CofFEE

Centre of Full Employment and Equity

Developing Spatial Measures of Residential Segregation using Kernel Density Estimation

Martin Watts[1]

December 2007

1.   Introduction

Residential segregation can be defined as 'the extent to which individuals of different groups occupy or experience different social environments' (Reardon and O'Sullivan, 2004: 122-123). In the USA research has demonstrated a powerful role for residential segregation by ethnic group and gender in reinforcing inequality and promoting labour market segregation, which is manifested in income polarisation and differential rates of unemployment (see, for example, the review by Charles, 2003, but see Peach, 2006 for a discussion of the positive effects of residential segregation). Three explanations of persistent residential segregation feature in the US literature, namely objective differences in socioeconomic status, prejudice and housing market discrimination (Charles, 2003: 176).

Thus the impact of residential segregation by race and gender on the spatial distribution of un/employment is an important area of empirical enquiry. However, residential segregation is the outcome of the complex interplay of social and economic processes. Massey and Denton (1988: 309-310) identify five dimensions of segregation. *Evenness* is the degree to which populations are distributed uniformly across space (areal units). *Exposure (isolation)* measures the extent to which the different (same) population groups share common areas. *Clustering* identifies the extent to which members of a minority group are located close to each other. *Concentration* refers to the degree to which a group agglomerates in space. *Centralisation* measures the extent to which a group resides close to the centre of an urban area.

The most frequently used measures of segregation, namely evenness and isolation are aspatial because they fail to take into account the proximity of population groups between, as well as within, areal units which gives rise to the so called checkerboard problem (White, 1983). Also neither spatial nor aspatial indexes typically address the Modifiable Areal Unit Problem (MAUP), even though the former take account of proximity, because the data are typically aggregated into population counts across specific areal units which are sensitive to the chosen boundaries, rather than observations being precisely located in space (Reardon and O'Sullivan, 2004: 123-124). These areas are normally administratively determined and based on high levels of internal homogeneity which distorts measurement (O'Sullivan and Wong, 2007: 148). In addition, the spatial scale over which the segregation process occurs is not known *a priori* (Lee et al, 2006).

With the exception of Healy and Birrell (2003) and a series of articles by Forrest, Johnston and Poulsen, in varying order (2001, 2002a,b, 2004, 2006, 2007), there has been limited research on the measurement and interpretation of ethnic segregation by residence in Australian cities.

This paper has two objectives: i) an assessment of the use of kernel density estimation as a means of spatially smoothing population data prior to the measurement of residential segregation (Martin et al, 2000; O'Sullivan and Wong, 2007); and ii) a brief exploration of the merits of different spatial measures of residential segregation, given the arguments of Reardon and O'Sullivan (2004) and other contemporary literature. These issues will be illustrated by the measurement of residential segregation in the Sydney Commuting Area in 2001, using an index of unevenness and one of exposure with a range of different parameter values for the underlying kernel density estimation to test the sensitivity of the magnitudes.

2.   Spatial smoothing

Reardon and O'Sullivan (2004: 153-154) note that population data are normally defined across discrete areas, rather than by individual location. Population counts at centroids can be replaced by uniform densities for the population groups across each area, but this approach can cause sharp discontinuities at the boundaries, which is likely to distort the computation of segregation measures.

The authors (2004: 128-131) develop a coherent conceptual framework within which to incorporate space into aspatial indexes of evenness and exposure and provide multi-group analogues to the construction of pairwise indexes. They overcome the rigid boundaries of the enumeration units through the calculation of local rather than areal unit based populations to underpin index computation.

 Population groups live in J areas within region R. Typically they are assumed to be located at the corresponding areal centroids, that is at points p,q,r… A non-negative proximity function $\phi(p,q)$ is defined between all pairs of these discrete points where $\phi(p,q) = \phi(q,p)$ and $\phi(p,p) = \phi(q,q)$ for all $p,q \in R$ and a larger value of the function denotes closer proximity. Population measures are redefined within the local environment of each sample point via the proximity functions. These redefined measures can be used in the standard aspatial segregation measures to convert them to spatial measures (Reardon and O'Sullivan, 2004: 136-141). The proximity function could be constructed to reflect specific geographic or other features of the region which enhance or inhibit proximity and hence access.

In their study of the impact of scale on the magnitude of segregation, Lee et al (2006) first calculate population counts based on track densities across a grid of 50m * 50m cells. These counts are then smoothed (Tobler, 1979), prior to the application of the proximity function.

2.1   Kernel density estimation

An alternative form of data smoothing is based on kernel density estimation (KDE) (Martin et al, 2000; O'Sullivan and Wong, 2007). KDE entails the estimation of a probability density function (pdf) from a data sample. Under KDE, a particular functional form is applied across the sample data and summed to yield a single empirical estimate of the underlying pdf, y(x) (O'Sullivan and Wong, 2007: 152). This can be written as

$$y(x) = (1/n)\sum_{i=1}^{n} K((x - X_i)/h)$$
(1)

where X_i (i = 1,2,…,n) denotes the sample data.

Only sample observations which lie within the bandwidth, h, of x are included in the estimated y(x). A wide bandwidth leads to over-smoothing, so that key features of the pdf are hidden, whereas under-smoothing leads to an accentuation of the peaks of the estimated pdf. There a number of common univariate kernel functions including the Gaussian, triangular, triweight, epanechnikov and the quartic. KDE is crucially dependent on parameter values, in partiucular the magnitude of the bandwidth, h.

The extension of a univariate KDE to the spatial (two dimensional) case is straightforward with the kernel function normally assumed to be radially symmetric, so that all sample points within a distance h of x contribute to the estimated y(x). When population counts are located at a discrete number of points, typically centroids, they are counted as separate data points (O'Sullivan and Wong, 2007: 153).

Assuming that the bandwidth is chosen appropriately, so that it crosses the boundaries of the areal units, smoothing takes place which addresses the problems caused by non-interaction across boundaries. While, in principle, the pdf can be computed for any $x \in R$, in practice a finite resolution grid must be imposed on the region which defines the locations at which pdf estimates are made (O'Sullivan and Wong, 2007: 153). These points can be used in index calculations.

Martin et al (2000: 345) adopt a weighting schema which is used for the construction of standard UK surfaces:

$$w_{ij} = [(k^2 - d_{ij}^2)/(k^2 + d_{ij}^2)]^\alpha \qquad d_{ij} < k$$
$$= 0 \qquad otherwise \tag{2}$$

where k is the kernel width and d_{ij} denotes the distance between the sample data point, j and i, a point on the resolution grid. Then the estimated population at point i can be written as:

$$\hat{P}_i = \sum_{j=1}^{N} P_j w_{ij} \tag{3}$$

so that only sample points within distance k of grid point i are included.

The parameter α impacts on the shape of the distance decay function. Martin et al (2000: 345) note that values of α greater than unity generate a more peaked kernel, a value of unity yields an approximately uniform decline from the centroid to the kernel's edge, whereas a parameter value of less than unity results in a flatter kernel with rapid decay close to the edge. O'Sullivan and Wong (2007: 153-154) recommend that the bandwidth lie between 2 and 5 times the grid resolution. There are also other forms of spatial data smoothing, including pycnophylactic smoothing (Tobler, 1979) and dasymetric mapping (Holt et al, 2004).

3. Segregation indexes

Research on the extent of residential segregation can take three broad forms: i) a time series study based on a given region; ii) a cross section study of different regions; and iii) an intra-regional study based on local measures of segregation (Wong, 2002, Brown and Chung, 2006). We do not consider local measures in this paper.

Brown and Chung (2006) claim that the Massey and Denton schema can be simplified by locating concentration at the other end of the continuum from evenness. The shrinking of the dimensions of residential segregation remains an empirical rather than a theoretical question, however. A Concentration index compares the distribution of areas by size with the corresponding distribution of a particular racial group. A measurement of Evenness such as the Dissimilarity Index for blacks and whites could be zero, yet the corresponding concentration index could be very high (or zero) because the populations exhibit significant variation in their spatial density (or exhibit uniform density) across space. Also Brown and Chung (2006: 129) argue that spatial clustering is representative of low exposure to other groups, but Reardon and Firebaugh (2002a: 126) argue that it is clustering and evenness that appear at opposite ends of the continuum. Brown and Chung (2006) provide empirical support for the above claims in the form of high correlations between evenness and concentration, and clustering and exposure, respectively. The index measure of Centrality is increasingly redundant given the emerging polycentric, multimodal and sprawling cities (Brown and Chung, 2006: 126).

Johnston, Poulsen and Forrest (2007) examine the five dimensions of segregation by measuring twenty indexes based on data for US metropolitan areas for the period 1980-2000 and using principal components factor analysis. They conclude that residential segregation can be measured across two dimensions, namely separation which encompasses unevenness, isolation and clustering, and location which captures centralisation and concentration.

Building on Reardon and Firebaugh (2002b) and Grammis (2002), and the earlier reviews byJames and Taeuber (1985) and Massey and Denton (1988), Reardon and O'Sullivan (2004: 131-136) recast criteria for aspatial indexes into an equivalent form for aggregate multi-group spatial indexes (see also Watts, 2005 for a detailed discussion of both aspatial and spatial indexes). Until these papers, the development of spatial indexes of segregation had been fragmented and lacked a coherent conceptual basis (see, for example, Jakubs, 1981; Morgan, 1983; Morrill, 1991; and Wong, 1993, 2002, 2003).

1. Scale interpretability:

If the group proportions are the same in the local environment of each individual, then the spatial segregation index should take the value zero, whereas a segregation index should reach its maximum value (typically normalised at unity), if the local environment of each individual is mono-racial. This criterion is motivated by the desire to make universal comparisons of index magnitudes across time and space. However, if the chosen index is not margin-free (see below), then simple index comparisons can be misleading.

2. Arbitrary boundary independence:

A spatial segregation measure should be independent of the definitions of areal units. This demanding criterion requires that all individuals be precisely located in space and there is a unique and exhaustive set of spatial proximities for all pairs of locations. The MAUP would be overcome, but the scale at which segregation is measured would remain unresolved.

3. Location equivalence:

Points p,q,… within a tract can be aggregated if their population compositions are the same within their local environments, and they satisfy $\phi(p,q) = c$ and $\phi(p,s) = \phi(q,s)$ for all points s lying outside the tract. These conditions are demanding and are unlikely to be satisfied. Location Equivalance is the spatial analogue of Organisation Equivalence, which enables the aggregation of units, such as occupations and schools, which have the same gender or race composition (Reardon and O'Sullivan, 2004: 132-133).

4. Population density invariance:

If the population density of each group m at each point p is multiplied by a constant factor, then segregation is unchanged.

5. Composition invariance:

If the number of individuals in a particular group increases uniformly across all locations within the region, and the numbers and spatial distribution of all other groups remain unchanged, then the measure of segregation is unchanged.

6. Transfers:

The principle is based on the Pigou-Dalton principle which appears in the income inequality literature (e.g. Shorrocks and Wan, 2005). A group m individual is transferred from point p to q, without replacement. If the proportion of group m in the local environment of all points closer to p than q is greater than the proportions of group m in points closer to point q, then segregation is reduced (Reardon and O'Sullivan, 2004: 134-135). This criterion is problematic. First the point densities are upset by the one-way transfer. Second points close to both p and q may all be dominated by group m, (relative to its overall population share), so that the transfer reduces the group m dominance of points close to p and increases dominance for points close to q. As a consequence the criterion requires that the index be non-linear. Finally none of the indexes analysed by the authors satisfy the criterion (p.151), so it is ignored.

7. Exchanges:

Reardon and O'Sullivan (2004: 135) identify two forms of exchange in which a group m individual at p is replaced by a group n individual from q. In the first case points close to p have a greater proportion of group m and points near point q have a greater proportion of group n. Compliance with this condition implies a non-linear index. The second version involves the same exchange but the proportions of group m at points near to and including point p are greater than the corresponding proportions of group n at these points, and conversely for point n.

The Information Theory and Relative Diversity Indexes only satisfy both exchange criteria, if spatial symmetry is assumed, which can be satisfied by two forms of geography, both of which are highly restrictive (Reardon and O'Sullivan, 2004: 156-157). A weak version of both types of Exchange is satisfied by the Index of Dissimilarity, but again spatial symmetry must be assumed. Consequently this criterion is also ignored.

8. Additive Spatial Decomposability

This criterion requires that the grouping of points (areal units) into fewer, larger areas enables total spatial segregation to be represented by the sum of within area and between area components. Reardon and O'Sullivan (2004: 147-149) acknowledge that an interaction term would need to be included in the decomposition to reflect the impact on the within area (say k) measure arising from points, q outside k via the proximity function, $\phi(p,q)$ where $p \in k$.

If significant in size, the interaction term clouds the interpretation of the other two components of the decomposition. Also, the within group index measures use the within group population shares as the benchmark, rather than the regional population shares, so that their interpretation is not straightforward (Brown and Chung, 2006).

9. Additive Grouping Decomposability

If M groups are clustered in N (<M) supergroups, then a segregation measure should be decomposable into a sum of independent within- and between-supergroup components. Again local benchmarks are being employed in the within group component. This condition is only satisfied by the Information Theory Index.

The set of criteria which spatial segregation indexes should satisfy must be aligned with the particular objectives of the empirical research. In particular, any time series analysis of spatial segregation across a region would be difficult to interpret due to the impact on the index magnitude of changing overall population shares and the associated changes in the population densities across the local environments. An index that neutralises the impact of these changes is said to be margin-free (independent of the margins). This requires both Composition Invariance (see above) as well as Unit Invariance, that is the index measure is independent of the population density within the local environment of a point as long as the group composition of the local population stays unchanged.

Dawkins (2004: 837) rejects unit invariance as a criterion at least for the measurement of residential segregation. He notes that Fluckiger and Silber (1999) do not support the requirement that the index be margin free, since the margin totals impact on both the dispersion and shape of the distribution of segregation ratios, which contribute to the overall pattern of segregation. He also argues that the margin free requirement is mainly advocated by researchers studying labour force segregation, which can be more easily specified in 'standardised units of analysis' whereas residential segregation is often measured across neighbourhood units that are defined in 'highly idiosyncratic ways' (Charles and Grusky, 1995). This is an unpersuasive argument because these invariance principles are designed to enable rigorous comparisons over time.

Dawkins is indirectly highlighting the point that cross-sectional comparisons of residential segregation are highly problematic, even if the MAUP is ignored, given the vagaries of the underlying spatial units which cannot be considered as equivalent say across cities, because they are not uniform in area or in population and may be unequal in number too. As noted above, choosing an index which exhibits Scale Interpretability does not solve the problem of making meaningful cross-section comparisons. These problems remain for cross-national or cross-regional studies of (aspatial) occupational segregation, unless a uniform occupational classification is adopted.

Unless a smoothing procedure is adopted, even time series comparisons of residential segregation measures for a particular region are likely to be unsatisfactory due to the changing spatial definitions of the underlying units and possibly the number of units making

up a particular regional area. By imposing a consistent locational grid on the region, KDE does overcome the problem of changing spatial definitions, which the simple construction of local environments around the sample points does not. However the unit invariance issue is not neutralised through the construction of a locational grid, because the point population densities not only differ but they change in relative terms over time.

Watts (2003) showed that the aspatial analogues of Locational Equivalence and Unit Invariance are in fact inconsistent. This means that the adoption of a margin free index is unsatisfactory. The aspatial log index recommended by Grusky and Charles (1998) for the analysis of occupational segregation has serious deficiencies (Watts, 1998b). The alternative is to adopt a decomposition procedure, as argued by Karmel and MacLachlan (1988) and Watts (2003) with respect to occupational segregation. This approach can be adopted with non-margin-free indexes defined over smoothed spatial data.

In a series of papers (2001, 2002a,b, 2006, 2007), Forrest, Johnston and Poulsen address the measurement issue and explore the pattern of ethnic segregation in cities, including Sydney, Auckland and New York.. They argue that global indexes over-simplify complex phenomena, and that making cross section or time series comparisons based on index measurement is difficult, due to indexes being relative, in particular those representing the five dimensions of segregation (Poulsen, Johnston and Forrest, 2002: 246).

Measures of ethnic concentration should i) maintain a close link between measurement and theory; ii) enable comparability both within a city and across ethnic groups and among cities at the national and international level; iii) represent absolute indicators which directly illuminate the segregation issue (Johnston, Poulsen and Forrest, 2002: 246).

Poulsen, Johnston and Forrest (2002) argue that there should be a focus on three aspects of the residential patterns of ethnic groups, namely i) residential concentration, ie the areas in which an ethnic group predominates; ii) the extent of assimilation, that is the sharing of residential space with the host society; and iii) encapsulation, that is the residential isolation of a particular ethnic group from the host group and other ethnic groups.

Johnston, Poulsen and Forrest (2002) construct a concentration schedule depicting the total percentage of the particular ethnic population (vertical axis) residing in tracts where this ethnic group represents more than a prescribed share of the local population. The schedule extends Peach's work beyond a single threshold. Johnston, Poulsen and Forrest (2002: 248) claim that the schedule provides data as to the extent of spatial segregation at a specified threshold, which is comparable over time and space. This metric is alleged not to be affected by either the group's size or changes in its size over time. Poulsen, Johnston and Forrest (2002: 233) acknowledge that the thresholds are to some extent arbitrary. In a later paper, (Poulsen and Johnston, 2006) a typology with 6 ranges of concentration which are mutually exclusive is adopted.

Watts (2007) argues that both representations of concentration are sensitive to overall population shares and conflate the dimensions of unevenness and isolation, because time series comparisons of concentration reflect both changes in population shares of the ethnic groups, and changes in the spread of the particular ethnic group across the tracts. Also, both representations are aspatial, because by treating space just as a source of difference, they fail to capture the spatial relationships between areas, as well as being sensitive to the MAUP. While in a later study, (Johnston et al, 2007), they argue that the five dimensions of segregation can be reduced to two, it still makes little sense to construct a measure which conflates the key dimensions of unevenness and isolation.

4. Index measurement and data

In this paper, in the absence of 2006 Census data, which would allow time series comparisons, we will adopt simple bivariate indexes of unevenness and exposure, which are considered the key dimensions of spatial segregation. The Index of Dissimilarity (ID)

(Duncan and Duncan, 1955) and the Exposure index (Lieberson, 1981) are used, mainly because they are relatively easy to interpret and are frequently used. Despite enthusiastic endorsement by Reardon and O'Sullivan (2004), the Information Theory index, which draws on Theil (1972), is not employed, because the additional criteria (as compared to ID) that it unconditionally satisfies, namely Exchanges (type 2) and the two Additive Decomposability Properties are of limited value.

ID can be written as

$$ID = \sum_{i=1}^{L} \left| H_i / H - N_i / N \right| / 2 \tag{4}$$

where H_i, N_i respectively represent the size of the Host and Minority populations within area i. ID measures the total percentage of one group which must be removed without replacment to equalise the population distributions across all locations 1,2,…L. It has a maximum value of unity, if all areas are monoracial, and a minimum value of zero.

The Exposure index measures the degree of potential contact, or interaction, between a minority and a host group (; Massey and Denton, 1988: 287). Thus, if members of the two groups share common areas, the index will be quite large. The exposure of the host population to the minority population can be written as:

$$_H P_N = \sum_{i=1}^{L} (H_i / H) N_i / t_i \tag{5}$$

where t_i denotes the total population in area, i. The first term is the probability that a member of the host population chosen at random resides in area i and the second term denotes the probability that a member of the minority population is encountered in area i. The index has a maximum value of N/T where T is the total population of the region. This occurs if the minority population is distributed uniformly across the L areas so that (N_i/t_i) is uniform (i = 1,2,…L) and hence equal to N/T. Thus this index is sensitive to overall population shares.

We utilise 2001 Census data of population by country of birthplace for 6991 Collector Districts spread across 52 Statistical Local areas which are largely self-contained with respect to commuting patterns. Thus the residents of this region work within the region. Seven CDs (1210715, 1280214, 1291808, 1312403, 1410215, 1412610, 1421904) had zero population leaving 6984 observations. The remaining CDs had a minimum population of seven and a maximum of 2336. Consistent with the work of Poulsen et al (2004), we define the host population as native born Australians and residents who were born in English speaking countries, namely Canada, Ireland, New Zealand, United Kingdom and United States. NESB residents represent 19.13% of the local population, which represents the maximum value of $_H P_N$. Poulsen et al (2004) also utilise ancestry data, but those data will be employed in later work.

The latitude and longitude are available for the centroid of each Collector District. Using M_Map v1.4 code in Matlab devised by Rich Pawlowicz at the University of British Columbia, the latitudes and longitudes were projected onto Cartesian space using the Lambert Conformal Conic, which appears well suited for this relatively small area. . The location grid is defined within minimum and maximum values of the x coordinate (-0.0136, 0.0101) and y coordinate (-0.0189, 0.0129). Each range is divided into an equal number of segments ranging from 30 to 80, yielding a locational grid of between 900 and 6,400 points. The bandwidth is adjusted in line with changes in the number of segments to maintain a ratio to the grid resolution on the x axis of between 2.5 and 5.0. Values of the kernel parameter α between 0.25 and 2.50 are adopted in increments of 0.25.

The weighting procedure (2) does not maintain the regional population over the locational grid. Uniform adjustment of the respective population numbers across each location to bring them into line with the raw data was undertaken, but due to Composition Invariance, the ID

index is unaffected, whereas the Index of Isolation is sensitive to relative population numbers.

5. Results

Tables 1-3 shows the magnitudes of the ID and the exposure indexes for the different parameter values. A number of general observations can be made about the computed ID measures. First, increasing the number of points in the locational grid leads to a consistent increase in the ID index, which is a well known result. Second, reducing the bandwidth around each point for a given number of points on the locational grid, increases the ID index due to reduced smoothing, which is a well-known result. Third, higher values of the exponent, α, yield a more peaked kernel and a lower value of the ID index, except for a parameter value of 2.50. Fourth, the final choice of location grid yields 6,400 points, relatively close to the number of sample observations, but the levels of the computed index ID for the different levels of α are significantly below the original ID value of 0.3759 for the 6984 CDs, which demonstrates the impact of smoothing.. Local values of an appropriate index of unevenness over the location grid would also be informative.

Turning to the interaction indexes: First, increasing the number of points in the locational grid along with changing the bandwidth consistently reduces the index magnitude. Second, higher values of the exponent, α, yield lower values of the interaction indexes. Third, in contrast to the ID index, the exposure indexes assume greater values than those associated with the 6,984 CDs of 0.1651 and 0.6976.

6. Conclusion

Historically the measurement of residential segregation has been contested because there was no unanimity about its conceptualisation although writers now increasingly focus on measures of unevenness and exposure. However differences persist with respect to both the appropriateness and importance placed on different criteria, which reflect, in part, the absence of a clear articulation of the empirical application for a particular measure.

In this paper it is argued that meaningful comparisons across space of the extent of segregation cannot be necessarily rigorous, even if the chosen index lies between 0 and 1. On the other hand, an index that can be decomposed to neutralise the impact of the margins can yield insights about overall trends in segregation, if the relevant data can be analysed across a fixed grid of points in space, following a smoothing process. Smoothing will also address the MAUP and the checkerboard problem.

This paper has examined a particular form of spatial smoothing used in the UK (Martin et al, 2000), after projection of the data onto Cartesian coordinates. Across different parameter values, the sets of index values are relatively consistent, but there is no guidance a priori as to the appropriate parameter values. The analysis of data from both the 2001 and 2006 Censuses will provide a better indication as to whether the underlying Kernel specification yields consistent time series results. In addition, the scale of segregation, the edge problem associated with Kernel estimation (Cowling and Hall, 1996) and the adoption of integration of index values will be addressed in future work.

References

Brown, L.A. and Chung, S.Y. (2006) 'Spatial Segregation, Segregation Indices and the Geographical Perspective', *Population, Space and Place*, 12, 125-143.

Blackburn, R.M., Jarman, J. & Siltanen, J. (1993) 'The Analysis of Occupational Gender Segregation over Time and Place: Considerations of Measurement and Some New Evidence', *Work, Employment and Society*, 7(3), 335-362

Charles, C.Z. (2003) 'The Dynamics of Racial Residential Segregation', *Annual Review of Sociology*, 29, 167-207.

Charles, M. and Grusky, D.B. (1995) 'Models for Describing The Underlying Structure of Sex Segregation', *American Journal of Sociology*, 100(4), 931-971.

Cowling, A. and Hall, P. (1996) 'On Pseudodata Methods for Removing Boundary Effects in Kernel Density Estimation' *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(3), 551-563.

Dawkins, C.J. 2004. 'Measuring the Spatial Pattern of Residential Segregation.' *Urban Studies*, 41(4), 833-851.

Deming, W. and Stephan, F. (1940) 'On a least squares adjustment of sampled frequency table when the expected marginal totals are known.' *Annals of Mathematical Statistics*, 11, 427-44.

Duncan, O.D. and Duncan, B. (1955) 'A Methodological Analysis of Segregation Indexes.' *American Sociological Review*, 20, 210-217.

Flückiger, Y. and Silber, J. (1999) *The Measurement of Segregation in the Labour Force*, Physica-Verlag: Heidelberg.

Grannis, R. (2002) 'Discussion: Segregation Indices and Their Functional Inputs', in Stolzenberg, R.M. (ed.) *Sociological Methodology*, 32, 69-84, Boston, MA: Blackwell Publishing.

Grusky, D.B. and Charles, M. (1998) 'The Past, Present and Future of Sex Segregation Methodology.' *Demography* 35, 497-504.

Healy, E. and Birrell, B. (2003) 'Metropolis Divided: The Political Dynamic of Spatial Inequality and Migrant Settlement in Sydney', *People and Place*, 11(2), 65-87.

Holt, J.B., Lo, C.P. and Hodler, T.W. (2004) 'Dasymetric Estimation of Population Density and Areal Interpolation of Census Data', *Cartography and Geographic Information Science*, 31(2), 103-121.

Jakubs, J.F. (1981) 'A Distance-Based Segregation Index', *Journal of Socioeconomic Planning Services*, 15, 129-136.

James, D.R. and Taeuber, K.E. 1985. 'Measures of Segregation.' In Tuma, N.B.(ed.) n *Sociological Methodology*, 14: 1-32, San Francisco, CA: Jossey-Bass.

Johnston, R., Forrest, J. and Poulsen, M. (2001) 'Sydney's Ethnic Geography: New Approaches to Analysing Patterns of Residential Concentration', *Australian Geographer*, 32(2), 149-162.

Johnston, R., Poulsen, M. and Forrest, J. (2002) 'Rethinking the Analysis of Ethnic Residential Patterns: Segregation, Isolation or Concentration Thresholds in Auckland, New Zealand?', *Geographical Analysis*, 34(3), 245-261.

Johnston, R., Poulsen, M. and Forrest, J. (2007) 'Ethnic and Racial Segregation in US Metropolitan Areas, 1980-2000: The Dimensions of Segregation Revisited', *Urban Affairs Review*, 42(4), 479-504.

Karmel, T. & Maclachlan M. (1988) 'Occupational Sex Segregation - Increasing or Decreasing.' *Economic Recor,d* 64, 187-195.

Lee, B.A., Reardon, S.F., Firebaugh, G. Farrell, C.R., Matthews, S.A. and D. O'Sullivan (2006) 'Beyond the Census Tract: Patterns and Determinants of Racial Residential Segregation at Multiple Scales', *mimeo*, Department of Sociology. Penn State University, PA, USA 16802.

Lieberson, S. (1981) 'An Asymmetrical Approach to Segregation', in *Ethnic Segregation in Cities,* edited by Peach, C., Robinson, V. and Smith, S., 61-82 Croom-Helm.

Martin, D., Tate, N.J, and Langford, M. (2000) 'Refining Population Surface Models: Experiments with Northern Ireland Census Data', *Transactions in GIS*, 4(4), 343-360.

Massey, Douglas S., and Nancy A. Denton (1988) 'The Dimensions of Residential Segregation', *Social Forces* 67, 281–315.

Morgan, B.S. (1983a). 'An Alternate Approach to the Development of a Distance-Based Measure of Racial Segregation.' *American Journal of Sociology* 88, 1237–49.

Morgan, B.S. (1983b) 'A Distance-Decay Interaction Index to Measure Residential Segregation', *Area* 15, 211–16.

Morrill, R. L. (1991) 'On the Measure of Spatial Segregation.', *Geography Research Forum*, 11, 25–36.

O'Sullivan, D. and Wong, D.W.S. (2007) 'A Surface-Based Approach to Measuring Spatial Segregation', *Geographical Analysis*, 39, 147-168.

Pawlowicz, R. (2005) *M_Map: User's Guide v1.4*, University of British Columbia, http: //www2.ocgy.ubc.ca/~rich/private/mapug.html#p2

Peach, C. (1996) 'The Meaning of Segregation', *Planning Practice and Research*, 11(2), 137-150.

Poulsen, M., Johnston, R. and Forrest, J. (2002) 'Plural Cities and Ethnic Enclaves: Introducing a Measurement Procedure for Comparative Study', *International Journal of Urban and Regional Research*, 26(2), 229-243.

Poulsen, M., Johnston, R. and Forrest, J. (2004) 'Is Sydney a Divided City Ethnically', *Australian Geographical Studies*, 42(3), 356-377.

Poulsen, M. and Johnston, R. (2006) 'Ethnic Residential Segregation in England: getting the right message across', *Environment and Planning A*, 38, 2195-2199.

Reardon, S.F., and Firebaugh, G. (2002a) 'Measures of Multigroup Segregation', in Stolzenberg, R.M. (ed.) *Sociological Methodology*, 32, 33-67, Boston, MA: Blackwell Publishing.

Reardon, S.F., and Firebaugh, G. (2002b) 'Response: Segregation and Social Distance – A Generalised Approach to Segregation Measurement', in Stolzenberg, R.M. (ed.) *Sociological Methodology*, 32, 85-101, Boston, MA: Blackwell Publishing.

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

Shorrocks, A. and Wan, G. (2005) 'Spatial Decomposition of Inequality', *Journal of Economic Geography*, 5, 59-81.

Tobler, W.R. (1979) 'Smooth Pycnophylactic for Geographical Regions', *Journal of the American Statistical Association*, 74, 519-529.

Watts, M.J. 1992. 'How Should Occupational Sex Segregation be Measured?.' *Work, Employment and Society*, 6, 475-487.

Watts, M.J. (1998a) 'Occupational Gender Segregation: Index Measurement and Econometric Modelling.' *Demography*, 35(4), 489-96.

Watts, M.J. (1998b) 'The Analysis of Sex Segregation: When is Index Measurement not Index Measurement?', *Demography*, 35(4), 505-508.

Watts, M.J. (2003) 'The Evolution of Occupational Gender Segregation in Australia: Measurement and Interpretation', *Australian Journal of Labour Economics*, 6(4), 631-655

Watts, M.J. (2005) Conceptualising and Measuring Spatial Segregation: The State of Play', paper presented to *Joint Paradigms Forum: Methodology, Tools and Techniques*, ARC Research Network in Spatially Integrated Social Science, Newcastle, NSW, 15-17 June.

Watts, M.J. (2007) 'Correctly Measuring Ethnic Residential Segregation: A Critique', *mimeo*, Centre of Full Employment and Equity, The University of Newcastle.

White, M.J. (1983) 'The Measurement of Spatial Segregation', *American Journal of Sociology*, 88, 1008-18.

White, M.J. (1986) 'Segregation and Diversity: Measures of Population Distribution', *Population Index*, 52, 198-221.

Wong, D.W. (1993) 'Spatial Indices of Segregation', *Urban Studies* 30, 559–572.

Wong D.W. (2002) 'Spatial Measures of Segregation and GIS', *Urban Geography* 23, 85–92.

Wong D.W. (2003) 'Spatial Decomposition of Segregation Indices: A Framework toward Measuring Segregation at Multiple Levels', *Geographical Analysis* 35, 179–194.

Table 1 ID Index Values for Different Parameter Values of the KDE.

	B/width	Exponent αof Kernel equation (2)									
		0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50
30*30	* 5	0.1929	0.1966	0.2011	0.2057	0.2103	0.2145	0.2181	0.2212	0.2242	0.2276
	* 2.5	0.2488	0.2533	0.2571	0.2603	0.2630	0.2657	0.2683	0.2706	0.2727	0.2744
40*40	* 5	0.2178	0.2221	0.2260	0.2307	0.2348	0.2384	0.2414	0.2446	0.2477	0.2507
	* 2.5	0.2668	0.2698	0.2723	0.2750	0.2777	0.2802	0.2825	0.2848	0.2868	0.2887
50*50	* 5	0.2356	0.2396	0.2436	0.2474	0.2511	0.2544	0.2575	0.2604	0.2631	0.2654
	* 2.5	0.2784	0.2814	0.2838	0.2860	0.2883	0.2903	0.2921	0.2937	0.2952	0.2967
60*60	* 5	0.2482	0.2525	0.2563	0.2598	0.2628	0.2657	0.2683	0.2706	0.2729	0.2749
	* 2.5	0.2864	0.2892	0.2915	0.2935	0.2954	0.2973	0.2994	0.3013	0.3033	0.3050
70*70	* 5	0.2582	0.2623	0.2658	0.2689	0.2716	0.2742	0.2764	0.2785	0.2804	0.2822
	* 2.5	0.2919	0.2944	0.2970	0.2993	0.3015	0.3035	0.3055	0.3072	0.3088	0.3104
80*80	* 5	0.2664	0.2697	0.2727	0.2756	0.2783	0.2806	0.2827	0.2847	0.2866	0.2883
	* 2.5	0.2978	0.3004	0.3026	0.3049	0.3072	0.3093	0.3111	0.3128	0.3146	0.3162

Notes: Bandwidths set at multiples of 2.5 and 5.0 of the grid resolution.
Source: Table B06 (Collector Districts), Basic Community Profiles 2001 Census

Table 2 Exposure Index ($_HP_N$) for Host Population, H, based on Different Parameter Values for KDE

	B/width	Exponent α of Kernel equation (2)									
		0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50
30*30	* 5	0.1852	0.1849	0.1846	0.1843	0.1840	0.1837	0.1835	0.1832	0.1829	0.1827
	* 2.5	0.1813	0.1809	0.1805	0.1802	0.1798	0.1795	0.1792	0.1790	0.1788	0.1786
40*40	* 5	0.1837	0.1833	0.1830	0.1826	0.1823	0.1820	0.1817	0.1814	0.1811	0.1809
	* 2.5	0.1796	0.1792	0.1788	0.1785	0.1782	0.1780	0.1777	0.1775	0.1773	0.1771
50*50	* 5	0.1824	0.1820	0.1816	0.1813	0.1809	0.1806	0.1803	0.1800	0.1798	0.1796
	* 2.5	0.1783	0.1780	0.1777	0.1774	0.1771	0.1769	0.1767	0.1765	0.1763	0.1761
60*60	* 5	0.1813	0.1809	0.1805	0.1802	0.1798	0.1795	0.1792	0.1790	0.1788	0.1786
	* 2.5	0.1774	0.1771	0.1769	0.1766	0.1764	0.1761	0.1759	0.1757	0.1755	0.1754
70*70	* 5	0.1804	0.1800	0.1796	0.1793	0.1789	0.1787	0.1784	0.1782	0.1780	0.1778
	* 2.5	0.1768	0.1765	0.1762	0.1759	0.1757	0.1755	0.1753	0.1751	0.1749	0.1747
80*80	* 5	0.1796	0.1792	0.1788	0.1785	0.1782	0.1780	0.1777	0.1775	0.1773	0.1771
	* 2.5	0.1762	0.1759	0.1756	0.1754	0.1751	0.1749	0.1747	0.1745	0.1743	0.1742

Notes & Source: see Table 1.

Table 3: Exposure Index ($_NP_H$) for Minority Population, N based on Different Parameter Values for KDE

	B/width	Exponent α of Kernel equation (2)									
		0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50
30*30	* 5	0.7827	0.7816	0.7803	0.7791	0.7778	0.7766	0.7754	0.7742	0.7732	0.7721
	* 2.5	0.7664	0.7647	0.7630	0.7615	0.7601	0.7588	0.7576	0.7565	0.7556	0.7547
40*40	* 5	0.7764	0.7749	0.7734	0.7720	0.7705	0.7692	0.7679	0.7667	0.7656	0.7646
	* 2.5	0.7589	0.7573	0.7558	0.7545	0.7533	0.7522	0.7512	0.7503	0.7495	0.7487
50*50	* 5	0.7709	0.7693	0.7677	0.7661	0.7647	0.7634	0.7621	0.7610	0.7599	0.7589
	* 2.5	0.7537	0.7522	0.7509	0.7498	0.7487	0.7478	0.7469	0.7460	0.7452	0.7445
60*60	* 5	0.7663	0.7646	0.7630	0.7614	0.7600	0.7588	0.7576	0.7565	0.7556	0.7547
	* 2.5	0.7500	0.7487	0.7475	0.7464	0.7454	0.7444	0.7435	0.7427	0.7419	0.7412
70*70	* 5	0.7624	0.7607	0.7591	0.7576	0.7563	0.7551	0.7541	0.7531	0.7522	0.7514
	* 2.5	0.7472	0.7459	0.7447	0.7436	0.7426	0.7416	0.7407	0.7399	0.7392	0.7384
80*80	* 5	0.7590	0.7574	0.7559	0.7545	0.7533	0.7522	0.7512	0.7503	0.7495	0.7487
	* 2.5	0.7446	0.7434	0.7423	0.7412	0.7402	0.7393	0.7384	0.7376	0.7369	0.7362

Notes & Source: see Table 1.

[1] Research Associate, Centre of Full Employment and Equity, University of Newcastle.