index, $D$ | Global Inequality, $A$ | Local Inequality $A$ |
| .0 | 0.05 | 0.05 | 0.05 |
| .3 | 0.11 | 0.07 | 0.06 |
| .6 | 0.20 | 0.10 | 0.07 |
| .9 | 0.38 | 0.12 | 0.11 |

The table shows the percentage of 1,000 replications that gave a larger test statistic than the value corresponding to their permutation test five-percent significance level.

Table 4.4.1 shows the results of the simulation experiment for all test statistics. Table 4.4.1 reports the percentage of the replications that gave a larger test statistic than the value corresponding to the 'true' five-percent significance level (based on 1000 realizations of income vectors). The rows of the table organize Monte Carlo draws that are based on varying degrees of spatial dependence for generating the values $y_i$.

The size of the spatial dependence parameter, $p$, has a positive influence on the variance of the test statistics for the all three inequality measures, but especially for the dissimilarity index. This suggests that spatial dependence makes it harder to identify whether a school zone segregates children using global measures because of the increasing likelihood of a Type I error.

## 4.5 Summary

We don't know the true data generating process causing school zone income inequality. However, even if there is evidence of no intent in the design of a school zone to segregate, this does not preclude some segments of a boundary from reinforcing the negative outcomes of segregated residential patterns. Therefore this chapter proposed a local inequality measure aimed at identifying harmful boundary segments in terms of outcomes, instead of inferring unfairness. Measures following this local approach could allow one to judge if specific segments along a boundary could be perceived by a local resident to divide a child's surrounding neighborhood into poor blocks within the zone and rich blocks outside the zone.

The proposed metric can be applied to political discussions on changing the configuration of a school zone which is suspected of contributing to the isolation of minority school children. As an exploratory tool, the metric could also be applied to finding pockets of high differentials in terms of health outcomes along other types of boundaries that divide health services, such as for countries, states, counties. The advantage of a local approach is that it controls for many other factors shared by people living next to one another, which we might not have data for, such as genetics and climate. For example, Gladwell (1998) explains how the health differentials among members of the Pima Indian tribe living on different side of the Mexican-American border provides a natural experiment for understanding factors of obesity. The method proposed here could be used as an exploratory tool to find similar types of observations.

As a future direction of research, significance testing can be based on conditional spatial permutations. Instead of shuffling only the income values of contiguous areas surrounding each boundary segment to get a test statistic for that segment, we can randomly pull income values from the full geographic extent and shuffle them among areas along different segments. This type of statistical test may allow us to cope with mean income heterogeneity on the landscape.

The significance of this measure is that we can assess inequality in terms of local suffering caused by local boundary segments instead of focusing on assessing the institutional unfairness of an entire school attendance zone design.

# Chapter 5

# Conclusion

Chapters 2, 3, and 4 dealt with measuring inequality using the distribution of incomes across neighborhoods, states, and school attendance zones, respectively. Each of these chapters illustrated an approach to assess inequality that can be divided into 3 steps or objectives. The first objective was to develop a spatial view of inequality. The second objective was to incorporate this spatial view of inequality into the inequality assessment using a spatial weights matrix. The third objective of each chapter was to understand the empirical significance of incorporating a spatial weights matrix.

The dissertation started with an introductory chapter that introduced the preliminary ideas of understanding a spatial view of inequality. Our intention to quantify inequality using statistical measures is based on a practical goal to communicate information on public concerns in an objective manner. The ultimate goal of inequality statistics is to inform society's awareness of suffering and judgments of institutional unfairness. It is important then that the formulation of our aggregative measures validly matches the reality of the situation in terms of how different income gaps hold different significance to different concerns. In more concrete terms, it is important that the objectively measured size of an overall change upwards (or downwards) of an inequality statistic actually corresponds to when a situation has become normatively worse (or better).

In chapter 2, the motivating normative concern for assessing neighborhood income inequality was related to suffering. Sudden income gap changes can effect a reduction of a poorer community's ability to obtain absolute economic capabilities such as health care and education for which they may be in competition with other communities. A key empirical finding was that different geographic and racial patterns of changing inequality do not move in concordance with one another, and may even move in opposite directions, putting into question the relevance of aggregate assessments of neighborhood income inequality changes.

In chapter 3, the motivating concern for assessing state income inequality related to unfairness. Assessments of changes in the distribution of income across regions of a political system can provide information for appraising those institutions that regulate the conflicting economic interests of the different geopolitical units of the political system. Fan and Casetti (1997) used an observation of an increase in the dispersion of incomes of states of the United States in the 1980's as evidence that the ideas behind the Neoclassical economic convergence hypothesis are obsolete. This conclusion can be interpreted in terms of the expected fairness associated with the structure of the political economic system towards initially poor regions. To examine how spatial aspects of the data complicate matters, I incorporated a spatial weights matrix into a convergence data generating process. The idea was to model how the distributions of state incomes can be changing both from an aspatial catch-up convergence component and a random spatial error component. The results of this spatial error model were then used to filter out the spatial component from a time-series of income dispersion. The results showed the divergence in the 1980's may not have necessarily been caused by a large change in the underlying convergence process, but rather could plausibly have been caused from a geographically patterned economic disturbance. My results conflict with Fan and Casetti (1997)'s statements that the idea of convergence to describe the economic has become obsolete. The results of chapter 3 showed that assessing state income inequality as unfairness is more complex when we take a spatial view of how income gaps change. There can simultaneously be an underlying convergence process that characterizes the fairness of the overall institutional structure along with expected periods of divergence due to random shocks over time that are not related to changes in the convergence mechanism, but whose significance in terms of unfairness appears magnified by their spatial patterning.

In terms of suffering, for chapter 4, the concern regarding school segregation is how the attendance zone boundaries may act as a barrier to potential inter-racial social interaction. For this issue, the spatial view taken is that nearby pairs of neighborhoods are more important to children's potential encounters than far away neighborhoods. From this local spatial perspective of a child resident a local measure highlighting particular segments of an institutional boundary was formulated. The significance of this measure is that we can assess inequality in terms of local suffering caused by local boundary segments instead of focusing on assessing the institutional fairness of an entire school attendance zone design.

This dissertation contributes to understanding how we think geographically about our inequality concerns. Our brains naturally organize information about human relations geographically. In doing so our geographic representations of the world may be a source of bias in how me measure human welfare. Geography, as a way of gaining and communicating knowledge about human welfare, categorizes individuals in terms of abstract classes such as regions, states, cities, and neighborhoods. A common map with political boundaries is a simple, common example of how these conventional classes are used to represent a collective perception of the real world spatial relations amongst people. Questions of geographic thought addressed by this dissertation are, Does our conventional geographic understanding of how people are associated with one another realistically match the concrete reality of suffering and unfairness in societies, and, To what degree are the results of our measurements of human welfare robust to the choice of geographic associations we use?

Unlike fields such as geology or astronomy, which also map spatial relations, in human geography we are often mapping relations that are unobservable. There is no empirical way to objectively come to an agreement on the best geographic representation of an inequality problem. We must rely on common notions of suffering and unfairness, and descriptions of our intuitive notions regarding geographic associations. Furthermore, by being forced to highlight a chosen subset of associations over some other subset, a single analysis must inevitably mask suffering or obscure unfairness to some degree. Therefore, exploratory data analysis that experiments with multiple spatial weights matrices for a single problem may be a fruitful direction. We cannot be certain that incorporating a spatial weights matrix is a better approximation to human associations than not doing so. Nevertheless, in an exploratory sense, spatial analysis results provide us with information to test whether plausible alternative spatial views have empirical significance. If they do, then this alludes to the fact that we should think deeper about how our implicit generalizations on spatial structure match the concerns we wish to prioritize.

Formulas for computing inequality statistics are not transparent in how they make implicit spatial generalizations. Or, in other words, as Temkin (1993) says, in a different context, it is "not written on their sleeves". So at a minimum, incorporation of a spatial weights matrix into inequality assessments highlights how some sort of spatial presumption of human economic associations is behind our quantitative assessments. In addition, evaluation of this spatial presumption cannot be done in a manner that is as objective as it may appear to be by the equations within which they are embedded. This is because such evaluations must be informed by normative considerations(Sen, 1995, 1977). In summary, it is important not to ignore the spatial arrangement of our income observations when they are referenced geographically in our data.

# References


Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review*, 26(2):156–166.

Anselin, L. and Getis, A. (2010). Spatial statistical analysis and geographic information systems. *Perspectives on Spatial Data Analysis*, pages 35–47.

Azzoni, C. R. (2001). Economic growth and income inequality in Brazil. *Annals of Regional Science*, 31(1):133–152.

Barro, R. (2002). The U.N. is dead wrong on poverty and inequality. *BusinessWeek*, May 6:24.

Barro, R. and Sala-i-Martin, X. (2003). *Economic Growth*. McGraw-Hill, London, 2nd edition.

Barro, R. and Sala-i-Martin, X. (2004a). *Economic Growth*. MIT Press, Cambridge, MA, 2nd edition.

Barro, R. and Sala-i-Martin, X. (2004b). *Economic Growth*. MIT Press, Cambridge, MA, 2nd edition.

Bernard, A. B. and Durlauf, S. N. (1996). Interpreting tests of the convergence hypothesis. *Journal of Econometrics*, 71(1-2):p161 – 173.

Blau, J. R. and Blau, P. M. (1982). The cost of inequality: Metropolitan structure and violent crime. *American Sociological Review*, pages 114–129.

Blau, P. (1977). *Inequality and heterogeneity*. Free Press New York.

Borjas, G. J. (1995). Ethnicity, neighborhoods, and human-capital externalities. *The American Economic Review*, 85(3):365–390.

Carree, M. A. and Klomp, L. (1997). Testing the convergence hypothesis: a comment. *Review of Economics and Statsitics*, 77:683–686.

Census Bureau (2013). U.S. Census Bureau: FAQs.

Chakravorty, S. (1996). A measurement of spatial disparity: The case of income inequality. *Urban Studies*, 33(9):1671–1686.

Chen, J. and Fleisher, B. M. (1996). Regional income inequality and economic growth in china. *Journal of Comparative Economics*, 22:141–164.

Clark, W. A. (1991). Residential preferences and neighborhood racial segregation: A test of the schelling segregation model. *Demography*, 28(1):1–19.

Cliff, A. D. and Ord, J. (1973). *Spatial Autocorrelation*. Pion, London.

Clotfelter, C. (2004). *After Brown: The Rise and Retreat of School Desegregation*. Princeton University Press.

Coughlin, C. C. and Mandelbaum, T. B. (1988). Why have state per-capita incomes diverged recently? *Review of the Federal Reserve Bank of St. Louis*, 70:24–36.

Cowell, F. A. (1995). *Measuring Inequality*. Prentice Hall, London.

Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, pages 1241–1299.

Cressie, N. A. C. (1991). *Statistics for spatial data*. New York.

Cuadrado-Roura, J., Garcia-Greciano, B., and Raymond, J. (1999). Regional convergence in productivity and productivity structure: The Spanish case. *International Regional Science Review*, 22:35–53.

Deaton, A. (2003). Health, inequality, and economic development health, inequality, and economic development health, inequality, and economic development. *Journal of Economic Literature*, 41(1):113–158.

Delaney, D. (2001). The boundaries of responsibility: Interpretations of geography in school desegregation. In Blomley, N., Delaney, D., and Ford, R. T., editors, *The Legal Geographies Reader: Law, Power, and Space*, pages 54–68. Blackwell Publishers.

Diez-Roux, A., Link, B., and Northridge, M. (2000). A multilevel analysis of income inequality and cardiovascular disease risk factors. *Social science & medicine*, 50(5):673–687.

Drennan, M. P. and Lobo, J. (1999). A simple test for convergence of metropolitan income in the United States. *Journal of Urban Economics*, 46:350–359.

Drennan, M. P., Lobo, J., and Strumsky, D. (2005). Unit root tests of sigma income convergence across US metroplotan areas. *Journal of Economic Geography*, 4:583–595.

Duncan, O. and Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, 20(2):210–217.

Duque, J. C. (2004). *Design of Homogenous Territorial Units. A Methodological Proposal and Applications*. PhD thesis, Universitat de Barcelona.

Durlauf, S. N. (2004). A theory of persistent income inequality. *Journal of Economic Growth*, 1(1):75–93.

Epple, D. (2003). Modeling population stratification across locations: an overview. *International Regional Science Review*, 26(2):189–196.

Epple, D. and Platt, G. J. (1998). Equilibrium and local redistribution in an urban economy when households differ in both preferences and incomes. *Journal of Urban Economics*, 43(1):23–51.

Epple, D. N. and Romano, R. (2003). Neighborhood schools, choice, and the distribution of educational benefits. In *The economics of school choice*, pages 227–286. University of Chicago Press.

Fan, C. C. and Casetti, E. (1994). The spatial and temporal dynamics of U.S. regional income inequality. *Annals of Regional Science*, 28:177–196.

Fan, C. C. and Casetti, E. (1997). The spatial and temporal dynamics of us regional income inequality, 1950-1989. *The Annals Regional Science*, 28:177–196.

Foster, J. E. and Sen, A. (1997). *On Economic Inequality*. Oxford University Press, New Delhi.

Friedman, M. (1992). Do old fallacies ever die? *Journal of Economic Literature*, XXX:2129–2132.

Galbraith, J. K. and Hale, J. T. (2008). State income inequality and presidential election turnout and outcomes. *Social Science Quarterly*, 89(4):887–901.

Getis, A. (1991). Spatial interaction and spatial autocorrelation: a cross product approach. *Environment and Planning A*, 23(9):1269 – 1277.

Getis, A. (2008). A history of the concept of spatial autocorrelation: a geographer's perspective. *Geographical Analysis*, 40(3):297–309.

Gezici, F. and Hewings, G. (2004). Regional convergence and the economic performance of peripheral areas in Turkey. *Review of Urban and Regional Development Studies*, 16(2):113–132.

Gladwell, M. (1998). The Pima Paradox. *The New Yorker*.

Hubert, L. J., Golledge, R. G., and Constanza, C. M. (1981). Generalized procedures for evaluating spatial autocorrelation. *Geographical Analysis*, 13:224–231.

Isard, W. (1951). Interregional and regional input-output analysis: a model of a space-economy. *The review of Economics and Statistics*, 33(4):318–328.

Janikas, M. (2006). *Comparative Regional Income Dynamics: Clustering, Scale, and Geocomputation*. PhD thesis, San Diego State University and University of California, Santa Barbara.

Janikas, M. V. and Rey, S. J. (2005). Spatial clustering, inequality, and income convergence. *Région et Développement*, 25:41–64.

Jargowsky, P. and Kim, J. (2005). A measure of spatial segregation: The generalized neighborhood sorting index. Technical report, University of Texas at Dallas.

Jarkowsky, P. (1997). *Poverty and Place: Ghettos, barrios, and the American city*. Russel Sage Foundation, New York.

Jian, T., Sachs, J. D., and Warner, A. M. (1996). Trends in regional inequality in china. *China Economic Review*, 7(1):1–21.

Kanbur, R. and Zhang, X. (1999). Which regional inequality? The evolution of rural-urban and inland-coastal inequality in China, 1983-1995. *Journal of Comparative Economics*, 27:686–701.

Kennedy, B., Kawachi, I., Glass, R., and Prothrow-Stith, D. (1998). Income distribution, socioeconomic status, and self rated health in the united states: multilevel analysis. *BMJ: British Medical Journal*, 317(7163):917.

Kennedy, P. (2003). *A Guide to Econometrics*. The MIT Press, Cambridge, MA, fifth edition.

Kozol, J. (1991). *Savage Inequalities*. Crown, New York.

Kozol, J. (2005). *The Shame of the Nation: The Restoration of Apartheid Schooling in America*. Crown Publishers.

Krieger, N., Williams, D. R., and Moss, N. E. (1997). Measuring social class in us public health research: Concepts, methodologies, and guidelines public health research:measuring social class in us public health research: Concepts, methodologies, and guidelines. *Annu. Rev. Public Health*, 18:341–378.

Krugman, P. and Venables, A. (1995). Globalization and the Inequality of Nations. *The Quarterly Journal of Economics*, 110(4):857–880.

Le Gallo, J. and Ertur, C. (2002). Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980-1995. *Papers in Regional Science*, Forthcoming.

Levernier, W., Partridge, M. D., and Rickman, D. S. (1998). Differences in metropolitan and nonmetropolitan us family income inequality: A cross-county comparison. *Journal of Urban Economics*, 44(2):272–290.

Levernier, W., Partridge, M. D., and Rickman, D. S. (2000). The causes of regional variations in us poverty: A cross-county analysis. *Journal of Regional Science*, 40(3):473–497.

Levernier, W., Rickman, D. S., and Partridge, M. D. (1995). Variation in us state income inequality: 1960-1990. *International Regional Science Review*, 18(3):355–378.

Lichtenberg, F. R. (1994). Testing the convergence hypothesis. *The Review of Economics and Statistics*, 76(3):576–579.

Lieberson, S. and Carter, D. K. (1982). A model for inferring the voluntary and involuntary causes of residential segregation. *Demography*, 19(4):511–526.

Lucas, R. E. (2002). *Lectures on Economic Growth*. Harvard University Press.

Massey, D. (1979). In what sense a regional problem? *Regional Studies*, 13(2):233–243.

Massey, D. S. and Denton, N. A. (1987). Trends in the residential segregation of blacks, hispanics, and asians: 1970-1980. *American Sociological Review*, pages 802–825.

Massey, D. S. and Denton, N. A. (1993). *American Apartheid: Segregation and the Making of the American Apartheid: Segregation and the Making of the Underclass*. Harvard University Press, MA.

Morrill, R. L. (1991). On the measure of spatial segregation. *Geography Research Forum*, pages 25–36.

Mossi, M. B., Aroca, P., Fernandez, I., and Azzoni, C. R. (2003). Growth dynamics and space in Brazil. *International Regional Science Review*, 26(3):393–418.

Nissan, E. and Carter, G. (1999). Spatial and temporal metropolitan and non-metropolitan trends in income inequality. *Growth and Change*, 30(3):407–415.

Oppenshaw, S. and Taylor, P. (1991). The modifiable areal unit problem. In Wrigley, N. and Bennett, R., editors, *Quantitative geography*, pages 60–69. Routledge and Kegan Paul, Cancer, 1991:126–58. Lon- don, United Kingdom.

Partridge, J. S., Partridge, M. D., and Rickman, D. S. (1998). State patterns in family income inequality. *Contemporary Economic Policy*, 16(3):277–294.

Partridge, M. D. and Rickman, D. S. (2007). Persistent rural poverty: Is it simply remoteness and scale? *Applied Economic Perspectives and Policy*, 29(3):430–436.

Partridge, M. D., Rickman, D. S., and Levernier, W. (1996). Trends in us income inequality: evidence from a panel of states. *The Quarterly Review of Economics and Finance*, 36(1):17–37.

Quah, D. T. (1996a). Empirics for economic growth and convergence. *European Economic Review*, 40:1353–1375.

Quah, D. T. (1996b). Regional convergence clusters across Europe. *European Economic Review*, 40:951–958.

Quah, D. T. (2002). One third of the world's growth and inequality. London School of Economics, CEPR DP3316.

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.

Rawls, J. (1999). *A Theory of Justice*. Harvard University Press.

Rawls, J. and Kelly, E. (2001). *Justice as fairness: A restatement*. Harvard University Press.

Reardon, S. and O'Sullivan, D. (2004). Measures of Spatial Segregation. *Sociological Methodology*, 34:121–162.

Rey, S. (2004a). Spatial Analysis of Regional Income Inequality. In *Spatialy Integrated Social Science*, pages 280–299. Oxford University Press, USA.

Rey, S. and Dev, B. (2006a). Sigma-convergence in the presence of spatial effects. *Papers in Regional Science*, 85(2):217–234.

Rey, S. and Dev, B. (2006b). Sigma-convergence in the presence of spatial effects. *Papers in Regional Science Papers in Regional Science*, 85(2):217–234.

Rey, S. J. (2001). Spatial empirics for economic growth and convergence. *Geographical Analysis*, 33(3):195–214.

Rey, S. J. (2002). Spatial dependence in the evolution of regional income distributions. In Getis, A., Múr, J., and Zoeller, H., editors, *Spatial econometrics and spatial statistics*. Palgrave, Hampshire.

Rey, S. J. (2004b). Spatial dependence in the evolution of regional income distributions. In Getis, A., Múr, J., and Zoeller, H., editors, *Spatial econometrics and spatial statistics*. Palgrave, Hampshire.

Rey, S. J. and Dev, B. (1997). Integrating econometric and input-output models in a multiregional context. *Growth and Change*, 28:222–243.

Rey, S. J. and Folch, D. C. (2011). Impact of spatial effects on income segregation indices. *Computers, Environment and Urban Systems*, 35(6):431–441.

Rey, S. J. and Janikas, M. V. (2005). Regional convergence, inequality, and space. *Journal of Economic Geography*, 5(2):155–176.

Rey, S. J. and Montouri, B. D. (1999). U.S. regional income convergence: A spatial econometric perspective. *Regional Studies*, 33:143–156.

Rivas, M. G. (2007). The effects of trade openness on regional inequality in mexico the effects of trade openness on regional inequality in mexico. *The Annals of Regional Science*, 41(3):545–561.

Sala-i-Martin, X. (2006). The world distribution of income: Falling poverty and ... convergence, period. *Quarerly Journal of Economics*, 121(2):351–397.

Sala-i-Martin, X. X. (1996). The classical approach to convergence analysis. *The Economic Journal*, 106(437):1019–1036.

Sánchez-Reaza, J. and Rodríguez-Pose, A. (2002). The impact of trade liberalization on regional disparities in mexico. *Growth and Change*, 33(1):72–90.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of mathematical sociology*, 1(2):143–186.

Sen, A. (1977). Starvation and exchange entitlements: a general approach and its application to the great Bengal famine. *Cambridge Journal of Economics*, 1(1):33–59.

Sen, A. (1995). *Inequality reexamined*. Oxford University Press.

Sen, A. (1997). From income inequality to economic inequality. *Southern Economic Journal*, 64(2):384–401.

Sen, A. and Foster, J. (2006). *On Economic Inequality*. Oxford University Press.

Shorrocks, A. F. (1982). Inequality decomposition by factor components. *Econometrica*, 50(1):193–211.

Smith, A. (1776). An inquiry into the nature and causes of the wealth of nations. In Campell, R. and Skinner, A., editors, *An Inquiry into the Nature and Causes of the Wealth of Nations*. Clarendon Press, 1976 edition.

Solow, R. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70:65–94.

Subramanian, S. et al. (2004). The relevance of multilevel statistical methods for identifying causal neighborhood effects. *Social Science and Medicine*, 58(10):1961–1968.

Subramanian, S., Kawachi, I., and Kennedy, B. (2001). Does the state you live in make a difference? multilevel analysis of self-rated health in the us. *Social Science & Medicine*, 53(1):9–19.

Temkin, L. (1993). *Inequality*. Oxford University Press.

Tiebout, C. (1956). A pure theory of local expenditures. *Journal of Political Economy*, 64(5):416–424.

Tortosa-Ausina, E., Perez, F., Mas, M., and Goerlich, F. (2005). Growth and Convergence Profiles in the Spanish Provinces(1965-1997)*. *Journal of Regional Science*, 45(1):147–182.

Tukey, J. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.

U.S.551 (2007). Parents Involved in Community Schools v. Seattle School District. *Supreme Court of the United States*.

U.S.717 (1974). Milliken v. Bradley, 418. *Supreme Court of the United States*.

Walzer, M. (1983). *Spheres of justice: A defense of pluralism and equality*. Basic Books.

White, M. J. (1983). The measurement of spatial segregation. *American Journal of Sociology*, pages 1008–1018.

Wilson, W. (1987a). The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy.

Wilson, W. J. (1987b). *The Truly Disadvantaged*. University of Chicago Press, Chicago.

Wong, D. W. (1993). Spatial indices of segregation. *Urban studies*, 30(3):559–572.

Xue, J. (1997). Urban-rural income disparity and its significance in china. *Hitotsubashi Journal of Economics*, 38(1):45–59.

Yamamoto, D. (2008). Scales of regional income disparities in the USA, 1955 2003. *Journal of Economic Geography*, 8(1):79.

Zhang, J. (2004). A dynamic model of residential segregation. *Journal of Mathematical Sociology*, 28(3):147–170.