Identifying Spatial Commonalities: An Empirical Application to Inter and Intra-Country Data

WIDER Project Meeting "Spatial Inequality in Development"
May 29, 2003, Helsinki

Bettina Aten and Alan Heston[1]

Introduction

The persistence of spatial income inequalities existing within and between countries poses questions for any simple neo-classical model where factor movements tend to equalize returns, productivity and incomes. An economics major will have learned many, probably too many, explanations of why such convergence does not mirror the real world. Building on location theory, trade and industrial organization, Krugman provided a formulation that is less common in the development literature. This New Economic Geography is based upon economies of scale and transactions costs that produce geographic concentrations of specialization.

Still another strand of research has sought to explain growth and convergence going beyond the framework of Barro and Sala-i-Martin. Fingleton (1999) pursued the application of Markov processes that Quah used in explaining both personal and spatial income distributions over time. Fingleton looked at the regions of the European Union (EU) between 1975 and 1995, finding little or no support for convergence over these periods. Markov chains are often eschewed because they lack a rationale in economic theory. However, if the Markov process approximates reality as well as alternatives, and at the same time lends itself to a plausible interpretation, then Fingleton would argue it has claim to our attention.

Lately a number of trade, development and macro-economists have been exchanging salvos related to persistent income disparities across space. Is it geography and climate; (Sachs, 2001), colonial institutions (Acemoglu, Johnson and Robinson, 2001), social capital (Jones and Hall, 1999), or trade (Dollar and Kraay, 2002)? Rodrik, Subramanian and Trebbi (2002) have recently argued that institutions "trump" both trade and geography as the explanation for income differences. Shleifer and colleagues have termed this line of research on institutions the New Comparative Economics, which is often hard to distinguish from the New Institutional Economics of North, Olson and Williamson.

In more recent exchanges Easterly and Levine (2002) argued that geography has an effect, though primarily through institutions, while Sachs (2003) offers in his defense evidence that geography has a direct effect on economic performance, after allowing for the role of institutions. The instruments chosen in these studies to

[1] Aten is on leave from Bridgewater State College, MA, at the Bureau of Economic Analysis, Washington D.C., and Heston is at the University of Pennsylvania.

represent institutions, policies and geography are often quite thoughtful and ingenious, but it would hardly be surprising if subsequent studies using alternative instruments turn up different results. One is tempted to regard these controversies as analogous to rivalries among competing sects seeking members and funding; hopefully an ecumenical movement will produce some synthesis of these streams of thought.

This paper picks up on one aspect of measuring inequalities of income, namely on how to identify the persistence of income inequality over space. Are there regions that appear poor relative to their surrounding areas regardless of what scale we measure these areas? For example, is there a cluster of poverty within a 200 mile radius? Similarly, are there areas of affluence that spillover into neighboring areas but not beyond them? A part of the inquiry builds on what Krugman (1991) has termed the core-periphery problem, namely that concentrations of economic specialization have relatively little effect in raising surrounding regions to their level of affluence. We use a measure of income that is adjusted by estimates of regional price variations. These price levels are in turn based on a model specification that holds constant a number of factors such as geography, climate and openness, and allows for spatial spillovers between administrative units. Using these adjusted estimates, and comparing them with the original income values, we analyze the scale and sensitivity of the income distribution to changes in the spatial structure.

The research approach is discussed in Part I, the application in Part II, and an analysis of the results in Part III. The analytic techniques used in this paper represent some of the newer applications of spatial statistics, particularly the use of local Moran measures that Anselin has developed. The innovation in our approach is in examining the robustness of results from using alternative measures of distance and closeness.

Part I Research Strategy

In Aten and Heston (2003) a data set was developed that was built up from two components. For 1996 a set of national currency estimates of regional product or income were assembled for 740 sub-national units of 36 countries, and GDP for an additional 131 countries with no sub-national breakdown, a total of 871 observations. In addition there were real product and purchasing power parity (PPP) estimates for all 168 countries.[2] One part of that paper developed relationships between the purchasing power parities of the 168 countries and variables such as geographical location, climate and openness to trade. It used a model with an autocorrelated error structure that permitted estimates of PPPs for the 740 sub-national units. When national currency estimates are converted to international dollars at national PPPs they are termed *nominal regional incomes*. When the estimated regional PPPs are used for the national currency conversions, the regional estimates are referred to as *real incomes*. These estimates permit an examination of both real and nominal income differences for all 871 observations.

The exploratory approach that we use to identify pockets of poverty or of affluence is a variation of the simple cross-product statistic (Upton and Fingleton,

[2] The estimates of real incomes and PPPs for the sample countries are available at http://pwt.econ.upenn.edu.

1985), also referred to as QAP (quadratic analysis paradigm) or gamma index (Anselin, 1995) given by:

$$r = \sum_i \sum_j W_{ij} Y_{ij}$$

Where W_{ij} is a measure of the spatial proximity of locations i and j, and Y_{ij} is a measure of proximity in some other dimension: in this paper, nominal and real incomes. A special case of this approach, where the 'distance' between incomes Y is measured as deviations from the mean, and the index is divided by the sample variance, is called Moran's *I*, and has been frequently used to test for spatial autocorrelation. That is,

$$I = \frac{n \sum_i \sum_j W_{ij} Y_{ij}}{S_0 \sum_i Y_i^2}$$

$$\text{where } Y_{ij} = (y_i - \bar{y})(y_j - \bar{y})$$

$$Y_i^2 = (y_i - \bar{y})^2$$

$$\text{and } S_0 = \sum_i \sum_j W_{ij}$$

Note that for a row-standardized spatial weights matrix, $S_0 = n$, and the expression simplifies to:

$$I = \frac{\sum_i \sum_j W_{ij} Y_{ij}}{\sum_i Y_i^2}$$

When there is no autocorrelation present the expectation of *I* is –1/(n-1), and when there is maximum positive autocorrelation ***I*** will approach 1 (Upton and Fingleton, 1985, Anselin 1995). Moran's *I* indicates the degree of linear association between incomes and the weighted average of the neighboring values. Inferences about the significance of this association are obtained using various assumptions about the data such as asymptotic normality, equal likelihood and an empirical distribution obtained from permutations[3]

The choice of the W matrix will obviously affect the value of the cross-product statistic, and we examine the sensitivity of Moran's *I* to the definition of spatial proximity using a number of W matrices. They range from simple distance measures (inverse distance and inverse distance squared), to contiguity measures and nearest-neighbor measures. One could also use *aspatial* proximity measures, for example, trade flows or a migration matrix, but such data are not available at levels below the national level for most countries.

As a first step we focus on the spatial aspects of the distribution of incomes, particularly on the changes in the statistic as we move between spatial scales. We then search for pockets of instability in the global pattern by looking at outliers at each

[3] Moran's I and the various tests of significance, as well as all computations of the W matrices and corresponding indeces were carried out using SpaceStat 1.90© 1995, Luc Anselin.

spatial scale. To do so we decompose the global Moran's *I* into the contributions of individual observations. The local cross-product statistic is given by:

$$r_i = \sum_j W_{ij} Y_j$$

While the local Moran is equal to:

$$I_i = \frac{Y_i \sum_j W_{ij} Y_j}{\sum_i Y_i^2}$$

$$\text{where, } Y_i = (y_i - \bar{y})$$

The mean of the local Morans is equal to the global Moran (Anselin 1995)[4]. If the global measure is positive, as is the case for most of the *W* matrices, a negative local Moran will indicate local instability, such as a low-income area surrounded by high incomes.

A second approach to looking at the local variation is to examine the outliers by means of a bivariate plot of *Wy* against *y*, called the Moran scatterplot. The slope of this regression line is equal to Moran's *I* (Anselin, 1980) and persistent outliers are defined as observations that are far from the regression line for all the W matrices, where far is defined as two standard deviations from the mean or more than1.5 times the interquartile range of the distribution. The scatterplot also enables us to distinguish between two types of spatial association captured by Moran's *I*, namely whether there is positive spatial association: large values of y_i surrounded by large Wy_i or small values of y_i surrounded by small Wy_i or negative spatial association: large values of y_i surrounded by small Wy_i, or small values of y_i surrounded by large Wy_i. The latter are of particular interest as they represent areas of dissimilarity – pockets of poverty surrounded by affluence or vice versa.

One note of caution in interpreting the results should be mentioned. The estimates that we made of sub-national PPPs themselves involved using a measure of spatial interaction in the assumption about the error structure of the model[5]. Therefore, there is an element of tail chasing in this particular application of the real income data set for the regions. This is the reason that both nominal and real incomes have been presented throughout.

---

[4] See Anselin (1995) for a discussion of the moments and distribution of the local Moran under the null hypothesis.

[5] See Fingleton (1999) for a discussion of this model as an application to regional income convergence in the Eu. The autocorrelated errors model specification in matrix terms is:

$Y = Xb + e, \quad e = rWe + m, \text{ and } m \approx N(0, s^2 I)$

Part II: Application

A summary of the nominal and real income data (in international dollars[6]) is given in Table 1. It shows the mean, standard deviation, median and range of the distribution for the 871 country and regional units. Additionally, the distribution for incomes converted at exchange rates to the U.S. dollar are shown. The means of the nominal and real measures differ slightly because of differences in weights when regional incomes are averaged. The mean income at exchange rates is also close to the real and nominal incomes, which reflects the fact that for the world of 1996 the distribution of the ratio of the PPPs to the exchange rates to the US dollar across all regions are close to unity.

Table 1: Summary Characteristics of Income Data, 1996

Incomes (I $)	Mean	Std. Deviation	Median	Range
Nominal	11,468	9,029	7,652	51,567
Real	11,422	8,922	7,555	46,802
Exchange Rates (US$)	*11,536*	*12,501*	*4,276*	*71,346*

Source: Aten and Heston (2003)

However, the distribution of incomes does vary substantially between the three measures. The real income distribution has the smallest range, as might be expected, although this is not true within all countries (see Aten and Heston, 2003,for a more detailed discussion). The five lowest income areas in nominal and real terms are Akure, Abeokuta and Ibadan in Nigeria, Zaire, and Tanzania, although their respective positions change (Ibadan and Tanzania are 4th and 5th lowest in real terms but 5th and 4th lowest in nominal terms). Similarly, the four highest income areas are the same (Trenton, NJ, Hartford, CT, and Washington, DC, in the United States, and Oslo, Norway which is highest in both nominal and real terms), but the fifth highest is Hamburg, Germany in nominal terms and Luxembourg in real terms. At exchange rates, the lowest income areas are Zaire, Ethiopia, Burundi, Tajikistan and Mozambique (ranging from US$ 74 to 175) and the highest are Tokyo, Japan at US$ 53,000 and Oslo, Norway at US$ 71,420. Because the spatial distribution of incomes is misleadingly large at exchange rates, they will not be further discussed.

Table 2 shows the weights matrices (the $W_{ij}s$) created from these observations. They can be divided into three groups. The first group contains two common distance measures used in the literature, inverse of distance and inverse of distance squared. In all of the applications, capital cities are taken as the central point for countries and for the regions, which of necessity are all administrative units. The inverse (and inverse squared) of the great circle distance between each pair of observations becomes the entry in the *W* matrix, and the row values are standardized to sum to one. This eliminates scale effects within regions, so that physically larger countries with fewer regions do not receive disproportionately less weight (the distances between regions would be greater and hence their inverse would be less) than smaller countries. The

[6] An International dollar (I$) is the PPP converted national currency relative to the United States dollar.

scale effects across regions are captured more precisely in the next group of matrices, the contiguity and nearest neighbor matrices.

The second group includes contiguity measures consisting of a set of 9 matrices that indicate whether regions are within a critical distance threshold of one another, with the distances ranging from a circle of radius 100 to 5000 miles. For each region (row entry in the *W* matrix), the regions (column entries) within, say 100 miles, are assigned a value of one, and regions outside that radius are assigned a zero value. Here, unlike the inverse distance matrices, there will be entire rows with zero entries, as the administrative center of some regions will have no observations within the critical distance threshold of 100 miles, for example, Hawaii, or the state of Amazonas in Brazil. The thresholds allow us to see the distances where there are changes in the overall pattern of spatial autocorrelation.

Finally, the third group of matrices is a set of 15 nearest neighbor matrices, from first-order nearest neighbors to k=15 nearest neighbors. Entries in the *W* matrix again consist of zero and ones, where the ones correspond to regions that are k-order neighbors. Unlike the distance and contiguity measures, the nearest neighbor entries are not necessarily symmetric – region A may be closest to B, but B is closer to C than to A. The difference between nearest neighbor matrices relative to the contiguity measures is that they are independent of scale – two large regions that are neighbors but are separated by a vast distance – such as some provinces in China, will have the same weight relative to one another as two neighboring prefectures in Japan.

Table 2 Summary Characteristics of Weight Matrices

W matrix 871 x 871	**Number of zero rows**	**Average number of links per row**	**Most Connected**	**Least Connected**
Distance				
Inverse	0	870	All	None
Inverse Squared	0	870	All	None
Contiguity				
100 miles	302	3	Hasselt, BEL (20 links)	133 obs (1 link)
200 miles	138	9	Arnsberg, GER (49 links)	102 obs (1 link)
300 miles	58	18	Amiens, FRA; Darmstadt, Freiburg, Mainz, Wiesbaden, Koblenz: GER (71 links)	57 obs (1 link)
400 miles	31	26	Hasselt, BEL (106 links)	28 obs (1 link)
500 miles	18	37	Munster, GER (131 links)	11 obs (1 link)
1000 miles	4	93	Slovakia; Graz, Eisenstandt: AUT (255 links)	Mauritius; Papua New Guinea; Seychelles (1 link)
2000 miles	1	192	Moldova (381 links)	Fiji (1 link)
3000 miles	0	276	Sohag, Al-Minya, Suez: EGY (455 links)	New Zealand (3 links)
5000 miles	0	460	Oman (654 links)	New Zealand (35 links)
Nearest				

W matrix 871 x 871	Number of zero rows	Average number of links per row	Most Connected	Least Connected
Neighbor				
K=1	0	1	All	None
K=2	0	2	All	None
K=3	0	3	All	None
K=4	0	4	All	None
K=5	0	5	All	None
K=6	0	6	All	None
K=7	0	7	All	None
K=8	0	8	All	None
K=9	0	9	All	None
K=10	0	10	All	None
K=11	0	11	All	None
K=12	0	12	All	None
K=13	0	13	All	None
K=14	0	14	All	None
K=15	0	15	All	None

Of these 26 matrices, a number are unlikely to capture meaningful spatial variations, for example the distance band of 5000 miles, or the k=15 nearest neighbor. However, by including a range of measures, we gain some insights into the relative importance of alternative assumptions, as described in Part III.

Part III Results

The value of Moran's I for each weights matrix is provided for nominal incomes in column (2) and real incomes in column (3) of Table 3. It was mentioned that there is some possible tail-chasing involved in the real measure and this may account for the fact that column (3) values are usually higher than for nominal incomes. However, the patterns displayed by Moran's I statistic are essentially the same for nominal and real income measure, so it makes little difference which is used.

Table 3 : Moran's I for Nominal and Real Incomes

W matrix	Moran's I Nominal Income*	Moran's I Real Income*
Distance		
Inverse	0.35	0.36
Inverse Squared	0.73	0.74
Contiguity		
100 miles	0.87	0.89
200 miles	0.86	0.88
300 miles	0.87	0.89
400 miles	0.85	0.87

W matrix	Moran's I Nominal Income*	Moran's I Real Income*
500 miles	0.84	0.85
1000 miles	0.73	0.74
2000 miles	0.44	0.45
3000 miles	0.23	0.23
5000 miles	0.10	0.11
Nearest Neighbor		
K=1	0.90	0.92
K=2	0.89	0.91
K=3	0.88	0.90
K=4	0.88	0.90
K=5	0.88	0.90
K=6	0.88	0.90
K=7	0.87	0.89
K=8	0.87	0.89
K=9	0.87	0.88
K=10	0.86	0.88
K=11	0.86	0.88
K=12	0.86	0.88
K=13	0.86	0.88
K=14	0.85	0.87
K=15	0.85	0.87

*All values are significant at the $p<0.001$ level (under the normal and randomization assumptions)

Figures 1 and 2 show the distribution of the I-statistics for the nominal and real incomes, separated into the distance and the nearest-neighbor matrices to show the more detailed variations. Note the change in the scale of the vertical axis between Figure 1 and 2 as the nearest-neighbor statistics are generally higher and have a smaller range than the distance-based measures. Some conclusions from examining the different measures are:

(a) The inverse distance measure is much less sensitive to picking up spatial autocorrelation than inverse distance squared. Since these are the only continuous variables used in the weight matrices, this may be of interest for other types of analysis. Fingleton (1999) used a mixed approach for the European Union, with inverse distance squared weights up to a distance band of approximately 900 miles, and zero weights beyond that distance. In this sample, the spatial autocorrelation for the inverse distance squared and the 1000 mile threshold are also about equal: 0.73 for nominal incomes and 0.74 for real incomes.

(b) There is a very consistent pattern with respect to the 9 distance groupings as illustrated in Figure 1 which plots the first 9 values in columns (2) and (3). For distance bands from 100 to 500 miles, there is little difference, but a consistent fall in value begins at 300 miles; and beyond 500 miles the distance bands display less and less spatial autocorrelation.

(c) For the number of nearest neighbors, there is a consistent but very gradual fall off in the value of Moran's I, as the number of neighbors is increased. Unlike the distance bands, there is a flattening of the statistic around k=5. This is illustrated in Figure 2.

(d) There is very little difference in the nominal and real measures (note that the scale in Figure 2 is over a smaller range, so that differences appear greater than in Figure 1, but the absolute difference in magnitudes is similar. See Table 3)

Figure 1. Moran's I for Inverse and Contiguity Distance Matrices

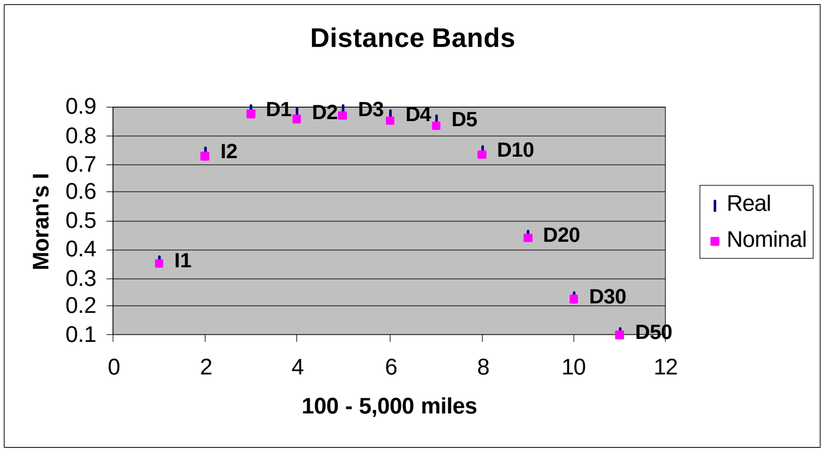


Figure 2. Moran's I for Nearest Neighbor Matrices

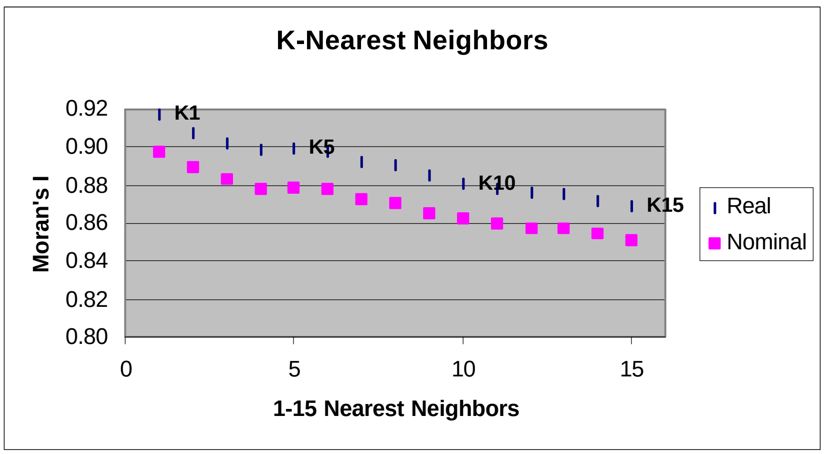


Table 4 is a summary of the local Moran's I distribution for selected weight matrices. The mean is equal to the global Moran, while the standard deviation and the median provide some information on the shape of the distribution. The ten most extreme negative values are shown, consisting of high income areas surrounded by low- income areas, or low-income areas surrounded by high incomes (underlined and italicized). These negative values represent areas of *dissimilarity* of incomes, or inequalities, and the distinction between pockets of poverty versus pockets of affluence will be discussed below.

Table 4. Local Moran's I-statistics – Areas of Inequalities

W	Nearest Neighbor K=1	Nearest Neighbor K=5	Contiguity Distance = 300 miles	Contiguity Distance = 500 miles
Mean	0.92	0.90	0.83	0.84
Standard Deviation	0.98	0.95	0.91	0.84
Median	0.73	0.72	0.65	0.67
Most extreme negative values:				
1	*Guangzhou ,CHN*	Hong Kong HKG	Singapore SGP	Singapore SGP
2	*Algeria DZA*	Australia AUS	Cyprus CYP	Hong Kong HKG
3	*Estonia EST*	Singapore SGP	Hong Kong HKG	Cyprus CYP
4	Helsinki, FIN	*Algeria DZA*	*Croatia HRV*	Israel ISR
5	*Johor Baharu, MYS*	Israel ISR	Israel ISR	*Croatia HRV*
6	Singapore SGP	Macao MAC	Macao MAC	Macao MAC
7	Bahamas BHS	*Croatia HRC*	*Algeria DZA*	*Estonia EST*
8	*Fuzhou, CHN*	*Fiji FJI*	*Latvia LVA*	*Latvia LVA*
9	Taiwan TWN	*Estonia EST*	*Estonia EST*	*Algeria DZA*
10	Israel ISR	*Latvia LVA*	*Guangzhou,CHN*	*Tunisia TUN*

For the k=1 matrix (first column in Table 4), the most extreme *negative* values are Guangzhou and Fuzhou in China, Algeria, Estonia, Helsinki in Finland and Johar Baharu in Malaysia, Singapore, the Bahamas, Taiwan and Israel. They are areas of dissimilarity, with their own incomes differing significantly from the values of their nearest-neighbors, such as Estonia and Helsinki, Johar Baharu and Singapore. Because the nearest neighbor relationship is not always symmetric, not all values in the list are paired.

A notable feature of Table 4 (and for all the weight matrices in general) is that most areas are associated with country borders[7]. This finding reinforces the literature that stresses the relative importance of inter-country differences versus intra-regional or national income distributions in determining the world distribution of personal income. Of the 40 cases in Table 4, 19 represent pairings with transition economies, and Israel appears under all four weights matrices, indicating the major role of geo-politics.

[7] In 1996 Hong Kong was still under lease to the U.K. and its administration and currency remain quite distinctive from the remainder of China.

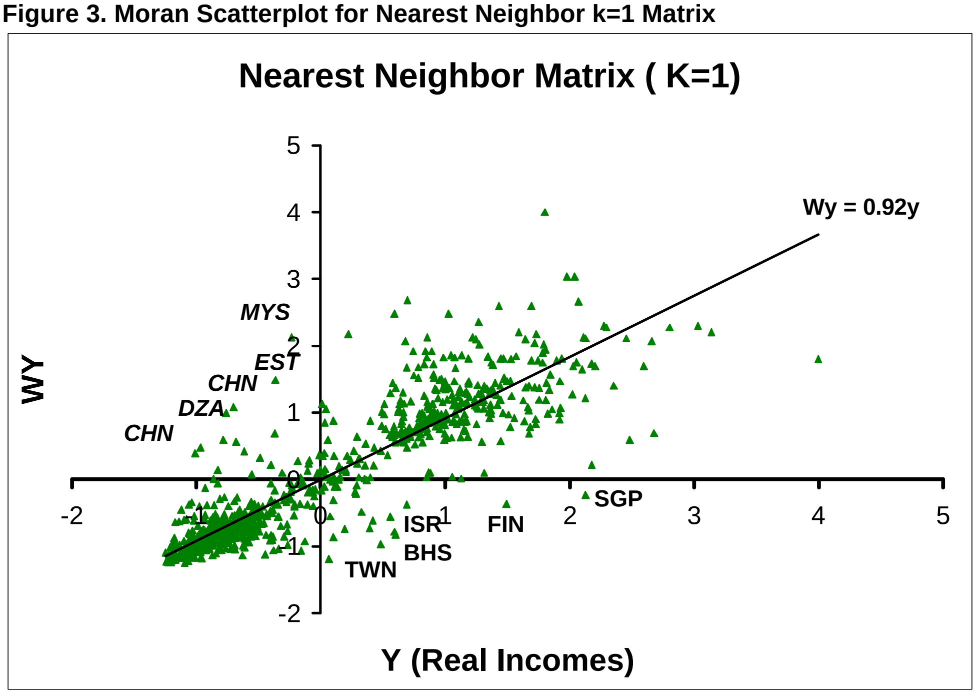
Table 5 illustrates the extreme income difference between country borders for the k=1 nearest neighbor matrix only.

Table 5. Most Extreme Pockets of Inequality relative to K=1 neighbors

	Area LOW	Real Income (Y)		Area HIGH	Real Income (Y)
1	***Johor Baharu (MYS)***	***9,312***	-	**Singapore (SGP)**	**30,403**
2	***Estonia (EST)***	***8,187***	-	**Helsinki (FIN)**	**24,729**
3	***Guangzhou (CHN)***	***5,177***	-	Hong Kong	26,274
4	***Algeria (DZA)***	***4,627***	-	Palma De Mallorca (ESP)	20,252
5	Haiti	2,033	-	**Bahamas (BHS)**	**16,751**
6	Jordan	4,428	-	**Israel (ISR)**	**16,708**
7	***Fuzhou (CHN)***	***2,783***	-	**Taiwan (TWN)**	**15,703**

We highlight the extreme negative values of Moran's I, but the statistic does not distinguish between pockets of affluence versus pockets of poverty. The distinction only becomes visible when we decompose the statistic into its numerator and denominator or depict it using a scatterplot of the *Wys* and *Ys*. For example, Helsinki has a positive (high income) *Y* and a *Wy* that is negative (its neighbors are low-income), while for Estonia, the reverse is true, with a *Y* that is negative and a *WY* that is positive. The bivariate scatterplots for real incomes and two distance matrices (k=1 and 300 miles) are shown in Figure 3 and 4.

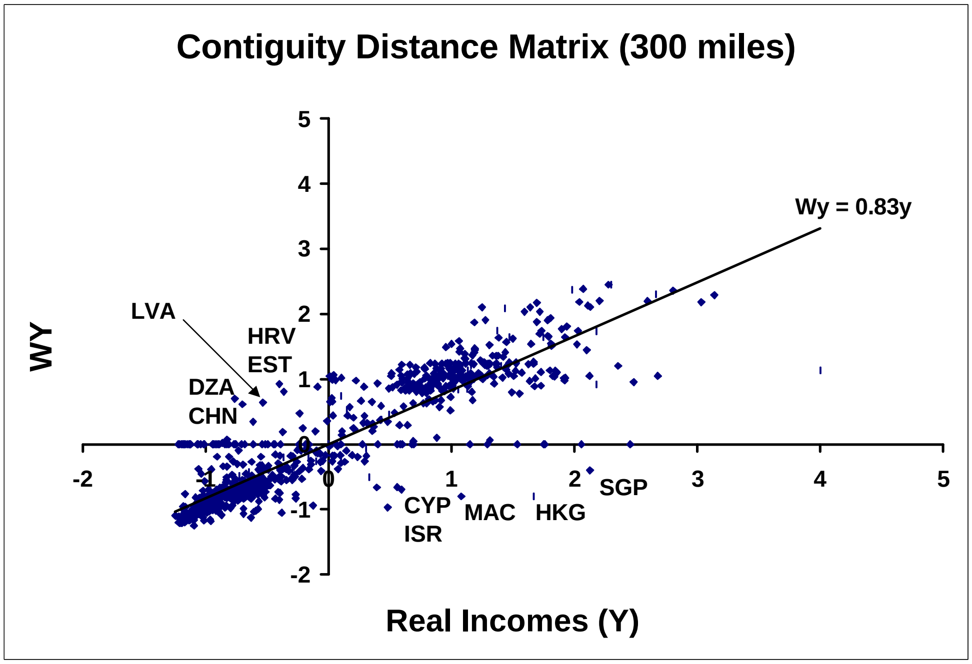
Figure 3. Moran Scatterplot for Nearest Neighbor k=1 Matrix



Areas in the upper left hand quadrant are low-income areas with high-income neighbors (negative *Y*, positive *WY*), and areas in the lower right-hand quadrant are high-income areas (positive *Y*) with low-income neighbors (negative *WY*). The ten largest negative Moran values are labeled with their country codes (listed in Table 4). They are the observations farthest from the origin in their respective quadrants. Note that because Moran's I is positive and equal to 0.92, the majority of observations are areas of similar incomes: high-income clusters in the upper right quadrant and low-income clusters in the lower left quadrant.

When we use the 300 miles radius instead of the first nearest neighbor matrix, some of the relationships change, as evidenced in Figure 4. Croatia and Latvia are now included as pockets of poverty and Macao, Hong Kong, and Cyprus are pockets of affluence. The spatial autocorrelation is lower (0.83 compared with 0.92 for k=1), and the standard deviation of the local Moran's is also smaller. One reason for this is that many observations (58 total) have zero weights and are on the horizontal axis, with closest neighbors outside the 300-mile distance band.

Figure 4. Moran Scatterplot for Contiguity Distance Matrix (300 miles)



In Table 6 there is more evidence of regional, rather than inter-country inequalities for the 300-mile radius. For example, in Spain (Sevilla and Merida relative to Ceuta and Toledo), Portugal, with Coimbra and Evora relative to Lisbon, and Ch'uncho'on relative to Seoul in the Republic of Korea are all in the upper left quadrant. Guanzhou in China is present as an extreme value in both the first order nearest neighbor and the 300-mile distance measures (and in many of the other matrix measures as well) as its income level contrasts strikingly with that of Hong Kong. Johor

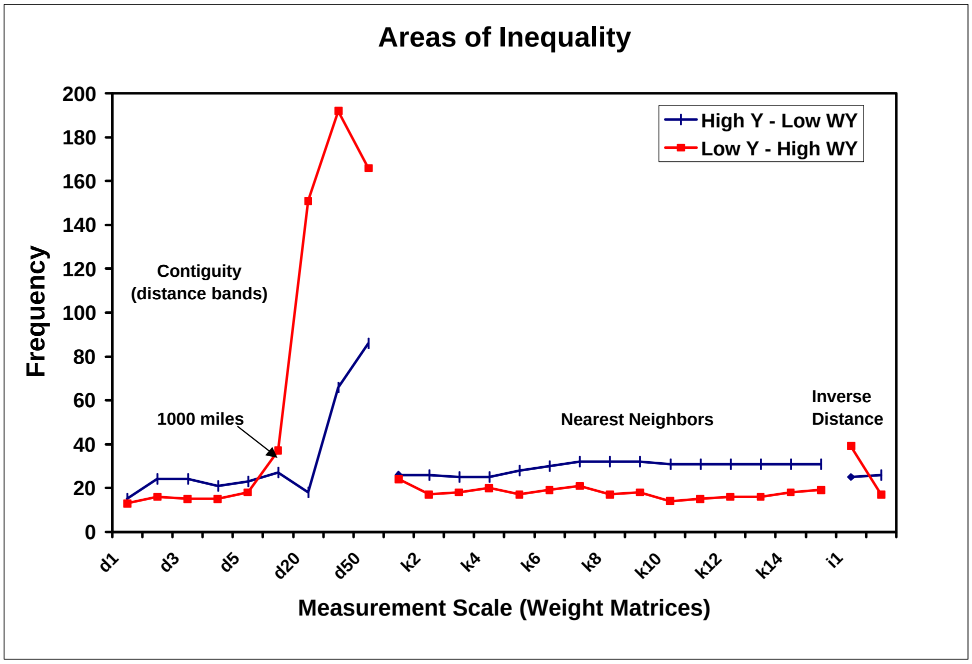
Baharu in Malaysia disappears because other, poorer, regions in Malaysia are included within the 300-mile radius, whereas in the nearest-neighbor measure only high-income Singapore is included. However, Singapore continues to have low-income neighbors within 300 miles, as evidenced by its extremely negative local Moran value (Table 4), and its position in the lower right-hand quadrant of the Moran scatterplot of Figures 3 and 4.

Table 6 Pockets of Poverty and Affluence relative to d=300 miles

	Area of POVERTY	**Real Income (Y)**	**Main Areas of AFFLUENCE within 300 miles**
1	Albania (ALB)	3,733	Naples (ITA), Athens (GRC)
2	Morocco (MAR)	4,041	Sevilla (ESP), Malaga (ESP)
3	Algeria (DZA)	4,627	Palma de Mallorca (ESP), Barcelona (ESP)
4	Guangzhou (CHN)	5,177	Hong Kong (HKG), Macao (MAC)
5	Latvia (LVA)	6,647	Helsinki (FIN), Stockholm (SWE)
6	Croatia (HRV)	7,857	Vienna (AUT), Trieste, Milano, Venezia (ITA)
7	Evora (PRT)	8,108	Lisboa (PRT), Madrid (ESP)
8	Estonia (EST)	8,187	Helsinki (FIN), Stockholm (SWE)
9	Merida (ESP)	9,544	Lisboa (PRT), Madrid (ESP)

The final chart shown in Figure 5 is the frequency distribution of the areas of dissimilarity (negative Moran's I) across all the weight matrices. It is simply a count of the observations in the upper-left quadrant (in red) and the lower right-hand quadrant (in blue). The large increase in pockets of poverty occurs beyond the 1000-mile distance band, which coincides with a sharp drop in the value of the spatial autocorrelation coefficient (Table 3). One interpretation is that we are simply capturing clusters of high-income areas, such as the smaller areas in Western Europe, and that such large distances are not effective in measuring localized difference effects.

Figure 5 Frequency Polygon of Scale Effect



Conclusion

This paper reports on the use of a particular form of spatial autocorrelation to group countries and regions according to how similar or dissimilar their incomes are relative to surrounding areas. We looked at per capita incomes in 1996 for 871 administrative units using nominal values, obtained by converting the income of countries and regions within countries to international dollars at national PPPs, and real values, where regional estimates of PPPs of 32 larger countries allowed variation in price levels within countries. Either income measure produced similar results.

The analysis included 26 different measures of distance between the administrative center of each region and its neighbors. They ranged from a continuous distance-decay measure (inverse distance and inverse distance-squared) representing the expectation that regions farther away from each other will have less interaction than closer regions, to weighting only regions within a certain radius (from 100 miles to 5000 miles), to regions that are the closest nearest neighbors – from first order to fiftheenth-order neighbors.

The spatial autocorrelation measure that we used is a special form of a simple cross-product statistic, with a weight matrix representing spatial similarity and an income matrix representing deviations from the mean. Both the global form (Moran's I) and the local indicator of spatial association (LISA) format were examined. The results from the global Moran supported the received view that country borders are associated with the larger proportion of income differences across space. In over half the cases examined, the difference between two regions, or between two countries, represented transition versus market economies. However, when local Moran values were examined, the outliers were more often regions within countries.

A surprising result of the analysis is that the usual suspects in regional polarity scenarios were not in evidence. Individual states in Northeast Brazil did not emerge consistently as pockets of poverty, nor states like Bihar and Orissa in India. This is in part due to the existence of clusters of poverty rather than individual areas surrounded by very affluent areas. For example, Recife is one of the largest cities in the Northeast, and it is very poor (I$ 4,291) compared to São Paulo (I$6,674), but Recife's nearest neighbors include João Pessoa with a real income of I$ 3,010 and Maceió at I$3,601. The same is true of Italy's south, as many of the distance measures that incorporate the Milano-Turino complex also include Croatia, Albania, sometimes Greece and parts of North Africa as well. Thus the disparity within countries is often diffused by the dissimilarity in incomes across countries. Is this cause to rewrite the book on regional disparities between countries or does it only indicate a limitation of the methodology that we have employed?

Before reaching a final conclusion, there are variations on our approach that deserve examination. For example, the local Moran analysis does point to a number of regional disparities that have not received as much attention in the literature as others, namely within Spain, Portugal, Turkey, the Ukraine and South Korea. These disparities are more sensitive to small changes in distance, and thus contiguity bands that are less than 100 miles should be employed. These smaller distances and separate regional matrices would capture exclusively within-country disparities, as would a measure reflecting transport costs, especially the border-transition cost. There is still another approach which we have not had time to sufficiently explore, but which we will mention.

In the previous scatterplots of *WY* on *Y*, the slope of the regression line equals the value of the global Moran's I for that weight matrix. Regions of inequality are depicted in the upper-left and lower-right quadrants, corresponding to negative values of the local Moran statistics, and more specifically, poor regions surrounded by rich areas and rich regions surrounded by poor areas, respectively. If instead of the regression line we draw the 45-degree line through the origin, the observations on the line represent cases where the region has exactly the same standardized income as its neighbors (*Y* is equal to *WY*). An observation above the line indicates that the region's neighbors have higher incomes than the region itself, regardless of whether the region is poor or affluent. In other words, we can disregard the vertical axis and focus on the 45-degree line to highlight additional areas of dissimilarity.

In terms of the usual suspects, now the northeastern states of Brazil that include the cities of Teresina, João Pessoa and Maceió do show up but Recife is below the line in all but the very large distance measures that would take in Brasilia and São Paulo . Similarly, two of the southern states, Florianópolis and Pôrto Alegre that are often thought of as well-off actually show up above the line because they are close to Buenos Aires and Uruguay, as well as to Curitiba, one of the highest-income areas in Brazil. In India, Rajastan, Bihar, Orissa and Uttar Pradesh also show up above the line. Preliminary analysis of China, Italy and some other countries suggest the value of employing confidence intervals in further pursuing this approach.

The use of a simple cross-product statistic at two levels allowed us to make inferences about spatial inequalities of income across a disparate group of observations. It is also suggestive of a couple of possible areas of research. First, the sensitivity of the results to differences in how one defines 'space' may help determine the choice of global weight matrices in exploring autocorrelated structures in regional and international comparisons. Secondly, the persistence of relative poverty or of affluence between regions in different countries, regardless of the scale or measure of proximity that was employed, suggests that a different type of weight matrix may be needed to capture the more subtle inequalities within one country. Examples of such matrices might include ones that more accurately depict transport costs, differences in climate and environmental conditions and geo-political structures.

References

Acemoglu, Daron, Simon Johnson and James A. Robinson (2001) "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91(5), December.

Anselin, Luc (1995), "Local Indicators of Spatial Association – LISA", *Geographical Analysis*, Vol.27:2, April.

Aten, Bettina and Alan Heston (2003) "Regional Output Differences in International Perspective", WIDER-LSE-Cornell Conference, March 2002.

Dollar, David and Art Kraay (2002) "Institutions, Trade and Growth," Carnegie-Rochester Conference Series on Public Policy, forthcoming.

Easterly, William and Ross Levine (2002), "Tropics, Germs, and Crops: How Endowments Influence Economic Development" NBER WP# 9106.

Fingleton, Bernard (1999), "Estimates of Time to Economic Convergence: An Analysis of Regions of the European Union," *International Regional Science Review*, April.

Getis, Art (1991), "Spatial Interaction and Spatial Autocorrelation: A Cross-Product Approach", *Environment and Planning A* 23, 1269-77.

Jones, Charles I. and Robert E. Hall (1999) "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *Quarterly Journal of Economics*, 116(1), 83-116.

Krugman, Paul (1991), *Geography and Trade*, Cambridge, MIT Press.

Rodrik, Dani, Arvind Subramanian and Francesco Trebbi (2002) "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development," NBER WP#. 9305.

Sachs, Jeffrey (2001), "Tropical Underdevelopment," NBER WP# 8119.

Sachs, Jeffrey (2003), "Institutions Don't Rule: Direct Effects of Geography on Per Capita Income," NBER WP# 9490.

Upton, Graham J. and Bernard Fingleton (1985), *Spatial Analysis by Example, Vol. I, Point Pattern and Quantitative Data*, John Wiley and Sons, New York.