Urban Studies
http://usj.sagepub.com/

# Are Big Cities More Segregated? Neighbourhood Scale and the Measurement of Segregation

Douglas J. Krupka

*Urban Stud* 2007 44: 187

DOI: 10.1080/00420980601023828

The online version of this article can be found at:
http://usj.sagepub.com/content/44/1/187

Published by:

SAGE

http://www.sagepublications.com

On behalf of:

Urban Studies Journal Foundation

**Additional services and information for *Urban Studies* can be found at:**

**Email Alerts:** http://usj.sagepub.com/cgi/alerts

**Subscriptions:** http://usj.sagepub.com/subscriptions

**Reprints:** http://www.sagepub.com/journalsReprints.nav

**Permissions:** http://www.sagepub.com/journalsPermissions.nav

**Citations:** http://usj.sagepub.com/content/44/1/187.refs.html

>> Version of Record - Jan 1, 2007

What is This?

Downloaded from usj.sagepub.com at Glasgow University Library on October 17, 2014

Urban Studies, Vol. 44, No. 1, 187–197, January 2007

Routledge
Taylor & Francis Group

# Are Big Cities More Segregated? Neighbourhood Scale and the Measurement of Segregation

Douglas J. Krupka

*[Paper first received, January 2006; in final form, May 2006]*

**Summary. Many studies show that larger metropolitan areas are more segregated than smaller ones. To some extent, this tendency is part of the conventional wisdom. However, the reason for this tendency is not apparent. This paper suggests that the correlation between segregation and metropolitan scale is spurious. Segregation measures based on census data will tend to rank larger cities higher because larger cities have more neighbourhoods that are big enough to 'fill up' entire census tracts, while smaller cities with equally homogeneous, but smaller, neighbourhoods have to pair neighbourhoods to fill up a census tract. This bias will be reduced at smaller levels of spatial aggregation. This prediction is tested by comparing segregation measures computed at several levels of spatial aggregation and with American Housing Survey data. The results suggest that spatial aggregation effects are important: the correlation between city size and measured segregation appears to be at least partly spurious.**

## 1. Introduction

It is probably an understatement to say that racial and economic segregation are important issues in America. One stylised fact about segregation is that large cities tend to have more segregated residential patterns. Farley's (1991) ranking of the most and least racially segregated metropolitan areas in America shows across the middle of the century that large cities (most famously Chicago) tend to fill out the ranks of the top 10, while smaller cities (such as Lawrence, Kansas), fill the ranks of the least segregated. Across 20 years of data, this pattern does not change. Jargowski (1996) examines economic segregation, finding that larger Metropolitan Statistical Areas (MSAs) are indeed more economically segregated than smaller ones: all else equal, a 10 per cent increase in population is associated with about a five percentage point increase in his measure of segregation. This paper advances the hypothesis that the correlation between city size and segregation is spurious because the 'neighbourhood', whatever it is, is not represented anywhere in the census data used to compare segregation across cities.[1]

To some extent, this empirical regularity is a part of the popular perception about segregation in America: small towns are supposed to be egalitarian, while the big city is supposed to be highly stratified. We are aware of no convincing economic argument why this would be the case.[2] While many authors have examined the causes[3] or effects[4] of segregation, as yet no author has explained why segregation should be correlated with urban scale. Why is it easier, more cost-effective or efficient to segregate 1 million Black or

*Douglas J. Krupka is in the Department of Economics, Georgia State University, 14 Marietta Street NW, Suite 427, Atlanta, Georgia 30303, USA. Fax: 404 651 0416. E-mail: dkrupka@gsu.edu.* *The author would like to thank Geoffrey Turnbull, Douglas Noonan and Erin Krupka for helpful comments. All errors are the author's.*

0042-0980 Print/1360-063X Online/07/010187–11 © 2007 The Editors of Urban Studies
DOI: 10.1080/00420980601023828

Downloaded from usj.sagepub.com at Glasgow University Library on October 17, 2014

poor people than 10 000 of them? While theories stressing the costs and benefits of neighbourhood formation, with some inclusion of a range of increasing returns to neighbourhood size, could explain the pattern,[5] it seems unlikely that these explanations seem relevant in the metropolitan context. After all, only two neighbourhoods are needed to segregate perfectly Blacks from Whites. While some very small towns may be described as having only two neighbourhoods (no doubt on opposite sides of the train tracks), certainly no metropolitan area in our census data suffers from such a neighbourhood constraint.

This paper contends that this empirical regularity is largely an artifact of census data collection. Census tracts are generally of a certain size: they tend to contain between 3000 and 5000 people. In a metropolis like Chicago, a segregated neighbourhood may contain scores of such homogeneous census tracts. In a smaller city like Peoria, IL, the segregated neighbourhood may not even be large enough to fill out an entire census tract. In essence, census-tract boundaries can be set tightly around the segregated group in a large metropolitan area and still be of sufficient size. In a smaller metropolitan area, the segregated area must be paired with a dissimilar one in order to produce an area with population sufficient to make up a census tract. The result is that the smaller metropolitan area looks more mixed than the larger one, even though both have their minorities perfectly segregated to one neighbourhood or more.

The implication of this data constraint is that, in comparing segregation across cities, big cities should look more segregated than small cities, but that the difference will not be as great when the spatial unit of analysis is smaller. At a sufficiently small area of analysis, there should be no difference between large cities and small ones.

We carry out just such a sequence of comparisons. We compute indices of segregation for both racial and economic segregation at several levels of spatial aggregation: the county, zip code, census tract and block group. Unfortunately, this sequence cannot be taken down to the block level or below with census data due to confidentiality issues. We are able to take the analysis a step further using a special wave of the American Housing Survey (AHS) in which clusters of around 11 closest neighbours were interviewed. These very small neighbourhood clusters offer a way to observe directly the probability that two neighbours will be of different races or income classes. Using these neighbourhood clusters, we can examine whether these small neighbourhoods really are more diverse in smaller metropolitan areas.

The rest of the paper is laid out as follows: Section 2 briefly elaborates on the conceptual discussion above. Section 3 describes and carries out the analysis of the census data, while section 4 carries out the analysis with the AHS data. Section 5 concludes.

## 2. Spatial Scale and Segregation Measures

Measuring racial and economic segregation is difficult because neighbourhoods form over space, which is continuous, while our data are constrained to be discrete. Our models of urban areas feature continuous functions, while real cities are usually somewhat lumpy. Nonetheless, measurement problems should not rule out measurement, only suggest caution in our interpretation.

There is a deep literature on the measurement of segregation.[6] This literature has focused on the measurement of racial segregation. This is unfortunate because urban economic theory provides much cleaner predictions about economic segregation. In measuring racial segregation, we follow the advice of Massey and Denton (1988) in using the dissimilarity index for two groups. While other measures—and indeed other dimensions—of segregation exist, in general they are highly correlated with the dissimilarity index, which is defined in equation (1).

$$D = \frac{\sum N_i |b_i - b|}{2Nb(1-b)} \tag{1}$$

where, $N_i$ is the neighbourhood population; $N$ is the metropolitan area population; $b_i$ is the neighbourhood percentage Black; and $b$ is

Downloaded from usj.sagepub.com at Glasgow University Library on October 17, 2014

the metropolitan area percentage Black. This index ranges from zero (when all neighbourhoods have the same racial mix) to one (when all Blacks live in all-Black neighbourhoods) and can be interpreted as the proportion of the minority population that would have to move to create an even racial distribution across all neighbourhoods.

To measure economic segregation, the dissimilarity index is not sufficient because income is not binary. Instead of making arbitrary cut-offs between 'rich' and 'poor' households, we use the neighbourhood sorting index introduced by Jargowski (1996)

$$J = \frac{\left(H^{-1}\sum_n H_n(\bar{y}_n - \bar{y})^2\right)^{1/2}}{\left(H^{-1}\sum_h (y_h - \bar{y})^2\right)^{1/2}} \tag{2}$$

where, $h$ indexes households; $n$ indexes neighbourhoods; $y$ is income; $H$ is the number of households; the bar signifies a mean and the lack of a subscript signifies a metropolitan-level variable. Like the dissimilarity index, Jargowski's index ranges between zero (when every neighbourhood has the same average income) and one (when every neighbourhood has zero variance in income). The index can be interpreted as the percentage of the standard deviation in income that comes from variation *across* neighbourhoods as opposed to *within* neighbourhoods. If the absolute value operator in the dissimilarity index were replaced with a squaring and the sum were raised to the half power, these two indices would be equilvalent.

Whatever desirable qualities these indices have, both suffer from the modifiable areal unit problem (MAUP).[7] Specifically, estimates of either of these segregation measures will be sensitive to the level of spatial aggregation at which the data are collected.[8] To see this, consider the case when we take the entire metropolitan area as our 'neighbourhood'. It is obvious that the one neighbourhood average must equal the metropolitan average and both indices must be zero. At the other extreme, when we take the individual or household as our 'neighbourhood', it is clear that, for $D$, every individual will be coded as either White or Black,[9] so the area is perfectly segregated. For $J$, each neighbourhood average in the numerator will be equal to the household measure in the denominator, and Jargowski's index must equal one.

The modifiable areal unit problem is truly a problem, in that there is no solution to it. We must just keep it in mind when we interpret our data. We might fail to keep this in mind when interpreting the empirical regularity, highlighted above, that larger cities tend to be more segregated.

Figure 1 demonstrates this with three metropolitan areas. At levels of spatial aggregation larger than seven counties, all of these MSAs will be perfectly integrated: none of them has more than one spatial unit or 'neighbourhood'. Once our spatial scale goes below the seven county mark, Chicago will have more than one neighbourhood. To the extent that Chicago's two neighbourhoods are different, it will have a positive measure of segregation, but the two smaller cities, still having only one neighbourhood each, continue to measure no segregation. As our level of spatial aggregation reduces, each of the smaller MSAs begins to register positive segregation, but the larger cities will already have higher segregation scores. The measure of segregation in all three cities will increase as our spatial units become smaller, and all three are constrained to become equal to one when the 'spatial' unit of analysis becomes the individual census respondent.

Since every city is constrained to register zero segregation when the "neighbourhood" is defined to be the entire MSA and total segregation when the "neighbourhood" is the individual or household, an important implication becomes evident: if two cities are simply scale doubles of each other, with equally stark racial and economic separation, but at different scales (i.e. the shape of the lines in Figure 1 are the same), then the larger city will appear more segregated at every level of spatial aggregation except the individual. The reason for this is that a smaller city's segregated neighbourhoods might be so small that they must share an areal unit with another neighbourhood, while a big city has segregated neighbourhoods so

Downloaded from usj.sagepub.com at Glasgow University Library on October 17, 2014

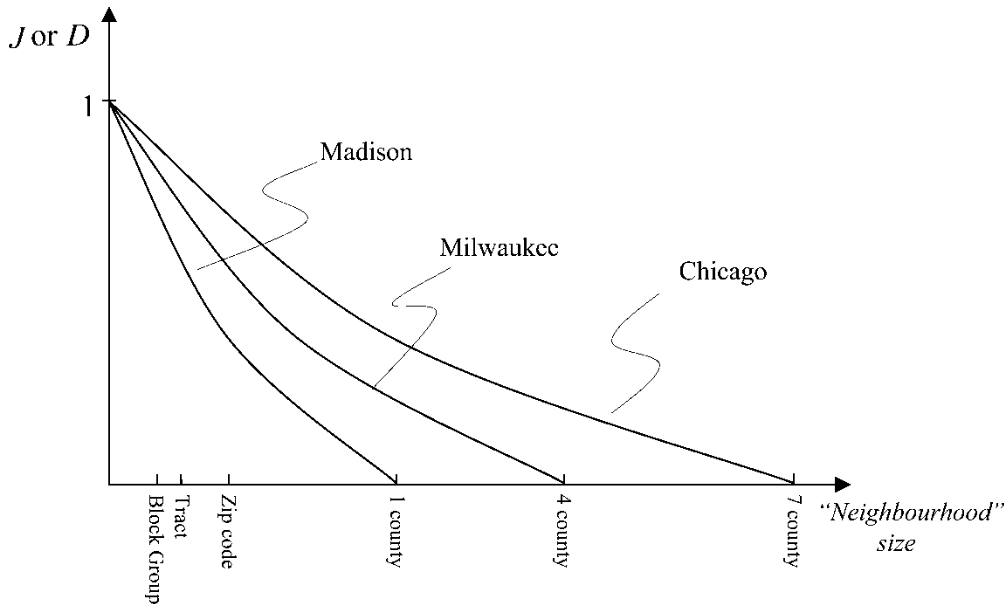


**Figure 1.** Urban scale and segregation measures.

large that they can have several units to themselves.

At the very least, this implication suggests that we should approach intermetropolitan comparisons of segregation cautiously. A more segregated city would be distinguished from a less segregated city not so much by its higher segregation score, but by a different space–segregation profile. A more segregated city would reach high segregation scores more quickly relative to its size than a less segregated city. This would be the case no matter which of the many measures of segregation was used.

This effect could be unimportant. For instance, if most MSAs are perfectly integrated at the county level, then larger MSAs do not have a 'head start' over smaller MSAs as depicted in Figure 1. Most multi-county MSAs do register some segregation at the county level, however. The spatial aggregation effect could also be unimportant if neighbourhoods are not larger in larger cities, so that large cities differ from small ones only in the number—and not the scale—of neighbourhoods. While harder to test, this seems an unlikely premise.

If it turns out that this spatial aggregation effect is important, then the higher segregation measures of large cities may not arise because of scale economies in neighbourhood formation, industrial mix or any other cause, but may be an artefact of data collection. The following two sections attempt to assess the practical importance of the spatial aggregation effect, first using census data at several different levels of spatial aggregation, then using small neighbourhood clusters (smaller even than a census block) from the American Housing Survey.

## 3. Census Data

This section uses 2000 census data to compute the measures of segregation defined in section 2 at several different levels of spatial aggregation. These levels of aggregation are the county, zip code, census tract and block group levels. At each level of aggregation, we compute a simple, bivariate correlation between metropolitan population and levels of measured segregation. If the spatial aggregation effect is important, we would expect that this correlation will become smaller as the spatial scale of analysis becomes finer. This expectation arises because as the spatial unit becomes smaller, it is less likely that

Downloaded from usj.sagepub.com at Glasgow University Library on October 17, 2014

dissimilar neighbourhoods are 'falsely' grouped into the same spatial unit. The implication is *loosely* interpretable as deriving from the lines in Figure 1 getting closer together as one moves to the left.

As an additional check, we run multivariate OLS at each level to assess the partial correlation between metropolitan population and measured segregation, controlling for the number of 'neighbourhoods' each MSA has. If the spatial aggregation effect is important, we would expect that number of 'neighbourhoods' will be a better predictor of measured segregation than metropolitan population.

The results suggest that the spatial aggregation effect is important. Table 1 reports the correlation between the segregation measures and metropolitan population across several levels of spatial aggregation. These correlations are calculated on 265 MSAs and CMSAs.[10] It shows that as the defined 'neighbourhood' becomes smaller, the correlation decreases (although it is still quite high). This relationship between level of spatial aggregation and the correlation is nearly monotonic: the only exception is for Jargowski's index and the log of the number of households in the area, where the unexpectedly low correlation at the county level makes the relationship non-monotonic.

We turn next to the regression results. Using the same population of 265 metropolitan areas, we regress the indices of segregation on log population (or household count) and the log of the number of 'neighbourhoods' in the metropolitan areas. The two independent variables are highly correlated (at between about 0.75 for counties and 0.99 for census tracts). If the size of the metropolitan area's population is a cause of more segregated residence patterns, then the log population should be correlated with higher measures of segregation, even when we hold the number of 'neighbourhoods' constant (the coefficient on log population should be positive). On the other hand, if the correlation between urban scale and segregation is an artefact of the data collection process, urban population should be insignificant and the number of neighbourhoods should be the primary determinant of measured segregation (the coefficient on the log of the number of spatial units should be positive).

Table 2 presents the racial segregation results for several levels of spatial aggregation. At all three census-based levels of spatial aggregation (counties, tracts and block groups), adding population to an MSA without adding more spatial units ('neighbourhoods') actually decreases measured segregation. These effects are mostly significant. On the other hand, increasing the amount of spatial units increases measured segregation by roughly the same magnitude: a 1 per cent increase in the number of counties, tracts or block groups, holding population constant, increases measured segregation by between 0.14 units and 0.19 units. The results for zip codes are in stark contrast: the magnitude of the spatial effect is much reduced and the

**Table 1.** Segregation and population correlations

| | Correlation | | | |
|---|---|---|---|---|
| 'Neighbourhood' | *D*, Pop | *J*, HH | *D*, lnPop | *J*, lnHH |
| County | 0.3819 | 0.5134 | 0.5751 | 0.621 |
| Zip | 0.3433 | 0.4802 | 0.5506 | 0.7292 |
| Tract | 0.2995 | 0.4498 | 0.4642 | 0.6481 |
| Block | 0.2629 | 0.4117 | 0.4009 | 0.572 |

Notes: *D* and *J* are the Index of Dissimilarity and Jargowski's Neighbourhood Sorting Index respectively. They are correlated with metropolitan population for *D* and number of households (HH) for *J*, and with each population measure's log. The correlation is based on 265 large and small MSAs and CMSAs

Downloaded from usj.sagepub.com at Glasgow University Library on October 17, 2014

**Table 2.** Regression results for racial segregation

| | County | | Zip | | Tract | | Block | |
|---|---|---|---|---|---|---|---|---|
| | Coefficient | Standard error | Coefficient | Standard error | Coefficient | Standard error | Coefficient | Standard error |
| log Pop | −0.0079 | 0.0058 | 0.0251 | 0.0147 | −0.1428 | 0.0373 | −0.1438 | 0.0313 |
| log Units | 0.1394 | 0.0082 | 0.0472 | 0.0176 | 0.1938 | 0.0374 | 0.1893 | 0.0321 |
| Constant | 0.1073 | 0.0699 | −0.1391 | 0.1338 | 1.3918 | 0.3143 | 1.2528 | 0.2263 |
| Adjusted $R^2$ | 0.678 | | 0.3166 | | 0.283 | | 0.254 | |

sign on log population shifts (although the coefficient is not significant) and the constant has changed sign.

The results for economic segregation are similar. Table 3 presents these results. The number of spatial units is strongly significant and positive at all three census-based levels of spatial aggregation. The number of households in the metropolitan area still has a negative effect, but the effect is no longer significant. Again, the results for zip codes are in strong contrast to the rest of the results: the effect of population becomes positive and significant, while the spatial effect becomes insignificantly negative.

For both racial and economic segregation, the explanatory power of the models falls monotonically as the level of spatial aggregation decreases. This mirrors the results for the bivariate correlations in Table 1. The high correlation between the independent variables in these regressions raises concerns about the stability of the results across different specifications. We do not wish to advance a theory of metropolitan-level segregation here, but it is possible that metropolitan economic performance is correlated with either segregation or its measure. Because growing areas will tend to have more populous counties, zip codes, census tracts and block groups than shrinking areas (because these geographies are changed only gradually and with a lag), metropolitan growth will also have an effect on the relationship between the population of an area and the number of spatial units in it. These relationships could be biasing the results presented here.[11] To test for this possibility, and especially to assess the stability of the results to the inclusion of a potentially important omitted variable, all of the above regressions were run controlling for absolute and percentage population gain (or loss) over the 1990s. None of the coefficients reported in Tables 2 or 3 changed in sign or significance when the additional controls were added. The magnitudes of the effects were also not affected. Metropolitan growth was in general found to be negatively and significantly related to measured racial segregation and weakly correlated positively with measured economic segregation, holding urban scale constant.[12]

Together, these results suggest that the observed tendency for larger metropolitan areas to be more segregated is a result of

**Table 3.** Regression results for economic segregation

| | County | | Zip | | Tract | | Block | |
|---|---|---|---|---|---|---|---|---|
| | Coefficient | Standard error | Coefficient | Standard error | Coefficient | Standard error | Coefficient | Standard error |
| log HHs | −0.0005 | 0.0030 | 0.0629 | 0.0085 | −0.0227 | 0.0204 | −0.0097 | 0.0173 |
| log Units | 0.0701 | 0.0043 | −0.0105 | 0.0102 | 0.0641 | 0.0204 | 0.0431 | 0.0176 |
| Constant | 0.0066 | 0.0336 | −0.4636 | 0.0695 | 0.3262 | 0.1528 | 0.2748 | 0.1089 |
| Adjusted $R^2$ | 0.694 | | 0.5301 | | 0.437 | | 0.337 | |

Downloaded from usj.sagepub.com at Glasgow University Library on October 17, 2014

bias introduced by the necessity to gather data in comparably populated spatial units. An equiproportionate increase in population and spatial units would appear to lead to an increase in measured segregation, however. Two sources of doubt remain.

First, the results for zip codes are not supportive of this interpretation. We do not have an explanation for these anomalous results. One possibility is that it is a result of the collinearity discussed above; however, the results change little when controlling for metropolitan growth. Another possibility is that zip codes are presumably set to minimise USPS costs—instead of express neighbourhood boundaries as in the case of tracts and block groups—and do not change frequently enough. However, the strongly confirmatory results for counties—which reflect neighbourhoods even more poorly and change even less frequently—would seem to contradict this explanation. A third possibility is that somehow zip codes are not shaped correctly: many zip codes have extremely convoluted shapes. In addition, zip codes do not sit properly within county boundaries, so the populations (of people) used in the zip code calculations are slightly different from those used in the county and census-based geographies. We recalculated the zip code results based on groups of block groups with internal points in each zip code to test for this effect, but the results were little changed: the racial segregation zip code results improved, while the economic segregation zip code results became more anomalous.

It is possible that these outlying results represent the true effects, although it is hard to think of a reason why this would be the case. One way forward on this point would be to combine block groups into contiguous groups of two, three and so on, and examine how the results change as we move discretely from the block group spatial scale (with between 45 and 14 809 spatial units per MSA) to the county (with between 1 and 33). However, such an analysis may be sensitive to how block groups are combined and is beyond the scope of this paper.

A second source of doubt about these results is that the analyses are carried out at the aggregate level: metropolitan-wide indices are regressed on metropolitan characteristics. To see if the implications of these results carry through to the micro-neighbourhood level, we move to the next section.

## 4. Neighbourhood Clusters

The above results are suggestive, but they can only go so far due to the large spatial aggregates they employ. It is hard to know exactly what people mean when they use a term like 'segregation', especially when there are several ways to measure it and all those measures are sensitive to data construction. One reasonable interpretation of the term would be something like the following: how similar are people who live near each other? If neighbours are very similar, then it would appear that that person lives in a segregated community. All else equal, if small clusters of neighbours are more diverse, segregation would appear to be less. Since most neighbourhoods are much bigger than a house and their immediate neighbours, even in small towns, such a definition might be robust to changes in metropolitan scale.

In this section, we use such a definition to address the main question of our paper: whether larger cities are more segregated than smaller ones. We use special 1993 data in the American Housing Survey (AHS) in which the 10 closest neighbours of around 600 houses were interviewed. We take these 600 neighbourhood clusters, located in over 80 metropolitan areas, as an observation of a micro-neighbourhood and compute two measures of neighbourhood diversity in this simplified setting. To measure racial diversity, we compute an entropy-based measure of neighbourhood diversity ($EN_{Neigh}$) over 10 racial categories.[13] To measure economic diversity, we calculate the within-neighbourhood standard deviation of household income ($SD_{Neigh}$), in 1993 dollars. Both these measures are increasing in neighbourhood diversity. We are seeking to determine if neighbourhoods are really more

Downloaded from usj.sagepub.com at Glasgow University Library on October 17, 2014

homogeneous in larger metropolitan areas. Thus we estimate equation (3)

$$D_{Neigh} = \alpha_0 + \alpha_1 MetSize + \varepsilon_1 \tag{3}$$

where, $D_{Neigh}$ is either $EN_{Neigh}$ or $SD_{Neigh}$; *MetSize* is the log of metropolitan population or households in the race and income equations respectively. If large cities are more segregated, $\alpha_1 < 0$ is expected, although if the city size/segregation relationship is spurious we should expect $\alpha_1$ to be insignificant or even positive.

It may not be fair to compare neighbourhoods in small, less diverse cities with neighbourhoods in large, diverse cities. A neighbourhood with 9 White households and 1 Black household is perfectly integrated in a small town with only a 10 per cent Black population, but is relatively segregated in a large city with 30–40 per cent Blacks. We might expect neighbourhoods in large diverse cities to have more diverse populations for this reason alone. We can account for this in two ways. First, we use a relative diversity index; where the neighbourhood's diversity is divided by the metro area diversity computed in the same manner from 1990 census data for the entire metro area. This is represented by equation (4)

$$RD_{Neigh} = \beta_0 + \beta_1 MetSize + \varepsilon_2 \tag{4}$$

where, $RD_{Neigh} = D_{Neigh}/D_{metro}$ and $D_{metro}$ is the metropolitan-level diversity measure. If large cities are really more segregated, $\beta_1 < 0$ is the expectation, while the contention of this paper is that $\beta_1$ should be positive or insignificant.

However, because large areas tend to be more diverse[14] and metro diversity is in the denominator of the left hand side variable in equation (4), $\beta_1$ will be biased downwards. To check for this effect, we control directly for the metropolitan-wide diversity in a third specification, equation (5)

$$D_{Neigh} = \gamma_0 + \gamma_1 MetSize + \gamma_2 D_{Metro} + \varepsilon_3 \tag{5}$$

If the standard view of large cities being more segregated is true, $\gamma_1 < 0$; however, this paper contends that $\gamma_1$ will be either insignificant or positive. Most theories would suggest that $\gamma_2 > 0$, as neighbourhoods in more diverse cities should be more diverse. We present in Table 4 the results for racial diversity and in Table 5 the results for economic diversity.

The results are consistent with expectations. In specification 3 of Table 4, larger metropolitan areas have more diverse neighbourhoods ($\alpha_1 > 0$). Specification 4 shows that controlling for metropolitan-level diversity by using neighbourhood *relative* diversity produces results suggesting that larger areas are more segregated (have more homogeneous neighbourhoods, $\beta_1 < 0$). Specification 5 controls for metropolitan-level diversity directly and the coefficient on metropolitan population becomes indistinguishable from zero ($\gamma_1 = 0$).

**Table 4.** Racial diversity and metropolitan population

| Dependent variable | Specification 3 $EN_{Neigh}$ | Specification 4 $EN_{Neigh}/EN_{Metro}$ | Specification 5 $EN_{Neigh}$ |
|---|---|---|---|
| ln(MetroPop) | 0.0505 | −0.1627 | 0.0011 |
| *T*-value | (3.99) | −4.83 | 0.07 |
| $EN_{Metro}$ | | | 0.2617 |
| *T*-value | | | 5.98 |
| Constant | −0.3480 | 3.1463 | 0.1947 |
| *T*-value | −1.83 | 5.92 | 0.90 |
| $R^2$ | 0.0228 | 0.0447 | 0.0766 |

Note: $N = 638$, in 88 metropolitan areas. Test statistics are in parentheses.

Downloaded from usj.sagepub.com at Glasgow University Library on October 17, 2014

**Table 5.** Economic heterogeneity and metropolitan size

| Dependent variable | Specification 3 $SD_{Neigh}$ | Specification 4 $SD_{Neigh}/SD_{metro}$ | Specification 5 $SD_{Neigh}$ |
|---|---|---|---|
| ln(MetroHHs) | 1879.481 | −0.01122 | −315.796 |
| $T$-value | (4.27) | −0.92 | −0.35 |
| $SD_{Metro}$ | | | 0.641245 |
| $T$-value | | | 2.83 |
| Constant | −3614.65 | 0.80433 | 4573.57 |
| $T$-value | −0.58 | 4.63 | 0.67 |
| $R^2$ | 0.0281 | 0.0013 | 0.0403 |

*Note*: $N = 633$, in 88 metropolitan areas. Test statistics are in parentheses.

The results for economic diversity (Table 5) are even more suggestive than those for race. In every case, the log of the number of households in the metropolitan area is either associated with greater neighbourhood heterogeneity or very far from significance. Also of note is the incredible lack of predictive power of this simple model. Economic segregation is explained very little by city size or by city size and city diversity. It seems that city scale does not have much of an effect on neighbourhood-level residential patterns. These low $R^2$ figures should be compared with those computed for the correlations and regressions reported in Tables 1–3 with census aggregates proxying for neighbourhoods, which were generally much higher. The explanatory power of metropolitan scale on segregation at the micro level is quite limited.

## 5. Conclusion

This paper has advanced the hypothesis that the correlation between metropolitan scale and measured segregation is spurious. The correlation is due to the fact that our geographical data are forced to proxy for neighbourhoods, but in large cities, higher density means smaller census tracts while larger population makes for larger neighbourhoods. The data analysed here support this contention: correlation between city size and measured segregation decreases as spatial data are collected with smaller areas; the number of spatial units in a metropolitan area is a stronger predictor of measured segregation than metropolitan scale; and, when data are gathered at the micro-neighbourhood level, there is no significant difference between large and small cities in terms of the homogeneity of their neighbourhoods. While there is always room for further analysis, the evidence thus far suggests that spatial aggregation effects bias estimates of the relationship between segregation and city size.

It is important to stress that up to this point, we have not done any modelling. The regressions do not represent or bear causal thinking: they are simply linear, unbiased estimates of conditional averages. The implications, however, are important to students of urban areas from economists to the public at large. First, this apparent empirical regularity appears not to be one that needs to be explained theoretically, as it is mostly a mirage. It would seem that there is nothing fundamentally different about big cities or big-city residents and neighbourhoods. Secondly, the focus of segregation-related policies has been predominantly in large urban areas. These results suggest that segregation may be just as much a problem in smaller areas that do not receive as much attention from academics or policy-makers. Finally, any study attempting to put together and empirically test an actual causal explanation of segregation patterns across urban areas or studying the effects of city-wide segregation on other outcomes should interpret its results carefully: measured segregation is not a benign proxy for true segregation because it is spuriously correlated with urban size.

Downloaded from usj.sagepub.com at Glasgow University Library on October 17, 2014

## Notes

1. Wong (2003) shows that this correlation is not particularly strong in his collection of 30 large metropolitan areas. However, it bears emphasising that if the hypothesis in this paper—that large cities are not as segregated as they look—is true, then a *lack* of correlation between city size and segregation would actually be masking the regularity that larger cities are *less* segregated.
2. A tradition in the sociology literature starting with Wirth (1938) and running through Fischer (1975, 1995) predicts that larger cities will be more heterogeneous. In the presence of increasing returns to neighbourhood formation as in Bayer *et al.* (2005), this would also lead to more segregation as long as city size increases faster than the number of groups. Thus, an explanation for a true relationship between segregation and urban scale is available. However, this paper should not be read as a test of any of these theories, as their implications are much broader than the relationship highlighted in this paper.
3. Bailey (1966), Yinger (1976), Kern (1981) and Schelling (1969) are early examples.
4. Wilson (1987), Ihlanfeldt and Sjoquist (1990), Duncan *et al.* (2004) and Sanbonmatsu *et al.* (2004) are a few.
5. Bayer *et al.* (2005) offer a promising start in this direction.
6. See Massey and Denton (1988) for a very helpful survey of this literature and assessment of the various measures.
7. See Fotheringham and Wong (1991) for a more detailed discussion of the modifiable unit problem. See Massey and Denton (1988) and Wong (2003) for some discussion of the virtues and shortcomings of the dissimilarity index.
8. This is termed the 'scale' effect of the MAUP. See Wong (1997), Wong *et al.* (1999) and Wong (2004) for several fine discussions and analyses of this aspect of the MAUP.
9. Multiracial individuals not withstanding.
10. The range of populations is from 57 813 (Enid, OK) to 19.4 million in NYC. For New English MSAs, these populations include only whole county partitions of MSAs. Thus, several smaller MSAs that did not contain entire counties were dropped from the analysis.
11. Thanks to an anonymous referee for pointing out this possibility.
12. This could potentially be explained by the fact that fast-growing areas will also experience large amounts of new housing construction. If new housing construction tends to be on the outskirts of cities, in standardised developments to a greater degree than the existing housing stock, large population growth could increase income stratification because a larger proportion of the area's housing stock is comprised of new, economically homogeneous sub-divisions. At the same time, the influx of new residents might increase racial mixing if new residents care less about preservation of old ethnic neighbourhood character, or gentrify inner-city neighbourhoods. These results and this explanation must, of course, be taken as provisional and speculative at this early stage, however.
13. The categories are White, Black, Indian, Asian and other, crossed with an indicator of 'Spanish' ethnicity. The entropy index is defined as $EN_{Neigh} = \sum p_r \ln (p_r^{-1})$ where, $r$ indexes racial categories and $p_r$ is the proportion of the neighbourhood population in each racial category.

    The index is maximised at ln(10) when every race is equally represented and approaches zero as any one $p_r$ approaches one. Because there are problems with calculating $EN_{Neigh}$ when some $p_r = 0$, we add $10^{-8}$ to all proportions so that $EN_{Neigh}$ is defined for all neighbourhoods.
14. Larger cities are more diverse: they have larger income variance and higher entropy indices. The correlation between metropolitan population and metropolitan-level heterogeneity is over 0.5 for both racial and economic measures.

## References


BAILEY, M. J. (1966) Effects of race and other demographic factors on the values of single-family homes, *Land Economics*, 42(2), pp. 215–220.

BAYER, P., FANG, H. and MCMILLAN, R. (2005) *Separate when equal? Racial inequality and residential segregation*. Working Paper No. 11507, NBER (http://www/nber.org/papers/w11507).

DUNCAN, G. J., CLARK-KAUFFMAN, E. and SNELL, E. (2004) *Residential mobility interventions as treatments for the sequelae of neighborhood violence*. Working Paper, Northwestern University Institute for Policy Research (http://www.northwestern.edu/ipr/publications/papers/2004/duncan/neighviolence.pdf).

FARLEY, R. (1991) Residential segregation of social and economic groups among Blacks, 1970–1980, in: C. JENCKS and P. E. PETERSON (Eds) *The Urban Underclass*, pp. 274–298. Washington, DC: Brookings Institution.


Downloaded from usj.sagepub.com at Glasgow University Library on October 17, 2014

FISCHER, C. S. (1975) Toward a subcultural theory of urbanism, *The American Journal of Sociology*, 80(6), pp. 1319–1341.

FISCHER, C. S. (1995) The subcultural theory of urbanism: a twentieth year assessment, *The American Journal of Sociology*, 101(3), pp. 543–577.

FOTHERINGHAM, A. S. and WONG, D. W.-S. (1991) The modifiable areal unit problem in multivariate statistical analysis, *Environment and Planning A*, 23(7), pp. 1025–1044.

IHLANFELDT, K. and SJOQUIST, D. (1990) Job accessibility and racial difference in youth employment rates, *American Economic Review*, 80(1), pp. 267–276.

JARGOWSKI, P. A. (1996) Take the money and run: economic segregation in US metropolitan areas, *American Sociological Review*, 61(6), pp. 984–998.

KERN, C. R. (1981) Racial prejudice and residential segregation: the Yinger model revisited, *Journal of Urban Economics*, 10(1), pp. 164–172.

MASSEY, D. S. and DENTON, N. A. (1988) The dimensions of residential segregation, *Social Forces*, 67(2), pp. 281–315.

SANBONMATSU, L., KLING, J. R., DUNCAN, G. J. and BROOKS-GUNN, J. (2004) *Neighborhoods and academic achievement: results from the Moving To Opportunity experiment*. Working Paper No. 492, Industrial Relations Section, Princeton University (http://www.irs.princeton.edu/pubs/working_papers.html).

SCHELLING, T. C. (1969) Models of segregation, *The American Economic Review*, 59(2), pp. 488–493.

WILSON, W. J. (1987) *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. Chicago, IL: University of Chicago Press.

WIRTH, L. (1938) Urbanism as a way of life, *The American Journal of Sociology*, 44(1), pp. 1–24.

WONG, D. W. S. (1997) Spatial dependency of segregation indices, *Le Géographe Canadien*, 41(2), pp. 128–136.

WONG, D. W. S. (2003) A comparison of traditional and spatial measures of segregation: some empirical findings, in: J. W. FRAZIER and F. W. MARGAI (Eds) M*ulticultural Geographies: The Changing Racial/Ethnic Patterns of the United States*, pp. 247–262. Binghamton, NY: Global Academic.

WONG, D. W. S. (2004) Comparing traditional and spatial segregation measures: a spatial perspective, *Urban Geographhy*, 25(1), pp. 66–82.

WONG, D. W. S., LASUS, H. and FALK, R. F. (1999) Exploring the variability of segregation index D with scale and zonal systems: an analysis of thirty US cities, *Environment and Planning A*, 31(3), pp. 507–522.

YINGER, J. (1976) Racial prejudice and racial residential segregation in an urban model, *Journal of Urban Economics*, 3(2), pp. 383–396.

Downloaded from usj.sagepub.com at Glasgow University Library on October 17, 2014