A spatial decomposition of the Gini coefficient

Article *in* Letters in Spatial and Resource Sciences · July 2012

DOI: 10.1007/s12076-012-0086-z

Sergio Rey
University of California, Riverside

Richard John Smith
Wayne State University

A spatial decomposition of the Gini coefficient*

Sergio J. Rey

GeoDa Center for Geospatial Analysis and Computation
School of Geographical Sciences and Urban Planning
Arizona State University

Richard J. Smith

School of Social Work
Wayne State University

February 20th, 2012

Abstract

The paper introduces a spatial decomposition of the Gini coefficient that supports the detection of spatial autocorrelation conjointly with an indicator of overall inequality. An additive pairwise decomposition based on a spatial weights matrix partitions inequality between observations that are geographically neighbors and those that are not. A framework for inference on the spatial decomposition is also suggested. The statistical properties of the decomposition measure are evaluated in a Monte Carlo simulation and an empirical illustration involving per capita income inequality in the US states is also provided.

Keywords: Spatial inequality, spatial autocorrelation

*This research was funded in part by NSF Award OCI-1047916, SI2-SSI: CyberGIS Software Integration for Sustained Geospatial Innovation. This is a working paper not for citation. The final publication is available at www.springerlink.com and may be cited as Rey, S., & Smith, R. (2012). A spatial decomposition of the Gini coefficient. *Letters in Spatial and Resource Sciences*, 1-16. Available at http://dx.doi.org/doi:10.1007/s12076-012-0086-z

1 Introduction

This paper proposes a way to combine the measurement of space and inequality. Inequality measures are measures of dispersion that tell us about the distribution of a value. For example, the Gini coefficient was developed to measure income inequality. It ranges from zero to one where zero, perfect equality, means all persons have the same income and one, perfect inequality, implies that one person in the population has all the income and everyone else has none. The Gini coefficient is a whole-map, locationally invariant measure of inequality. By whole map we mean that a single index serves as a summary measure quantifying the extent of the inequality in the attribute of interest. Locationally invariant implies that this whole-map measure is insensitive to the absolute and relative position of the value observations across the map. In other words, the measure remains unchanged if the values are permuted over the map. The Gini can tell us if inequality is happening, but not where it is happening within the region (Silber, 1989; Dawkins, 2006, 2004; Arbia and Piras, 2009; Arbia, 2001).

In order to get at the within region inequality, one needs a measure of concentration to determine the degree to which values are similar to others in or around a particular sub-region. If concentration increases over time and forms a bimodal distribution (i.e. twin peaks), where one sub-region has high values and another sub-region has low values, this is called polarization(Arbia and Piras, 2009). Early measures of concentration also locationally invariant and are insensitive to neighbors. They are better described as measures of disproportionality because they do not have a spatial component(Bickenbach and Bode, 2008). The Krugman (1991) locational Gini index is an example of a concentration measure of disproportionality that does not have a spatial component. It is calculated from the ratio of the share of a given industry employment in a region over the region's share of national employment. This measure ranges from zero, which means that the industry is not concentrated, to .5, which means that the industry is concentrated. Our proposed measure is motivated by the need for spatial measures of concentration and polarization.

Researchers have been making advances in considering space and inequality by applying the Gini coefficient to segregation. While many measurement problems plague the study of inequality and segregation, this paper will be concerned primarily with the checkerboard problem (White, 1983) and the modifiable areal unit problem (MAUP). Segregated units are said to cluster when they are adjacent in a single part of a region, but are not clustered if they are distributed like a checkerboard. The index of dissimilarity, a measure of the unevenness of segregation, is not sensitive to segregation that is evenly distributed across the region. Spatial autocorrelation (e.g. Moran's I) measures the degree of clustering, but not the degree of unevenness in the distribution. Spatial autocorrelation measures are also sensitive to the assumption of spatial stationarity in the mean and variance. The MAUP also plagues most areal research in that an arbitrary boundary shift may change a measure in cases where the within area distribution is heterogeneous. Also, the scale or spatial precision of the units may also change the measure. Our measure makes progress with the checkerboard problem by exploiting the contiguity matrix to create a decomposition of the Gini.

Some segregation researchers have suggested a joint measure of spatial segregation that weights a non-spatial segregation index by a contiguity matrix to reflect topological relationships between area units (Morgan, 1983a,b; Morrill, 1991; Wong, 1993). These measures that rely on a dissimilarity index for the segregation component generally fail to satisfy the principle of transfers which requires that a move of households from one neighborhood to another neighborhood where their own race is less represented should always result in a decline in measured segregation. In general, the Gini coefficient satisfies this principle(Allison, 1978; Massey and Denton, 1988a). The Wong (1993) decomposition is hierarchical and addresses the MAUP type problem by dis-aggregating from regions to tracts and to blocks. This solution is a different than what we are proposing.

The purpose of a Gini decomposition is to compare spatial patterns of segregation across regions with differing levels of segregation. Dawkins (2004) develops a Spatial Gini decomposition measure that relies on a partitioning of population into mutually exclusive subgroups (e.g. black, white) as well as multiple spatial scales, such as cities and neighborhoods. In our case the Gini decomposition is pairwise disjoint and mutually exclusive, not individually disjoint. That is for each pair i, j, membership in one of two groups is determined. Dawkins (2007) and our proposed decomposition both may be extended to spatial income inequality.

The plan of our paper is as follows. Section 2.1 of our paper reviews inequality measures and Section 2.2 describes criteria needed for a useful inequality measure. Because researchers have been making advances in considering space and inequality by applying the Gini coefficient to segregation, we discuss these recent advances in Section 2.3. Section 3 introduces the formal definition of the Gini coefficient. Our global decomposition is defined formally in Section 3.1 while a local Spatial Gini coefficient is defined in Section 3.2. We discuss the inference for these measures in Section 3.3. Section 4 provides the results of a simulation and we conclude with possible extensions and applications for this measure in Section 5.

2 Previous research

2.1 Inequality measures

Inequality and segregation have been of interest to economists, sociologists, geographers and policy makers. The two phenomena are related and have normative implications. For example, sociologists and welfare economists argue that under the assumption of the diminishing returns to utility of income, high levels of income inequality lead to a loss in social welfare (Allison, 1978; Clark et al., 1981). The policy implication would be to endorse a progressive income tax that has higher rates for those with the highest incomes. The tax revenues are transferred to those with lower incomes directly or through provision of public goods. Libertarian economists such as Friedman and Friedman (2002) disagree that income inequality is a concern because they argue that it is more important to protect personal freedom. In a perfect market, the observed income inequality would simply be a function of individual preferences for risk-taking, work and leisure and any redistribution would be suboptimal. The policy implication from this perspective would be to reject a progressive income tax, but have a negative income tax for those in poverty. While the later two examples are underpinned primarily by value laden assumptions of human nature, the normative implications of aggregate income inequality of regions compared to each other are further complicated by the movement of labor and firms. Investments in a place with high poverty, for example, may attract residential investment from persons with higher incomes and thus increase income inequality. Likewise, investments in firms in areas with high poverty may end up being capitalized in land values and only serve to increase costs for residents that in turn creates an incentive to leave (Levine, 1999; Quigley, 1994). Accordingly, transfers across the income distribution and space must be carefully targeted in order to increase social welfare to account for people in places (Ladd, 1994).

2.2 Criteria for a useful inequality measure

Good measures of inequality satisfy two criteria: a) they have to have scale invariance (i.e. compositional invariance) such that multiplying each member of a population by a constant should not change the level of inequality; and b) they should be sensitive to transfers, such that a transfer of income from a low income person to someone with more income should increase inequality (Allison, 1978). Such measures ideally should be able to be decomposed into within group and between group differences in inequality to allow comparisons. Allison (1978) notes that the Gini Index, coefficient of variation and Theil index satisfy these two criteria.

In contrast, prior to improvements in microprocessing, the Gini was not decomposable due to the complexity of the calculation. Other scholars such as Shorrocks and Wan (2005) add that because the Gini is not decomposable, it is not subgroup consistent. In other words, an increase in inequality in each subregion should lead to an increase in inequality as a whole. This property holds for the Theil and not the Gini. In other words, without a decomposition one could calculate a Gini for each subregion, but a reduction in the inequality in one region may not in fact lower the Gini for the full region. However, Silber (1989) created an influential decomposition of the Gini using matrix algebra that is able to identify the within and between group differences in inequality provided the groups are mutually exclusive and of adequate size.

2.3 Spatial applications of the Gini to segregation

The use of the Gini Index has been applied to the study of segregation and this in turn has led to a spatial decomposition of the measure. Segregation is caused by discrimination and income inequality and in turn replicates inequality. Classic sociological and economic segregation literature usually applies to residential segregation in neighborhoods by race (Duncan and Duncan, 1955; James and Tauber, 1985; Massey and Denton, 1988a), but the concept and measures have been generalized to any kind of group distributed across any kind of units. For example, one could apply the concept to gender segregation in the workplace by office or functional unit. The accurate calculation of within group and between group differences is needed to make the case for targeted policy interventions (Kanbur, 2006). The most obvious link between segregation and income inequality is articulated as the spatial mismatch hypothesis or jobs housing imbalance (Kain, 1968; Wilson, 1987). In other words, as jobs left the inner city, low income African Americans were at a comparative disadvantage to obtain these jobs because of increased commuting costs and the inability to move closer to jobs due to housing discrimination.

Inequality and segregation measures are by definition measures of dispersion as opposed to central tendency (e.g. mean, median, mode). While the Gini index is one of the measures used to quantify segregation, the preferred single measure used to document correlates of and changes in the segregation of African Americans and Whites is the Dissimilarity Index (James and Tauber, 1985). The research in the 1950s and 1960s had direct normative implications as policy makers and the courts began to question several decades of policies and practices to keep races separated. It was assumed that as civil rights legislation and court decisions overturned segregationist policy and racial covenants in deeds that neighborhoods would naturally integrate. Schelling's (1969) Nobel Prize winning tipping point model called into question the belief that segregation need be caused by preferences for it. He demonstrated using a simulation that even in situations when both whites and African American preferred integration, a segregated landscape would still emerge when a neighborhoods passed a threshold or tipping point. Methodological advancement in the 1980s and 1990s introduced multidimensional measures of segregation by using factor analysis to collapse a set of 20 segregation measures into the dimensions of *evenness*, *exposure*, *concentration*, *centralization*, and *clustering*. The Dissimilarity and Gini indices factored into the evenness dimension (Massey and Denton, 1988a,b; Massey et al., 1996).

Solutions to the spatial methodological problems related to segregation measures were not given tentative solutions until the accessibility of Geographic Information Systems (GIS) (Wong, 2003a,b). For example, Reardon and Firebaugh (2002) propose that decompositions of segregation should be additive for both groups and spatial units. Reardon and O'Sullivan (2004) recommend that the most satisfactory solutions to the checkerboard problem and the modifiable areal unit problem have not yet surfaced, except to say that one should use point data if available. Grannis (2005, 2002) extends the discussion of identifying the right unit of analysis by theorizing about the configuration of space that realistically shapes residential segregation and by implication inequality. He posits a unit of observation called the T-community, the area bounded by high volume streets. This theory builds on work that documents how residents are less likely to cross a street with faster traffic and that neighborhoods are often bounded by these streets. This is an interesting solution to MAUP in that he has a compelling theoretical explanation as to why the unit of observation is a good proxy for the unit of analysis.

With regard to the checkerboard problem, Dawkins has made progress developing a Spatial Gini index using the Silber (1989) decomposition method. This Spatial Gini[1] has been applied to income (Dawkins, 2007) and residential segregation (Dawkins, 2006, 2004). This decomposition satisfies the principle of transfers, scale invariance and the Lorenz criterion. He introduces two ways to introduce space into the Gini. The first rank orders by units of observation and then replaces the vector with its nearest neighbor in space. When using Census tract level data, Dawkins takes the centroid of the tract as the location in space. The second implementation invokes a monocentric city assumption and reorders the vector by distance from the central business district. For each version of his spatial Gini, the Spatial component and the non-spatial component sum to the Gini. Dawkins benchmarks his measures against (Massey and Denton, 1988a). While Dawkins, like Schelling, chose a one dimensional solution to the checkerboard problem, we propose a two dimensional solution that does not depend upon a monocentric city assumption. This should better fit data in poly-centric metropolitan regions.

In a spatial context, the Gini coefficient can be reinterpreted as measuring the amount of concentration of the values on the map. This is not the same as measuring spatial autocorrelation, or the lack of randomness in the geographical distribution. More specifically, the Gini asks if the values are concentrated (or dispersed) across the observations, irrespective of the location of these areas of concentration relative to each other. Measures of spatial autocorrelation consider the spatial distribution of the attribute values relative to neighboring locations. Unlike concentration measures such as the Gini, measures of spatial autocorrelation are not locationally invariant.

Both measures of inequality and autocorrelation can suffer from the equifinality problem. That is different spatial distributions can yield the same value for an inequality measure, while maps with different levels of inequality can result in the same level of spatial autocorrelation. As Arbia and Piras (2009) correctly point out, this suggests that neither a-spatial inequality measures or spatial autocorrelation measures alone are appropriate for measuring *spatial concentration*.

An additional complication is that conventional measures of inequality may be sensitive to spatial dependence. Rey and Dev (2006) show that the magnitude of the sample variance, a common indicator used to study regional income convergence, is a function of spatial dependence, mean and variance heterogeneity, as well as an a-spatial measure of dispersion.

The relationship between measures of inequality and spatial autocorrelation have been noted in the literature beginning with Arbia (2001) who suggests a number of approaches towards combining measures of spatial autocorrelation, such as Moran's I or the Getis-Ord statistic, together with an inequality measure, such as the Gini coefficient, to develop a joint index. This approach has been criticized by Bickenbach and Bode (2008) as ad-hoc, and they suggest a general taxonomy of measures for disproportionality, concentration, specialization and localization.[2]

In this paper we suggest an alternative approach towards considering the joint effects of

[1] Briant et al. (2010) use the term Spatial Gini in reference to the application of a Gini index for a variable (income, productivity) measured on spatial units.

[2] Related work developing joint measures of inequality and concentration is reported in Arbia and Piras (2009).

inequality and spatial autocorrelation that relies on a decomposition of the classic Gini coefficient. This is in the spirit of Arbia (2001) in that consideration of both autocorrelation and inequality is the focus, however, our decomposition reveals that rather than having to develop a new measure that combines a measure of inequality together with a measure of inequality, the classic Gini index actually nests a measure of spatial autocorrelation.

3 Spatial Gini coefficients

The Gini in relative mean difference form is given as:

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n^2 \bar{x}} \tag{1}$$

where x_i is the value for variable x observed at location $i = [1, 2, \ldots, n]$ and $\bar{x} = (1/n) \sum_i x_i$.

3.1 A global Spatial Gini coefficient

Our spatial decomposition exploits the fact that the sum off all pairwise differences can be rewritten as:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j| = \sum_{i=1}^{n} \sum_{j=1}^{n} (w_{i,j}|x_i - x_j| + (1 - w_{i,j})|x_i - x_j|) \tag{2}$$

where $w_{i,j}$ is an element of a binary spatial weights matrix expressing the neighbor relationship between locations i and j. Rewriting (1) gives us:

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}|x_i - x_j|}{2n^2 \bar{x}} + \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (1 - w_{i,j})|x_i - x_j|}{2n^2 \bar{x}} \tag{3}$$

The interpretation of this Spatial Gini coefficient[3] is that as the amount of positive spatial autocorrelation strengthens, the second term should grow relative to the first, since value similarity in space would be in effect.[4] For negative spatial autocorrelation, or the checkerboard case, the result is exact opposite (on a perfect lattice with rook contiguity) since the difference between non-neighbor pairs would be smaller than that between neighbor pairs. In either case, the degree of spatial autocorrelation can be assessed by comparison of the relative contributions of these two components.

In the case where the binary weights are row standardized:

$$wr_{i,j} = \frac{w_{i,j}}{\sum_j w_{i,j}} \tag{4}$$

the effect will be to shift more of the sum of absolute difference between pairs of x to the non-neighbor component relative to that when the unstandardized binary weights are used. More general forms of weights, such as inverse distance, may also be employed so long as they are first subject to row standardization prior to carrying out the decomposition in (3).

The weights play the role of determining membership of pairs of observations. For binary weights the membership is mutually exclusive. However, for row standardized weights, fuzzy membership obtains the value of the weight indicating degree of membership in the neighbor

[3]The original form of this statistic was first implemented in the package STARS: Space-Time Analysis of Regional Systems (Rey and Janikas, 2006), but until this paper has not been formally described.

[4]The numerator of the first term for (3) is similar to the spatial autocorrelation coefficient suggested by Sokal et al. (1993).

pair group, and the complement of this weight reflecting degree of membership in the non-neighbor pair group.

Spatial autocorrelation coefficients only implicitly consider the two groups simultaneously. That is, the measures compare the covariation between neighbor pair observations only.By decomposing the Gini coefficient into these two pair groups, we close the loop as it were between a measure of inequality and a measure of spatial concentration. Finally, we note the equivalence between (3) and (1) when $w_{i,j} = 1\ \forall i, j$ since the second component of (3) will be equal to 0.

3.2 Inference

Two general approaches have been suggested in the literature for inference on inequality measures. The first relies on analytical results to derive the sampling distribution of the index under a specific null hypothesis. Alternatively, bootstrap and block bootstrap has been suggested by Bickenbach and Bode (2008) as an approach for inference for the inequality measures in a regional context. Both the analytical and bootstrap approaches rely on the assumptions that the sampling units are fixed and independent. This is problematic given that spatial autocorrelation tends to be the rule rather than the exception in empirical regional data sets because neighboring units may contain spill over effects or within unit hetrogeneity.

As an alternative to these approaches, Rey (2004) has suggested the use of random spatial permutations as a basis for inference on spatial inequality indices. This approach randomly reassigns the observed values to different locations to develop a sampling distribution under the null that the variate is randomly distributed in space. A large number of such permutations are carried out, and for each one the spatial inequality measure is obtained. The value of the original inequality statistic is then compared to the computationally based distribution to assess its statistical significance.

It is important to note that the permutation approach allows for inference on the spatial decomposition of the Gini coefficient in (3), and not the value of the general Gini coefficient from (1). As mentioned earlier, the latter is a whole map statistic that is insensitive to the spatial arrangement of the data. In what follows we thus focus on the statistical significance of the decomposition alone. Future work will examine approaches for the joint distribution of the inequality measure and its spatial decomposition.

4 Illustration

To illustrate the Spatial Gini decomposition we carry out two sets of analyses. The first consists of a small Monte Carlo simulation designed to explore the properties of the Spatial Gini as a test for spatial autocorrelation in a controlled setting. With an understanding of these properties in hand, we then move to an application of the decomposition in an analysis of US per capita incomes over the period 1969-2009.

4.1 Monte Carlo simulation

The Monte Carlo simulation allows for a comparison of the properties of the Spatial Gini as a test for spatial autocorrelation against those of Moran's I - the most commonly encountered test for spatial autocorrelation. The Gini based test for autocorrelation is defined as:

$$SG = \frac{\sum_{j=1}^{n}(1 - w_{i,j})|x_i - x_j|}{2n^2\bar{x}G} \tag{5}$$

where G is the Gini from (3). SG can be interpreted as the share of overall inequality that is associated with non-neighbor pair of locations. Inference on this statistic relies on random spatial permutations of the data. More specifically, the value of (5) is first obtained from the original data. Next the values are spatially permuted to simulate spatial randomness and the test statistic is calculated for this new map pattern. Additional permutations are carried out and the original value of the statistic is then compared to the distribution of values obtained from the randomly permuted data. The pseudo p-value for the observed test statistic is then defined as:

$$p(SG) = \frac{1+C}{1+M} \tag{6}$$

where C is the number of the $M = 999$ permutation samples that generated SG values that were as extreme as the observed SG value for the original data. The definition of extreme depends on whether one is considering a one or two-tailed test. In the former a directional test holds in which case the alternative hypothesis is that there is positive (negative) autocorrelation and values of the test statistic that exceed (are less than) the observed value from the original example contribute to C.

Moran's I is given as:

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j z_i W_{i,j} z_j}{\sum_i z_i z_i} \tag{7}$$

where $z_i = x_i - \bar{x}$ and $S_0 = \sum_i \sum_j W_{i,j}$. In our comparison, we also base inference for I on the random permutation approach.

We vary the amount of spatial autocorrelation along with both the sample size and level of inequality (variance) in the data generating process. For a given set of parameter values and sample size, we generate 1,000 realizations from an underlying data generating process (DGP). For each realization we apply both tests for spatial autocorrelation using a significance level of $\alpha = 0.05$.

For each realization we generate a spatially autocorrelated variate as follows

$$x = \rho W x + \epsilon \tag{8}$$

where x is an $n \times 1$ vector, W is an $n \times n$ binary spatial weights matrix and $\epsilon \tilde{} N(0, \sigma^2 I)$. W is defined using the rook definition of contiguity for a regular square lattice.

Three different sample sizes ($n = [25, 49, 81]$) are used. For each of these, four different levels of inequality ($\sigma_2 = [1, 4, 9, 16]$), and 10 levels of spatial autocorrelation ($\rho = [0.0, 0.1, \ldots 0.8, 0.9]$) are examined.

We examine both the power and the size of the Spatial Gini test for spatial autocorrelation across the various DGPs. Power is defined for the cases of $\rho \neq 0.0$ as the percentage of the samples for which the test results in a rejection of the null hypothesis of no spatial autocorrelation. When $\rho = 0.0$ the rejection frequency is the size of the test, or the probability of a Type I error. If the size of the test differs from the operational significance level employed ($\alpha = 0.05$) then the test is said to be biased.

Figures 1 and 2 display the rejection frequencies for the Spatial Gini and Moran's I tests of autocorrelation across the DGPs. As expected, the power of each test increases with the strength of the spatial autocorrelation (ρ) for all levels of inequality and sample size. For both of the tests, power is also higher for larger sample sizes, although this difference becomes smaller at the extreme values for ρ.

The relative power, measured as the ratio of the rejection frequencies, of the tests is portrayed in Figure 3. With few exceptions, the power of the Spatial Gini test is below that of Moran's I for a given DGP, although the differences shrink with increasing sample size and levels of autocorrelation. This is not surprising as Moran's I was designed with the sole purpose of testing for spatial autocorrelation, while the Spatial Gini we suggest is derivative and based on the traditional Gini measure of inequality.

Examination of the values underlying Figures 1 and 2 reveals some evidence of small sample bias for each test.[5] For the spatial Gini, the empirical rejection frequencies are significantly different from 0.05 for the following cases: $(n = 25, \sigma^2 = 1)$, $(n = 25, \sigma^2 = 4)$, $(n = 49, \sigma^2 = 1)$, $(n = 49, \sigma^2 = 4)$, $(n = 81, \sigma^2 = 9)$, and $(n = 81, \sigma^2 = 16)$, and in each instance the rejection frequency exceeds 0.05. For Moran's I, significant differences are found in the cases of: $(n = 25, \sigma^2 = 1)$, $(n = 25, \sigma^2 = 16)$, $(n = 49, \sigma^2 = 1)$, $(n = 49, \sigma^2 = 9)$, $(n = 81, \sigma^2 = 9)$, and $(n = 81, \sigma^2 = 16)$ with the empirical rejection frequencies exceeding 0.05.

A more formal evaluation of the relative power of the two tests is contained in Table 1. Here the differences in the rejection frequencies (Moran's I - Spatial Gini) are tested against 0 for all DGPs. Interestingly, when $\rho = 0.0$ the rejection frequencies for the two tests are not significantly different from 0. In other words, the size properties of the two tests are identical.

Turning to differences in power, the two tests perform differently for moderate levels of spatial dependence for the larger sample sizes, and to a greater extent for $\rho = 0.8$ and $n = 25$. When dependence is strongest ($\rho = 0.9$), the power of the Spatial Gini as a test for autocorrelation matches that of the traditional test.

4.2 US per-capita incomes

Figure 4 displays the various measures of inequality for the 48 states over the period 1969-2009. The overall pattern of global inequality (G) displays four peaks (1965, 1987, 2000, and 2007). The non-neighbor inequality component (NG) tends to follow this overall pattern, indeed the two measures have a correlation of 0.991. The non-neighbor component accounts for an average of 0.981 of overall inequality (standard deviation of 0.008).

At first glance the dominance of the non-neighbor inequality component might suggest that our decomposition may be of limited value. However, also shown in Figure 4 are the mean value and the 95 percent confidence intervals for the non-neighbor inequality component under the null hypothesis that per-capita incomes are randomly distributed in space. For each year in the sample, the observed value of NG exceeds its expectation under the null, and in each case these differences are statistically significant as the observed values are never contained in the confidence bounds.

The results from Figure 4 suggest that the decomposition we are proposing may be capable of providing insights to the extent of spatial dependence along side a global measure of inequality. To explore this issue further, we compare the indications of spatial dependence based on the classic Moran's I statistic with one based on our decomposition approach. More specifically, we utilize 999 random permutations of the per-capita income values for each year and compare the observed values for Moran's I and the non-neighbor (NG) component of inequality to their distributions under the null of spatially random incomes.

Figure 5 displays the pseudo p-values for Moran's I and the NG component for the US per-capita income values. While the p-values for I are typically smaller than those for NG, both indicators generate statistically significant values for all years (at $\alpha = 0.05$). This suggests that our decomposition approach may have utility as a joint measure of inequality and spatial autocorrelation. Of course these results are based on an empirical regional time series where the underlying level of autocorrelation (and inequality) are unknown. They are qualitatively in agreement with the findings of our Monte Carlo simulations, however.

[5] The 95% confidence interval for the rejection frequencies under the null ($\rho = 0.05$) is based on the distribution under the null: $z = \frac{\hat{\alpha} - \alpha}{\sqrt{(\alpha(1-\alpha))/(M+1)}}$. With $\alpha = 0.05$ and $M = 999$ this yields and interval of $[0.0383, 0.0617]$.

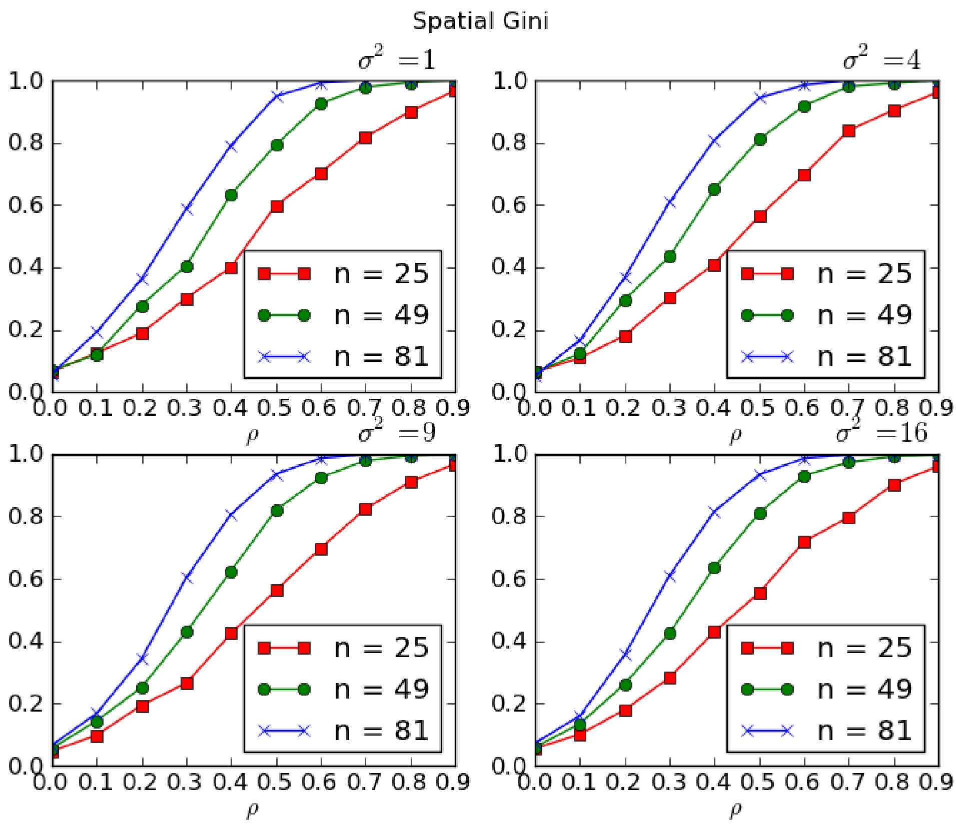

Figure 1: Power Function for Spatial Autocorrelation: Spatial Gini

5 Conclusion

In this paper, we propose a spatial decomposition of the Gini coefficient that supports the detection of spatial autocorrelation conjointly with an indicator of overall inequality. An additive pairwise decomposition based on a spatial weight matrix partitions inequality between observations that are geographically neighbors and those that are not. A framework for inference on the spatial decomposition is also suggested. The results of our simulation experiment indicate that the decomposition behaves similarly to a common indicator of spatial autocorrelation. Furthermore, in an empirical case study of US income dynamics, our decomposition provides that same qualitative and quantitative analysis on the level of spatial autocorrelation. Consequently, the approach has the advantage of providing insights into both inequality and autocorrelation dynamics using a single measure, rather than combining different indicators for each aspect.

While the spatial decomposition is promising, there are a number of extensions that can be explored. First the measure could also be employed in other instances where the dissimilarity index had been applied. For example, the spatial mismatch index measures the imbalance of jobs and housing in a regional economy. By using the spatial decomposition of the Gini index,

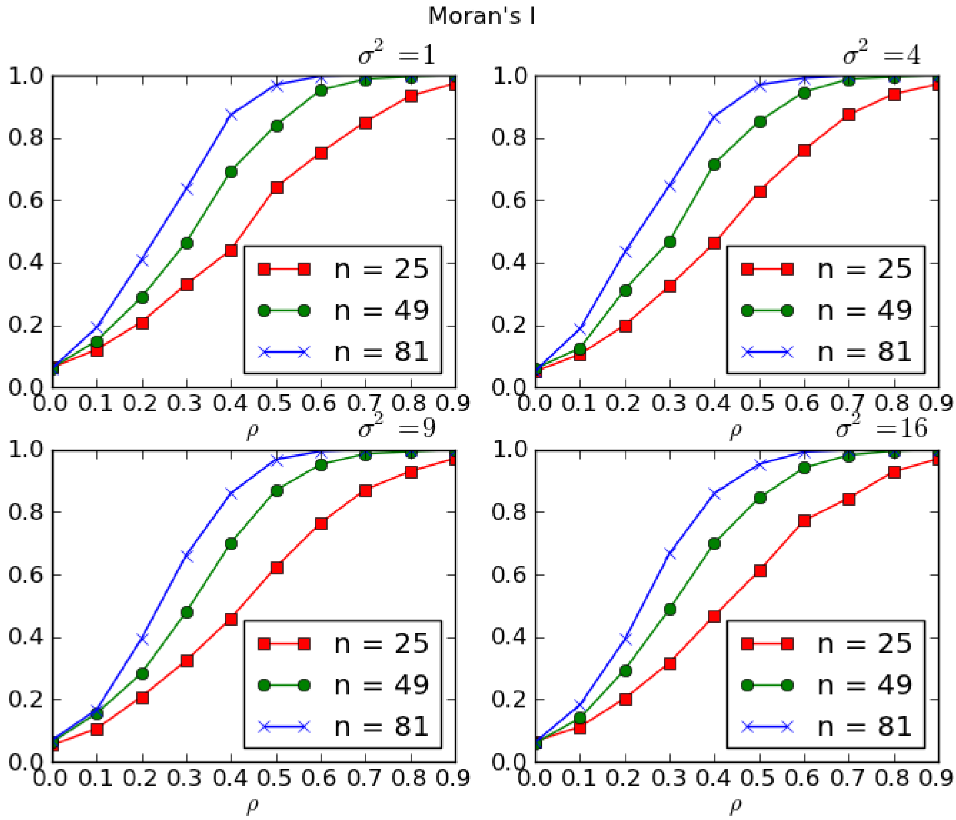

Figure 2: Power Function for Spatial Autocorrelation: Moran's I

one can account for the inequality and the autocorrelation of jobs and housing in areal subunits of a regional economy. Use of the Spatial Gini will be an improvement from the dissimilarity index because it respects the principle of transfers. Second, extending the inferential framework to consider both the level of autocorrelation in the decomposition together with the magnitude of the overall degree of inequality remains an important goal. Closely related to this is the ability to expand this framework to support testing for temporal changes in either the level of autocorrelation, inequality or their joint movement over time.

References

Allison, P. (1978). Measures of inequality. *American Sociological Review*, 43:865–880.

Arbia, G. (2001). The role of spatial effects in the empirical analysis of regional concentration. *Journal of Geographical Systems*, 3:271–281.

Arbia, G. and Piras, G. (2009). A new class of spatial concentration measures. *Computational Statistics and Data Analysis*, 53:4771–4481.

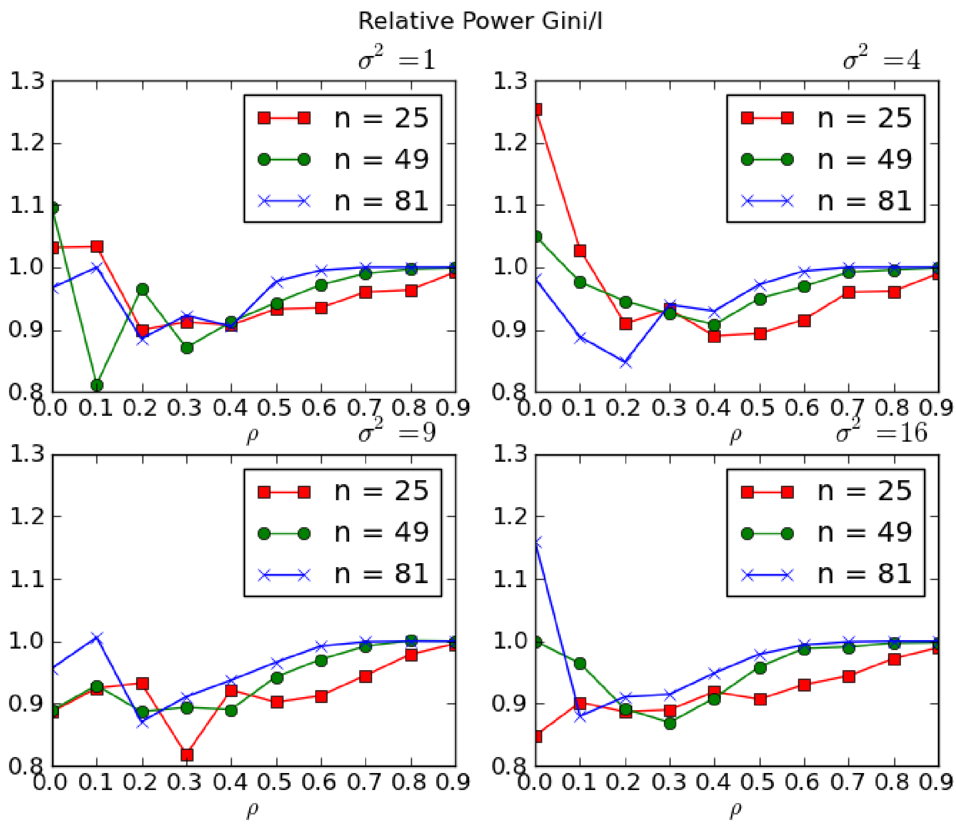

Figure 3: Relative Power Function for Spatial Autocorrelation: Gini/I

Bickenbach, F. and Bode, E. (2008). Disproportionality measures of concentration, specialization and localization. *International Regional Science Review*, 31:359–388.

Briant, A., Combes, P., and Lafourcade, M. (2010). Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? *Journal of Urban Economics*, 67(3):287–302.

Clark, S., Hemming, R., and Ulph, D. (1981). On indices for the measurement of poverty. *The Economic Journal*, 91(362):515–526.

Dawkins, C. (2004). Measuring the spatial pattern of residential segregation. *Urban Studies*, 41(4):833.

Dawkins, C. (2006). The spatial pattern of black-white segregation in us metropolitan areas: an exploratory analysis. *Urban Studies*, 43(11):1943.

Dawkins, C. (2007). Space and the measurement of income segregation. *Journal of Regional Science*, 47(2):255–272.

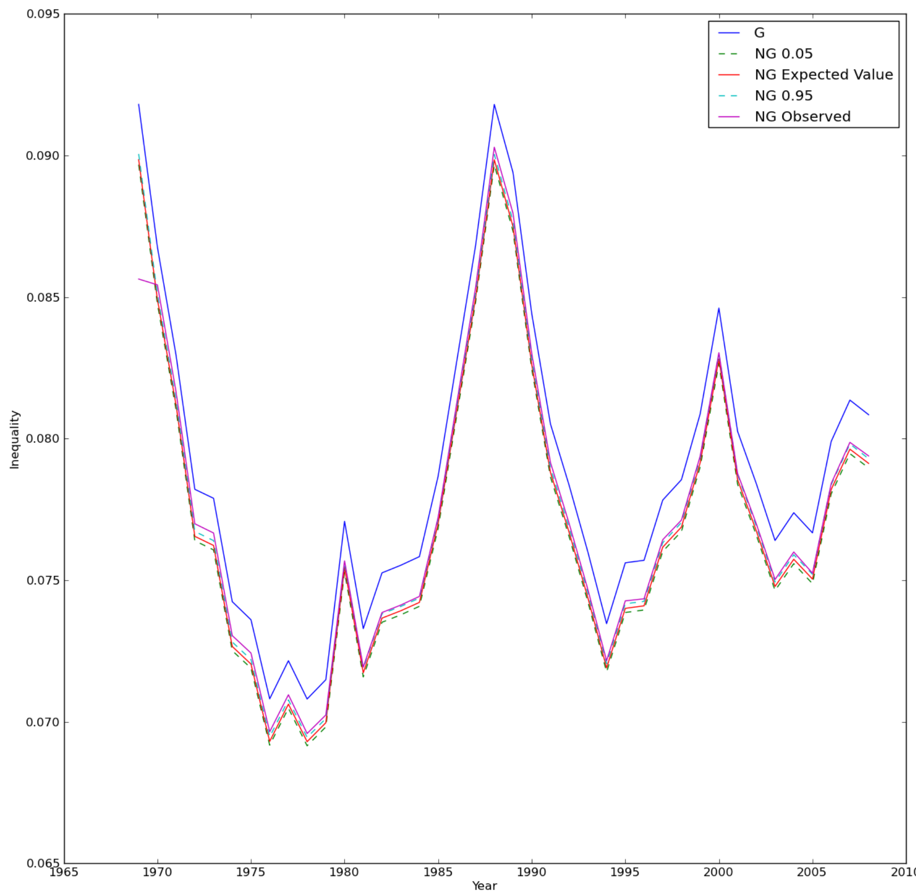

Figure 4: Comparisons of measures of spatial autocorrelation I and NG p-values 1969-2009

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

Friedman, M. and Friedman, R. (2002). *Capitalism and freedom*. University of Chicago press.

Grannis, R. (2002). Discussion: segregation indices and their functional inputs. *Sociological Methodology*, 32(1):69–84.

Grannis, R. (2005). T-communities: Pedestrian street networks and residential segregation in chicago, los angeles, and new york. *City & Community*, 4(3):295–321.

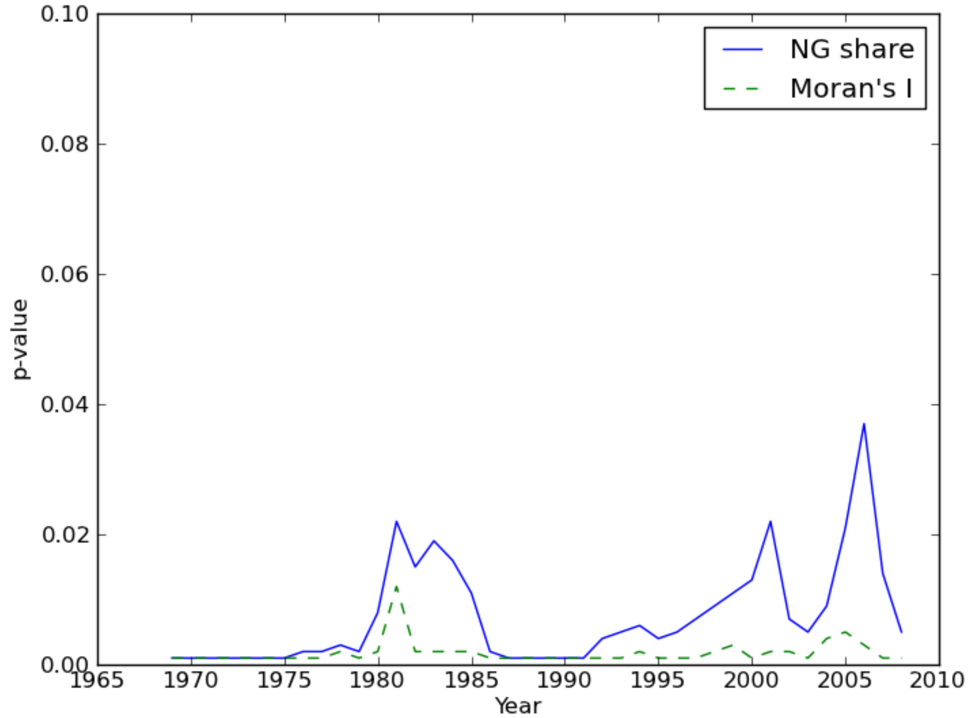

Figure 5: Nonneighbor Inequality Shares

James, D. R. and Tauber, K. (1985). Measures of segregation. *Sociological Methodology*, 15:1–32.

Kain, J. F. (1968). Housing segregation, negro employment, and metropolitan decentralization. *The Quarterly Journal of Economics*, 82(2):175–197.

Kanbur, R. (2006). The policy significance of inequality decompositions. *Journal of Economic Inequality*, 4(3):367–374.

Krugman, P. (1991). *Geography and trade*. The MIT Press.

Ladd, H. (1994). Spatially targeted economic development strategies: Do they work? *Cityscape*, 1(1):193–218.

Levine, D. (1999). Gilding the ghetto. *Across the Board*, 36(9):10.

Massey, D. and Denton, N. (1988a). Dimensions of residential segregation, the. *Social Forces*, 67:281.

Massey, D. and Denton, N. (1988b). Suburbanization and segregation in US metropolitan areas. *American Journal of Sociology*, pages 592–626.

Massey, D., White, M., and Phua, V. (1996). The dimensions of segregation revisited. *Sociological Methods & Research*, 25(2):172.

Morgan, B. (1983a). An alternate approach to the development of a distance-based measure of racial segregation. *American Journal of Sociology*, pages 1237–1249.

Morgan, B. (1983b). A distance-decay based interaction index to measure residential segregation. *Area*, pages 211–217.

Morrill, R. (1991). On the measure of geographical segregation. In *Geography research forum*, volume 11, pages 25–36.

Quigley, J. (1994). New directions for urban policy. *Housing Policy Debate*, 5(1):97–106.

Reardon, S. F. and Firebaugh, G. (2002). Measures of multigroup segregation. In Stolzenberg, R. M., editor, *Sociological Methodology*, pages 33–67. Blackwell Publishing, Boston.

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

Rey, S. J. (2004). Spatial analysis of regional income inequality. In Goodchild, M. and Janelle, D., editors, *Spatially Integrated Social Science: Examples in Best Practice*, pages 280–299. Oxford University Press, Oxford.

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

Rey, S. J. and Janikas, M. V. (2006). STARS: Space-Time Analysis of Regional Systems. *Geographical Analysis*, 38(1):67–86.

Schelling, T. (1969). Models of segregation. *The American Economic Review*, 59(2):488–493.

Shorrocks, A. and Wan, G. (2005). Spatial decomposition of inequality. *Journal of Economic Geography*, 5(1):59.

Silber, J. (1989). Factor components, population subgroups and the computation of the Gini index of inequality. *Review of Economics and Statistics*, 71:107–115.

Sokal, R. R., Oden, N. L., Thomson, B. A., and Kim, J. (1993). Testing for regional differences in means: distinguishing inherent from spurious spatial autocorrelation by restricted randomization. *Geographical Analysis*, 25:199–210.

White, M. (1983). The measurement of spatial segregation. *American Journal of Sociology*, 88:1008–18.

Wilson, W. (1987). *The Truly Disadvantaged: The Inner City, The Underclass and Urban Policy*. University of Chicago Press.

Wong, D. (2003a). Spatial decomposition of segregation indices: a framework toward measuring segregation at multiple levels. *Geographical Analysis*, 35(3):179–194.

Wong, D. S. (2003b). Implementing measures of spatial segregation in GIS. *Computers, Environment and Urban Systems*, 27:53–70.

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

n	ρ	$\sigma^2 = 1$	$\sigma^2 = 4$	$\sigma^2 = 9$	$\sigma^2 = 16$
25	0.0	-0.002	-0.013	0.006	0.010
	0.1	-0.004	-0.003	0.008	0.011
	0.2	0.021	0.018	0.014	0.023
	0.3	0.029	0.022	**0.059**	0.035
	0.4	0.041	**0.051**	0.036	0.038
	0.5	**0.043**	**0.067**	**0.061**	**0.057**
	0.6	**0.049**	**0.064**	**0.067**	**0.054**
	0.7	**0.034**	**0.035**	**0.048**	**0.047**
	0.8	**0.034**	**0.036**	0.020	**0.026**
	0.9	0.008	0.010	0.004	0.010
49	0.0	-0.006	-0.003	0.007	0.000
	0.1	0.028	0.003	0.011	0.005
	0.2	0.010	0.017	0.032	0.032
	0.3	**0.060**	0.035	**0.051**	**0.064**
	0.4	**0.061**	**0.066**	**0.077**	**0.064**
	0.5	**0.048**	**0.043**	**0.050**	**0.035**
	0.6	**0.027**	**0.029**	**0.028**	0.011
	0.7	0.010	0.008	0.008	0.009
	0.8	0.003	0.004	-0.001	0.003
	0.9	0.001	0.001	0.000	0.002
81	0.0	0.002	0.001	0.003	-0.010
	0.1	0.000	0.021	-0.001	0.022
	0.2	**0.047**	**0.066**	**0.051**	0.035
	0.3	**0.049**	0.039	**0.059**	**0.057**
	0.4	**0.083**	**0.061**	**0.054**	**0.044**
	0.5	**0.022**	**0.027**	**0.033**	0.020
	0.6	0.005	0.006	0.008	0.006
	0.7	0.000	0.000	0.001	0.001
	0.8	0.000	0.000	0.000	0.000
	0.9	0.000	0.000	0.000	0.000

Table 1: Difference in rejection frequencies: Moran's I - Spatial Gini. Bold indicates significant difference at $\alpha = 0.05$.