UNIVERSITY OF WAIKATO

Hamilton
New Zealand

Group-Size Bias in the Measurement of Residential Sorting

Mohana Mondal, Michael P. Cameron and Jacques Poot

Working Paper in Economics 09/18

August 2018

Corresponding Author
Mohana Mondal
School of Accounting, Finance and Economics
University of Waikato
Private Bag 3105, Hamilton 3240, New Zealand
Email: mm399@students.waikato.ac.nz

Michael P. Cameron
School of Accounting, Finance and Economics
and
National Institute of Demographic and Economic Analysis (NIDEA)
University of Waikato
Email: michael.cameron@waikato.ac.nz

Jacques Poot
National Institute of Demographic and Economic Analysis (NIDEA)
University of Waikato
and
Vrije Universiteit Amsterdam
Email: jacques.poot@waikato.ac.nz

Abstract

In this paper, we revisit a common issue with popular indices used for measuring residential sorting, that is, the extent to which a sub-group of the population is spatially distributed (sorted or segregated) differently from the remainder of the population. Specifically, we show that three common measures of residential sorting (namely, the *Index of Segregation*, the *Index of Isolation* and the *Entropy Index of Segregation*) are affected by group size, that is, the expected values of the indices are positive rather than zero under random sorting, and the size of this positive bias is related to group size. This is an important issue because it is common to compare sorting indices across groups of rather different sizes, both cross-sectionally and over time. Using New Zealand data, we demonstrate group-size impact on bias in measures of residential sorting by means of four methods: (1) plotting the relationship between group size and each residential sorting measure; (2) randomly allocating individuals across the area units, calculating the resulting residential sorting measures, and regressing these on group size; (3) showing that normalised/systematic indices of sorting are also related to group size; and (4) calculating the measurement bias for each index and plotting them against group size. Our empirical illustration uses microdata on the self-reported ethnicity of individuals (with multiple responses possible) from the New Zealand Census of Population and Dwellings (1991-2013) for the Auckland region, selected due to its high ethnic diversity. Our results demonstrate that the *Entropy Index of Systematic Segregation* measure of residential sorting is the measure that is the least affected by group size variation. As a result, we strongly recommend using this index of sorting as a preferred measure.

Keywords

residential sorting
segregation
group-size bias
entropy index

JEL Codes

C18, J19, Z13, R23

Acknowledgements

Access to data used in this study was provided by Statistics New Zealand under conditions of security and confidentiality provisions of the Statistics Act 1975. The first author acknowledges the support provided by the University of Waikato in the form of the generous University of Waikato Doctoral Scholarship. The first author also acknowledges the generous funding received from CaDDANZ (Capturing the Diversity Dividend of Aotearoa/New Zealand). We thank Arthur Grimes, participants at the NZAE (New Zealand Association of Economists) Conference, held in Auckland in June 2018, for their detailed comments and suggestions.

1. Introduction

Residential segregation or sorting[1] among ethnic groups has been a popular area of study since Duncan and Duncan's (1955) seminal contribution. By 2019, Google Scholar identified more than 2500 articles with 'residential segregation' in the title (and many more that cover residential segregation or sorting but where it is not explicit in the title). There has been a lot of debate about the correct index to use in measuring residential sorting (White 1983, Massey and Denton 1988, Carrington and Troske 199, Reardon and Firebaugh 2002, Fossett 2017), and extant studies mostly use the *Index of Dissimilarity* and/or the *Index of Isolation*. In this paper, we contribute to the methodological debate on the choice of a preferred index.

Specifically, we investigate a particular source of bias in many common measures and indices, arising from their sensitivity to the size of the groups for which the measures or indices are being calculated. Such bias arises when the expected value of the index is not zero but is strictly positive, even in the case in which the group of interest is randomly allocated across areas such that the expected value of its population share in every area is equal to its share of the total population.[2] Hence the aim of this paper is to show the sensitivity of popular measures of residential sorting to group size. We find that the *Entropy Index of Systematic Segregation* measure is least affected by this group-size bias and hence we recommend it as a preferred measure of sorting, even though this index has to date been far less commonly used than other sorting measures.

The bias is also a function of the granularity of the data. The smaller the spatial units, and therefore their expected population size, the greater the bias under random sorting. However, granularity is not addressed in the present paper.

We illustrate our results on group-size related bias by means of microdata on self-reported ethnicity of individuals (with multiple responses possible) from the New Zealand Census of Population and Dwellings (1991-2013) for the Auckland region, selected due to its high ethnic diversity. Hence, throughout this paper, by 'region' we mean the Auckland region, as defined by Statistics New Zealand, which is made up of about 400 area units that roughly represent suburbs or wards. Hence, the term 'area' refers to area units in the Auckland region. We refer to an individual's ethnic group as 'group'. The number of individuals belonging to a specific ethnic group is referred to as the 'group size'. The New Zealand census allows for multiple responses to the ethnicity question and, hence, individuals can belong to more than one group. The counts used in the paper refer to total responses, not total individuals. An ethnic

---

1 We use 'residential sorting' as a term that encompasses a range of measures of residential segregation that include dissimilarity, isolation, and concentration (for example, Massey and Denton, 1988). Our preferred term is not only broader, but carries none of the negative connotations associated with use of the word 'segregation'.

2 A randomised allocation is obtained when the number of persons of the group allocated to an area is given by a draw from a binomial distribution $B(n, p)$ with n equal to the area's population and p the fraction of the group in the total population.

group proportion in an area unit is the number of people residing in that area unit who are reporting that ethnicity divided by the aggregate count of all reported ethnicities in that area unit.

The remainder of the paper proceeds as follows. In Section 2, we briefly discuss some relevant studies on popular measures of residential sorting. Section 3 describes the data and Section 4 details the methods. Section 5 presents and discusses the results and Section 6 concludes.

2. Literature Review

Residential sorting is defined as the degree to which groups live away from each other (Denton and Massey 1988 Johnston *et al*. 2007). There have been thousands of studies of residential sorting, including several in the New Zealand context (for example, Johnston *et al.* 2002, 2005 and 2011, Maré *et al.* 2012). These studies mostly resort to one of several 'traditional' measures of residential sorting, of which the most common are the *Index of Dissimilarity*, the *Index of Segregation*, and the *Index of Isolation*.

Denton and Massey (1988) summarised the literature on residential sorting to that point in time, and concluded that residential sorting is a multidimensional concept that captures five distinct dimensions of spatial variation: (1) evenness; (2) exposure; (3) concentration; (4) centralisation; and (5) clustering. Each dimension brings out different features of the spatial distribution of social groups. While measures of evenness calculate the differential distribution of the subject population, measures of exposure reveal the extent of potential contact with other groups. Concentration refers to the relative physical space occupied by a group, whereas centralization indicates the extent to which a group is located near the centre of an urban area. Finally, the degree to which minority group members live disproportionately in adjacent areas is defined as clustering. Massey and Denton (1988) point out that these five dimensions overlap empirically (a group that is residentially sorted on one dimension will often also show some evidence of sorting on one or more of the other dimensions). However, the dimensions are conceptually distinct and have led to a considerable number of measures that each aim to quantify a specific dimension. For example, formulae for 17 segregation indices defined in Massey and Denton (1988) can be found in Iceland *et al.* (2002).

James and Taeuber (1985) presented a set of criteria for evaluating measures of sorting, being the principles of organisational equivalence, size invariance, transfers, and exchanges. By organisational equivalence, they mean that when a unit is subdivided, with the same group proportions as in the original unit, then the sorting measure should remain unchanged. A measure is size invariant if its value is unchanged when the number of persons in each group in each area is multiplied by a constant factor. According to the principle of transfers, if an individual is relocated from one unit to another unit, where the proportion of persons in the group is greater in the former unit, then sorting will decrease. The principle of exchanges states that if an individual in group g in area a is exchanged with an individual in a different group in a different area, the proportion of persons in the respective groups being greater in their original area units, then sorting will decrease.

The most important and well-known dimension of residential sorting is evenness (Johnston *et al.* 2002). The *Index of Dissimilarity* (Duncan and Duncan 1955) is a measure of evenness that reflects the proportion of people in a population subgroup that would have to relocate in order to make their distribution identical to that of the reference group. When the same index is computed between one group and all other groups combined, the index is sometimes referred to as the *Index of Segregation* (Maré *et al.* 2011), although in the literature the term 'segregation index' can also be the generic term that refers to any of the sorting measures. The Index of Dissimilarity and the Index of Segregation range between 0 (the two groups are identically distributed spatially) and 1 (in any area only one group or the other is represented but never both). A high value represents a high level of residential sorting - most of the group members live in an area where other groups are relatively absent (Duncan and Duncan 1955). In contrast, the *Index of Isolation* is a measure of exposure, and is used to measure the degree to which individuals locate with other members of their own group (Duncan and Duncan 1955).

Many studies have noted the weaknesses of using such measures of residential sorting, as they are sensitive to many factors (Duncan and Duncan 1955, White 1983, Carrington and Troske 1997, Fossett 2017). For example, the traditional measures of residential sorting described above are only global measures, because they summarise residential sorting for the entire region under study (Wong 2002). Hence, they do not capture differences in sorting between parts of the overall region.

White (1983) identified faults in using the *Index of Dissimilarity* to measure residential sorting. He stated that the values of this measure are sensitive to the group sizes, as well as to the size and number of the areal units. He added that all measures of residential sorting that are related to the *Index of Dissimilarity* have the same disadvantages. Moreover, the *Index of Dissimilarity* does not obey the principles of transfers and exchanges (White 1986, Reardon and Firebaugh 2002). Voas and Williamson (2000) note that even when there is random distribution, the *Index of Dissimilarity* can give highly misleading results when the area population is small or the group proportion is low. They add that the value of the index is also difficult to interpret when there are more areal units under consideration than minority individuals (the minimum value of the *Index of Dissimilarity* then rises very rapidly with the number of area units). Moreover, the *Index of Dissimilarity* does not capture changes in the level of residential sorting when population groups in different areal units are swapped (Wong 2002), demonstrating that it fails to obey the exchange principle.

Carrington and Troske (1997) note that the *Index of Segregation* and the *Index of Isolation* can suggest the presence of substantial residential sorting, even when there is an absence of residential sorting behaviour, in the case of there being many small spatial units and for groups that form a small proportion of the overall population. This can be easily demonstrated by simulating random sorting, as Maré *et al.* (2012) show in the appendix to their paper. The *Index of Isolation* is sensitive to group size as well as group settlement patterns, being generally low for small groups and rising with increases in group size, even though the group's level of sorting may actually remain the same.

In the New Zealand context, Johnston *et al.* (2011) also note that the *Index of Dissimilarity*, and hence the *Index of Segregation* as well, can give misleading results when there are small groups. They argue that the best approach to measuring residential sorting is therefore to report multiple indices. In their study, they calculate the *Index of Segregation* and the *Index of Isolation* for 25 ethnic groups in Auckland, using 1996 New Zealand Census data. They show that the smallest groups are the most segregated according to the *Index of Segregation* values, and that there is a close relationship between a group's size and the *Index of Isolation* values. Maré *et al.* (2012) show that, when they randomly allocate group members across spatial units, the *Index of Segregation*, *Gini coefficient* and the *Maurel and Sédillot Index of Concentration* all suggest the presence of substantial residential sorting even when there is none. However, despite the inappropriateness of the traditional measures, they continue to be used because of the simplicity of their calculation, their ease of interpretation, and their comparability with past studies.

The *Entropy Index of Segregation* (also called the *Information Theory Index*) was originally proposed by Theil (1972) as another measure of evenness, that is, this measure also suggests the degree to which groups are unevenly distributed among area units (Denton and Massey 1988). The *Entropy Index of Segregation* measures the average difference between an area unit's group proportion and the group proportion in the city or region as a whole (Theil 1972).

Reardon and Firebaugh (2002) evaluated a set of six multi-group segregation indices following the principles introduced by James and Taeuber (1985) that we outlined earlier. They found that the *Entropy Index of Segregation* is the only multi-group measure of residential sorting that obeys the principles of organisational equivalence, size invariance, transfers and exchanges. Moreover, this measure has the added advantage that it can be decomposed into a sum of between-group and within-group components (Theil 1972, Nijkamp *et al.* 2015). Despite having many favourable properties, until now relatively few studies have used the *Entropy Index of Segregation* as a measure of residential sorting. Most of those studies are based on U.S. data (Wright *et al.* 2014, Parry and Eeden 2015, Fowler *et al.* 2016, Lichter *et al.* 2017).

Though previous studies have identified the presence of group-size bias in the traditional measures of residential sorting, there has been to date relatively little systematic analysis of this. Group-size bias is an important issue, because the interpretation and comparison of groups and areas in terms of residential sorting is affected by the choice of the number (and hence size) of groups included within the calculation of the indices. Thus, in this paper we compare the traditional measures of residential sorting and the *Entropy Index of Segregation*, in terms of their sensitivity to group size. Specifically, we demonstrate in four different ways the group-size bias of each measure and show that the *Entropy Index of Systematic Segregation* (which has expected value zero under random sorting) is the least affected by this bias.

3. Data

Auckland is the most ethnically diverse region in New Zealand. According to the 2013 Census,[3] its ethnic composition consisted of European (59.3 percent), Asian (23.1 percent), Pacific Islander (14.6 percent) and Māori (10.7 percent) ethnicity (Statistics New Zealand 2013).[4] Auckland is also the most populous of the 16 regions in New Zealand. It alone accounts for about one third of the New Zealand population of close to five million. Auckland can be considered a very good example of a modern EthniCity (Johnston *et al.* 2002) or superdiverse city (Spoonley 2014; Vertovec 2019). It is therefore a suitable focus for our empirical analysis. We obtained population data from the 1991, 1996, 2001, 2006, and 2013 New Zealand Census of Population and Dwellings for the Auckland region of New Zealand.

The New Zealand Census of Population and Dwellings collects a range of socio-demographic information on each member of the New Zealand population present in New Zealand on census night. The Census provides information about each usually-resident individual such as location, age, sex, ethnicity, income level, occupation, education and marital status which can be aggregated to population statistics at the area unit level.[5] The Auckland region is made up of 413 land-based area units,[6] of which 409 had a non-zero usually resident population in 2013. Area units with no usually resident population were excluded from the analysis. Unit record data were accessed within Statistics New Zealand's secure data laboratory to meet the confidentiality and security rules according to the Statistics Act 1975. In accordance with the strict confidentiality rules laid down by Statistics New Zealand, the summary statistics,

---

[3] The most recent population census was held on March 6, 2018. At the time of collecting the data for this paper, the results of that census were not yet available. In any case, due to non-response issues, 2018 census data are of somewhat lesser quality than previous censuses with respect to variables such as ethnicity. Additionally, caution is needed in comparing results of the 2018 census with those of previous censuses. See 2018 Census External Data Quality Panel (2020) *Final report of the 2018 Census External Data Quality Panel*. Retrieved from www.stats.govt.nz.

[4] The sum of these percentages exceeds 100 percent, as people can report more than one ethnicity.

[5] A meshblock is the smallest geographic unit for which Statistics New Zealand collects statistical data. Meshblocks vary in size from part of a city block to large areas of rural land. The country is divided into about 50,000 meshblocks that are aggregated to about 2000 area units. Our analysis is based on data aggregated to the area unit level. Area units are non–administrative areas that are in between meshblocks and territorial authorities in size (Statistics New Zealand 2013). In urban areas, area units are approximately the size of individual suburbs, and in our dataset they have an average population of 1530.

[6] In this paper, we use 2013 area unit boundaries.

counts and calculations are based on data that have been suppressed for raw counts less than six, and otherwise randomly rounded to base three.[7]

Self-reported ethnic identification is collected in the Census, and each person can choose a single or multiple-ethnic response. An individual reporting more than one ethnicity is included in each ethnic group that they report (this is referred to as 'total count' ethnicity) (Statistics New Zealand 2015). According to the New Zealand Standard Classification of Ethnicity, ethnicity is classified in a hierarchy of four levels (Statistics New Zealand 2013). The main (Level 1) ethnic groups defined in the 2006 and 2013 Census by Statistics New Zealand are: New Zealand European; Māori; Pacific peoples; Asian; Middle East, Latin American and African (MELAA); and Others. Previous research on ethnicities in New Zealand, such as Maré *et al.* (2012), have only investigated ethnic residential sorting using Level 1 ethnic groups. As there is considerable diversity in the characteristics and choices within most of these broad ethnic groups, we use data on Level 2 ethnic groups (total responses) instead. The Level 1 and Level 2 classifications along with the number of total responses for each ethnic group in New Zealand are shown in Table 1.[8]

The format of the question about ethnicity in the Census of Population and Dwellings was inconsistent between the Censuses from 1991 to 2001. The format in 2001 was similar to that of 1991, but both differed to that of 1996.[9] Thus, comparability across Censuses is likely to be affected. Consequently, there were some significant changes in the responses in 1996 compared to 1991 or 2001 that were likely to have been caused by the change in the wording of the question. These included increased multiple response in 1996, a consequent reduction in single responses, and a tendency for respondents to answer the 1996 question on the basis of ancestry (or descent) rather than ethnicity (or cultural affiliation). For example, van der Pas and Poot (2011) noted that almost 48,000 people identified themselves as Dutch in the 1996 Census but at the time of the 2001 and 2006 census there were only close to 29,000 people in New Zealand who identified themselves as Dutch. According to van der Pas and Poot (2011),

---

7 Counts that are already a multiple of three are left unchanged. Those not a multiple of three are rounded to one of the two nearest multiples. For example, a one will be rounded to either a zero or a three. Each value in a table is rounded independently.

8 The sum of Level 2 total responses in Table 1 is greater than the sum of Level 1 total responses because some individuals reported multiple ethnicities at level 2 for which some or all belonged to the same ethnic group at level 1.

9 The ethnicity question in the 1996 Census had a different format from that used in 1991 and 2001. In 1996, there was an answer box for 'Other European' with additional drop-down answer boxes for 'English', 'Dutch', 'Australian', 'Scottish', 'Irish', and 'other'. These were not used in 1991 or 2001. Furthermore, the first two answer boxes for the question were in a different order in 1996 from 1991 and 2001. 'NZ Māori' was listed first and 'NZ European or Pakeha' was listed second in 1996. The 1991 and 2001 questions also only used the words 'New Zealand European' rather than 'NZ European or Pākehā' (Pākehā is the Māori word referring to a person of European descent). The 2001 question used the word 'Māori' rather than 'NZ Māori'. The format of the 2006 and 2013 questionnaire was the same as that of 2001 (Statistics New Zealand 2017).

this huge difference between the 1996 and the subsequent two Censuses was the result of the 1996 Census question on ethnicity including Dutch as a specific option. The resulting inconsistencies mainly appear for the 'European' ethnic groups (including 'New Zealand European') and the 'Māori' ethnic group. In the 1996 data, the counts for 'Other Europeans' were much higher and the counts for the 'New Zealand European' category were much lower than in the 1991 or 2001 data. This can be attributed to the fact that, in 1996, people saw the additional 'other European' category as being more suitable to describe their ethnicity than the 'New Zealand European' category (Statistics New Zealand 2017).

Table 1: Level 1 and Level 2 Classification and Counts of Ethnic Groups in New Zealand 2013

Ethnic group code (Level 1)	Ethnic group code description (Level 1)	Total responses	Ethnic group code (Level 2)	Ethnic group code description (Level 2)	Total responses
1	European	2,969,391	10	European nfd	26,472
			11	NZ European	2,727,009
			12	Other European	268,044
2	Māori	598,605	21	NZ Māori	598,605
3	Pacific Peoples	295,941	30	Pacific Island nfd	1,026
			31	Samoan	144,138
			32	Cook Island Māori	61,077
			33	Tongan	60,333
			34	Niuean	23,883
			35	Tokelauan	7,173
			36	Fijian	14,445
			37	Other Pacific Island	11,925
4	Asian	471,708	40	Asian nfd	4,623
			41	Southeast Asian	77,430
			42	Chinese	164,949
			43	Indian	154,449
			44	Other Asian	82,242
5	MELAA	46,953	51	Middle Eastern	20,406
			52	Latin American/Hispanic	13,182
			53	African	13,464
6	Other	67,752	61	Other ethnicity	67,752

Notes: Total responses all ethnic groups 4,450,350 and 4,542,633; nfd = not further defined.
Source: Statistics New Zealand (2013)

In addition, many people choose 'New Zealander' as their ethnicity in the Census. This term was introduced in the 2001 census. Its assignment in the classification has changed over time. In 2001, 'New Zealander' was counted in the New Zealand European category. But from 2006 onwards, New Zealander has instead been included as a new category, as part of the 'Other' ethnicities. The increase in counts for the New Zealand European category from 2006 to 2013 is attributed partly due to fewer people identifying themselves as 'New Zealander' in 2013.

The changing ethnic classifications can have an impact on the comparison of sorting measures across groups and over time. However, they should have little effect on our analysis of group-size effects. In any case, we will control for differences between censuses by means of time-fixed effects in our regression models.

4. Methodology

As stated in the introduction, the aim of this paper is to show the sensitivity of popular measures of residential sorting to group size. We achieve this aim using four techniques. First, we calculate the values of the *Index of Segregation*, *Index of Isolation* and the *Entropy Index of Segregation* using the formulas outlined in Table 2, applied to Census data for the Auckland region of New Zealand. High values of these indices represent more residential sorting. The values of these indices vary between 0 (when all areas have the same composition) and 1 (complete sorting). Each measure of residential sorting is calculated based on data aggregated to the area unit level. We calculate the values for all the Level 2 ethnic groups in Auckland for all census years from 1991-2013. We proportionally distributed the population counts of the 'not further defined' category for each Level 2 ethnic group into the rest of the Level 2 groups sharing the same Level 1 ethnic group.[10] We then use scatter plots to display the relationship between group size and the value of each index.

Table 2: Summary Measures of Residential Sorting

Index of Segregation	$ISeg_g = \frac{1}{2}\sum_{a=1}^{A}\left\lvert\frac{P_{ga}}{P_{g.}} - \frac{(P_{.a} - P_{ga})}{(P_{..} - P_{g.})}\right\rvert$
Index of Isolation	$IIsol_g = \frac{\left(\left[\sum_{a=1}^{A}\pi_{ga}\frac{P_{ga}}{P_{.a}}\right] - \frac{P_{g.}}{P_{..}}\right)}{\left(1 - {P_{g.}}/{P_{..}}\right)}$
Entropy Index of Segregation	$EIS_g = \sum_{a=1}^{A}\frac{P_{.a}}{P_{..}}\left(1 - \frac{E_a}{\bar{E}}\right)$

$$\text{Where}: E_a = -\frac{P_{ga}}{P_{.a}}\ln\left(\frac{P_{ga}}{P_{.a}}\right) - \left(1 - \frac{P_{ga}}{P_{.a}}\right) ln\left(1 - \frac{P_{ga}}{P_{.a}}\right)$$
$$\bar{E} = -\frac{P_{g.}}{P_{..}} ln\left(\frac{P_{g.}}{P_{..}}\right) - \left(1 - \frac{P_{g.}}{P_{..}}\right) ln\left(1 - \frac{P_{g.}}{P_{..}}\right)$$

Notes

P_{ga} refers to the population of group g (=1, 2,…G) in area a (= 1,2,….A). A subscript dot refers to the sum over that specific subscript. $\pi_{ga} = \frac{P_{ga}}{P_{g.}}$, hence $\sum_{a=1}^{A}\pi_{ga} = 1$.

The calculation of *EIS* requires that we define 0*ln(1/0)= $\lim_{q\to 0}[\,q(\ln({1}/{q})] = 0$ to account for any cases in which group g is not represented in an area a. These summary measures of residential sorting are defined in Iceland *et al.* (2002).

[10] We also ran the analysis with not further defined as a separate category, as well as dropping them completely. The ranking of groups, the trends over time and our key conclusions are not affected.

Secondly, following Maré *et al* (2012) we simulate 100 random allocations of the population using a binomial distribution for each ethnic group. The simulated number of group members in an area unit is based on the total number of draws being equal to the actual area unit population, while the expected probability is taken to be equal to each group's share of the total Auckland population. We then calculate the values of the indices in each of these 100 independently simulated random allocations. We take the average of these index values as our estimate of the sorting that would be observed had the allocation across area units been random. In the absence of a group-size bias, the expected value of a measure of sorting equal to zero, when we calculate the indices based on the randomised data. In other words, in the case of randomly allocating people across areas (but taking into account areas populations), there should be ideally no relationship between group size and measures of residential sorting. We use scatter plots and simple linear regression to show that this is not the case for the conventional measures of residential sorting.

Hence, in the third part of our analysis we calculate a modified version of each of the standard segregation measures, following Carrington and Troske (1997). These authors refer to such a modified sorting measure as an index of systematic segregation, which has an expected value of zero under random sorting. When such an index yields a positive value, it measures the amount of excess sorting that would occur if allocation across area units is not random.[11] We calculate the systematic index values *IS* for the sorting index *I*, where *I* is the *Entropy Index of Segregation* or the *Index of Segregation* by means of the formula:

$$IS = \frac{(I - I_R)}{(1 - I_R)}$$

where *I* is the index value based on actual data and I_R is the average of the index values based on randomised data. Following Maré *et al.* (2012), we calculate the *Index of Systematic Isolation* using the formula:

$$ISIsol := \left(\left[\sum_{a=1}^{A} \boldsymbol{\pi}_{ga} \frac{P_{ga}}{P_{.a}}\right] - \left(\sum_{a=1}^{A} \boldsymbol{\pi}_{ga} \frac{P_{ga}}{P_{.a}}\right)_R\right) \Big/ \left(\mathbf{1} - \left(\sum_{a=1}^{A} \boldsymbol{\pi}_{ga} \frac{P_{ga}}{P_{.a}}\right)_R\right).$$

The subscript *R* refers to the average of values based on randomised allocations. We run a simple linear regression to identify the relationship between group size and the different measures of systematic residential sorting. Finally, we define the bias[12] for each index as *I* – *IS*, where *I* is an index value based on actual data and *IS* the value of the corresponding index of systematic sorting. We calculate the bias for each index and plot these against group size (on a logarithmic scale).

---

[11] Fossett (2017) has introduced an alternative way of generating sorting measures that will have an expected value of zero under random sorting.

[12] The difference between the actual index values and expected values of the indices under random sorting.

5. Results and Discussion

As stated in the introduction, the aim of this paper is to show that the selected measures of residential sorting are sensitive to (and hence, biased by) group size and propose the best index among these to measure residential sorting. We calculate the values of the measures of residential sorting, for each ethnic group in Auckland, using 1991-2013 census data (Appendix Table A1). We have multiplied the index values by 100 for easy interpretability.

Next, for each population subgroup, we simulate 100 random allocations using a binomial distribution.[13] We see that under random spatial allocation the values of the sorting indices are always less than the values based on actual data. We now plot these index values based on actual data as well as the average values of sorting indices under random allocation, pooled across all five Censuses, against group size, in Figure 1. We use a logarithmic scale for group size. The panels in Figure 1 show that in the case of residential sorting indices based on both actual data and randomised allocation, there is a relationship between each residential sorting measure and group size. Panel (a) shows the relationship between the *Index of Segregation* values and group size. The scatter plot clearly shows that larger groups have lower *Index of Segregation* values, i.e. large groups are less residentially sorted than small groups in Auckland.

Similarly, Panel (b) shows the relationship between the *Index of Isolation* and group size. The scatter plot shows that in the case where the index value is based on actual data, for larger groups, values of this measure are larger.[14] We observe that, under random sorting, the *Index of Isolation* values appear to be almost zero irrespective of group size. When using a different scale on the vertical axis (see Appendix Figure A1), it can be shown that there is very little effect of group size on the *Index of Isolation* for small and medium group sizes under random spatial allocation. In contrast, the index is somewhat less for the largest group sizes.

The relationship between the *Entropy Index of Segregation* and group size is shown in Panel (c). As in the case of the *Index of Segregation*, the *Entropy Index of Segregation* values also decrease with increases in group size. This is not surprising, because the *Index of Segregation* and the *Entropy Index of Segregation* values are in applications often highly

---

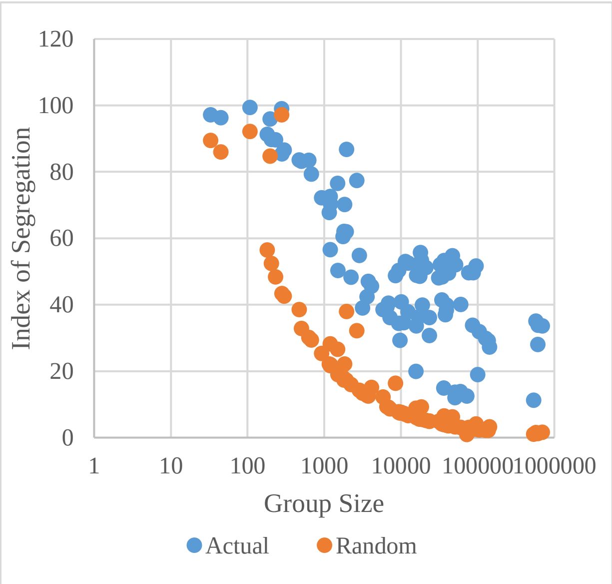
[13] Appendix Table A2 reports the average of index values obtained from the 100 simulations. We have multiplied the index values by 100 for easy interpretability.

[14] It can be easily shown by calculus that for a given spatial distribution of the group across areas, the *Index of Isolation* is non-decreasing in total group size. It should also be noted that the *Index of Segregation* is scale free in the total size in the group of interest for a given spatial distribution of this group. No simple mathematical result can be established in the case of the *Entropy index of Segregation*. This is because, even if E_a is scale-invariant for a given distribution of group g across areas, $\bar{E}$ and $\frac{P_{.a}}{P_{..}}$ depend on how relatively important the group g is in the population and in each area unit 'a' respectively. Empirically, however, the group size effect has been investigated previously by Fossett (2017) with US data.

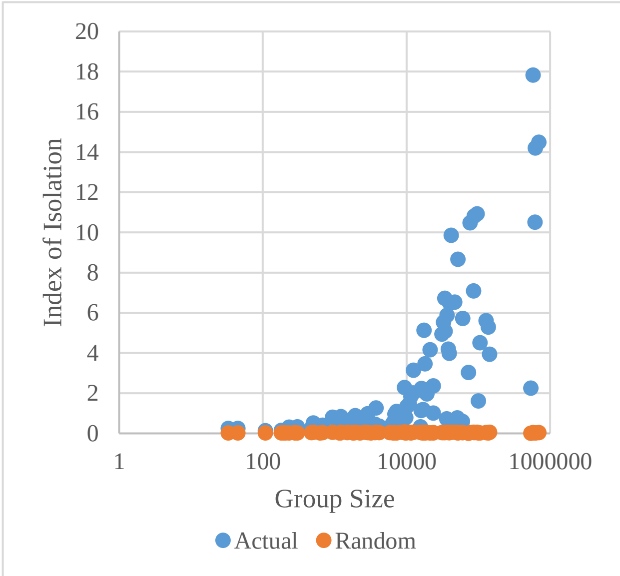
positively correlated. This can be seen in Table 3 for our Auckland data. With sorting observed for 18 groups in five census years, N = 90. The Pearson correlation coefficient between the *Index of Segregation* and the *Entropy Index of Segregation* is about 0.93. However, the *Index of Segregation* is weakly inversely correlated with the *Index of Isolation* (with a correlation coefficient of about -0.3), while there is no statistically significant correlation between the *Entropy Index of Segregation* and the *Index of Isolation*.

Figure 1: Scatterplots of Index Values and Group Sizes

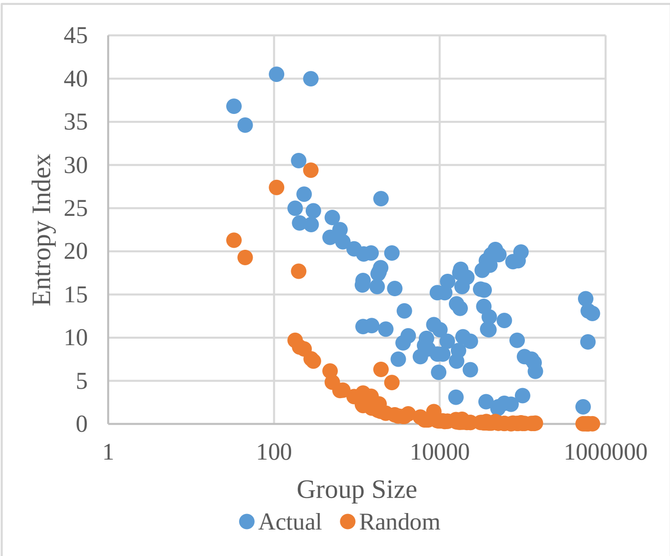
based on Randomised and Actual Data, Auckland Region, 1991-2013



1(a): *Index of Segregation* and Group Size



1(b): *Index of Isolation* and group size



1(c): *Entropy Index of Segregation* and Group Size

Table 3: Correlation between the Three Sorting Measures

	Index of Segregation	Index of Isolation	Entropy Index of Segregation
Index of Segregation	1.000		
Index of Isolation	-0.3027*** (0.0037)	1.000	
Entropy Index of Segregation	0.9306*** (0.000)	-0.0627 (0.5574)	1.000

Notes
N=90 (18 ethnic groups x 5 census years)
p-values in parentheses, * p<0.10, ** p<0.05, *** p<0.01

To check the statistical significance and size of the effect of group size in relation to the different index values, we ran a simple linear regression of each index value on group size (logarithmic scale), with census fixed effects added to the regression. The results are shown in Table 4.

Table 4: Statistical Significance of Group Sizes to the Different Indices

	Regression Results from Actual Data			Regression Results from Randomized Data		
	ISeg	IIsol	EIS	ISeg	IIsol	EIS
	(1)	(2)	(3)	(4)	(5)	(6)
Log Group Size	-8.466***	1.252***	-2.366***	-8.000***	0.00049	-1.676***
	(0.634)	(0.130)	(0.304)	(0.611)	(0.00043)	(0.201)
R^2	0.70	0.54	0.47	0.71	0.68	0.51

Notes
N=90 (18 ethnic groups x 5 census years).
The regressions include census fixed effects.
Standard errors in parentheses.
*p<0.10, **p<0.05, ***p<0.01
Iseg = Index of Segregation.
Iisol = Index of Isolation.
EIS = Entropy Index of Segregation.

From columns (1) to (3) of Table 4, we observe that group size is statistically significantly correlated with all the measures, at the 1% level of significance. However, in the case of the *Entropy Index of Segregation*, we see that the coefficient for group size (-2.37) is much smaller than for the *Index of Segregation* (-8.47), even though they are similar measures. We observe that the coefficient for group size for the *Index of Isolation* (1.25) is smaller than for the other two measures. However, we note that the *Index of Isolation* is not directly comparable to the *Entropy Index of Segregation*, as it measures a different aspect of the population distribution. The *Index of Isolation* for any group g measures the degree to which individuals of group g co-locate with other members of their own group, whereas the other index measure the extent to which group g is concentrated in particular areas.

When we check the statistical significance of group size (logarithmic scale) in relation to the different index values based on randomised data, we observe that group size is statistically significantly correlated with both *Entropy Index of Segregation* and *Index of Segregation* (Table 4, Columns (4) and (6)). However, we observe that the coefficient for group size is again much smaller for the *Entropy Index of Segregation* (-1.68) than for the *Index of Segregation* (-8.00) and thus, less biased by group size. We saw in Figure 1 that the *Index of Isolation* values after randomisation are almost zero and Table 4 shows that there is no statistically significant relationship of the isolation measure with group size.

Following Carrington and Troske (1997), we next calculate the *Index of Systematic Segregation* for each index (Appendix Table A3)[15] and then check the statistical significance of the relationship with group size (logarithmic scale) (Table 5). The results show that all three of the indices of systematic segregation are sensitive to group size, with the effect being statistically significant at the one percent level in all three cases. However, the coefficient of log group size in the regression for the *Index of Systematic Segregation* (-6.43) is much more negative than is the case for the *Entropy Index of Systematic Segregation* (-0.98). The *Index of Systematic Isolation* is positively related to log group size.

Table 5: Statistical Significance of Group Sizes to the Index of Systematic Segregation

	ISSeg (1)	ISIsol (2)	EISS (3)
Log Group Size	-6.432*** (0.648)	1.254*** (0.130)	-0.980*** (0.243)
R^2	0.57	0.54	0.20

Notes
N=90 (18 ethnic groups x 5 census years).
Standard errors in parentheses.
*p<0.10, **p<0.05, ***$p<0.01$
The regressions include census fixed effects.
ISSeg = Index of Systematic Segregation.
ISIsol = Index of Systematic Isolation
EISS = Entropy Index of Systematic Segregation.

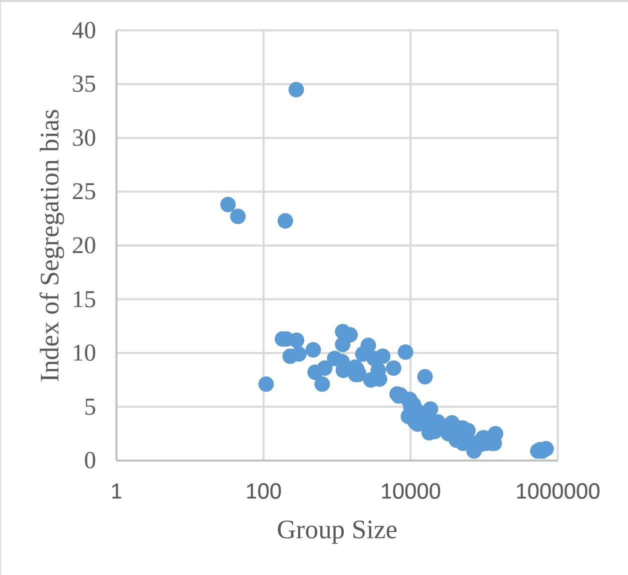
Comparing the values from Table 4, columns (1) to (3), with those of Table 5, we conclude that the *Entropy Index of Systematic Segregation* is the best measure, as the coefficient of group size for this measure (-0.98) is the smallest among all index values based on actual data.

Finally, we calculate the bias values for each of the three original and plot them against group sizes (on a logarithmic scale) in Figure 2. The bias decreases with increases in group size in the case of the *Index of Segregation* and the *Entropy Index of Segregation*. However, we note group size has a far less notable effect on the bias defined as the difference between

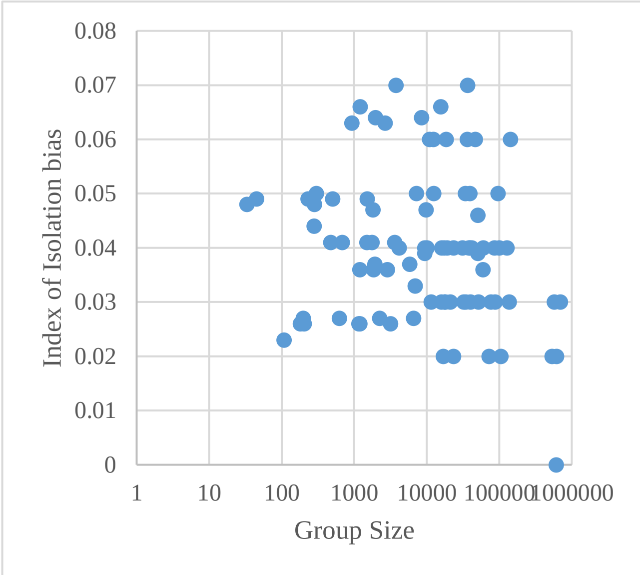
[15] We have multiplied the index values by 100 for easy interpretability.

the *Index of Isolation* and the Maré *et al.* (2012) modification of this original index. This is related to the fact that under random sorting the *Index of Isolation* values appear to be almost zero irrespective of group size (see Figure 1, Panel (b)).

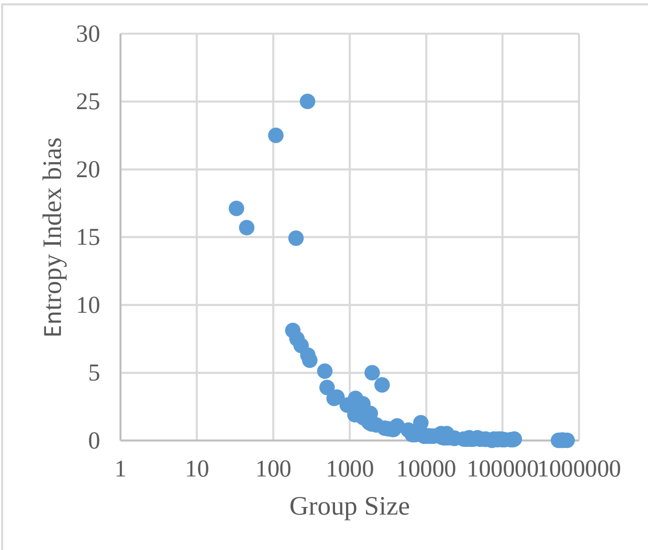
Figure 2: Scatter Plot of Index Bias and Group Size



2(a): Relationship between *Index of Segregation* Bias and Group Size



2(b): Relationship between *Index of Isolation* Bias and Group Size



2(c): Relationship between *Entropy Index of Segregation* Bias and Group Size

We run a simple linear regression, with census fixed effects, to see the relationship between the index bias and the group size (on a logarithmic scale), which is reported in Table 6. We find that group size is negatively related to the index bias values, with statistical significance at the one percent level in all three cases. Moreover, we observe that the coefficient for the *Entropy Index of Segregation* (-1.39) is smaller than the coefficient for the *Index of Segregation* (-2.03). Hence the *Entropy Index of Segregation* is the better measure among the two of evenness of spatial distribution.

Table 6: Statistical Significance of Group Size on Index Bias

Difference between Original Measures and Systematic Indices

	ISeg–ISSeg	IIsol–ISIsol	EIS–EISS
Log Group Size	-2.034*** (0.163)	-0.002*** (0.00033)	-1.387*** (0.167)
R^2	0.67	0.78	0.50

Notes

N=90 (18 ethnic groups x 5 census years).

Standard errors in parentheses.

*p<0.10, **p<0.05, ***p<0.01

The regressions include census fixed effects.

Iseg = Index of Segregation; ISSeg = Index of Systematic Segregation.

IIsol = Index of Isolation; ISIsol = Index of Systematic Isolation.

EISS= Entropy Index of Systematic Segregation

Overall, our results show that all sorting measures, including the *Entropy Index of Systematic Segregation*, are sensitive to group size. However, our results also show that the *Entropy Index of Systematic Segregation* appears to be the least affected by group size among the measures we considered.

6. Conclusions

The aim of this paper is to demonstrate the sensitivity of alternative measures of residential sorting to group size. The traditional measures included in our study are the *Index of Segregation* and the *Index of Isolation*. Both of these measures have positive bias in that their expected value under a random spatial distribution is positive rather than zero. We show that this bias is affected by group size empirically. As residential sorting is affected by not only the distribution of population but also the relative size of population groups, the interpretation and comparison of groups and areas in terms of residential sorting using these measures is problematic because of their sensitivity to group size.

In contrast, while the *Entropy Index of Segregation* measure of residential sorting is also biased and the bias is also affected by group size, our empirical data demonstrate that the relationship between group size and the *Entropy Index of Systematic Segregation* is the weakest among all the measures. We interpret the observed relationship between the empirical *Entropy Index of Systematic Segregation* values and group size as reflecting an underlying behavioural relationship observed in Auckland, in which larger groups are more evenly dispersed spatially rather than just evidence of statistical bias. Moreover, the *Entropy Index of Segregation* also is the only multi-group measure of residential sorting that obeys the principles of organisational equivalence, size invariance, transfers and exchanges (James and Taeuber 1985) and thus the same is true for the *Entropy Index of Systematic Segregation*.[16]

Our paper provides evidence that the *Entropy Index of Systematic Segregation* measure of residential sorting is the measure of residential sorting (among those we tested) that is the least biased by group size. However, our empirical results are based on an analysis within a single region of New Zealand. Therefore, these results should be corroborated by further analysis in other geographical contexts, and with different numbers of groups and areas. In the meantime, though, given the relationship we have identified between group size and measures of residential sorting along with the desirable properties of entropy measures identified in the literature (James and Taeuber 1985), we strongly recommend using the *Entropy Index of Systematic Segregation* for analyzing residential sorting. We also recommend that some conclusions of past studies of residential sorting should be re-interpreted in light of the potential for significant group-size bias in their results.

References

Carrington, W.J., & Troske, K.R. (1997). On measuring segregation in samples with small units. *Journal of Business & Economic Statistics*, *15*(4), 402–409. https://doi.org/10.1080/07350015.1997.10524718

Denton, N. A., & Massey, D. S. (1988). Residential segregation of blacks, hispanics, and asians by socioeconomic status and generation. *Social Science Quarterly (University of Texas Press)*, *69*(4), 797–817.

Duncan, O. D., & Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, *20*, 210-217.

Fossett, M. (2017). New Methods for Measuring and Analysing Segregation. Springer Verlag.

Fowler, C. S., Lee, B. A., & Matthews, S. A. (2016). The contributions of places to metropolitan ethnoracial diversity and segregation: Decomposing change across space and time. *Demography*, *53*(6), 1955–1977. https://doi.org/10.1007/s13524-016-0517-3

Iceland, J., Weinberg, D. H., & Steinmetz, E. (2002). *Racial and Ethnic Residential Segregation in the United States 1980-2000*. Washington, D.C., U.S: U.S. Census Bureau.

James, D. R., & Taeuber, K. E. (1985). Measures of segregation. *Sociological Methodology*, *15*, 1–32. https://doi.org/10.2307/270845

Johnston, R., Poulsen, M., & 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. https://doi.org/10.1111/j.1538-4632.2002.tb01087.x

Johnston, R., Poulsen, M., & Forrest, J. (2005). Ethnic residential segregation across an urban System: The maori in New Zealand, 1991–2001. *The Professional Geographer*, *57*(1), 115–129. https://doi.org/10.1111/j.0033-0124.2005.00464.x

Johnston, R., Poulsen, M., & Forrest, J. (2007). The geography of ethnic residential segregation: A comparative study of five countries. *Annals of the Association of American Geographers*, *97*(4), 713–738. https://doi.org/10.1111/j.1467-8306.2007.00579.x

[16] This is the case because the *Entropy Index of Systematic Segregation* is defined as $(E - E_R)/(1 - E_R)$ and the expected value of E_R is constant across different realisations of the actual spatial distribution of the group. Hence the *Entropy Index of Systematic Segregation* is a simple linear transformation of the *Entropy Index of Segregation.* Since the latter index satisfies the James and Taeuber (1985) criteria, the former does also.

Johnston, R., Poulsen, M., & Forrest, J. (2011). Evaluating changing residential segregation in Auckland, New Zealand, using spatial statistics. *Tijdschrift Voor Economische En Sociale Geografie*, *102*(1), 1–23. https://doi.org/10.1111/j.1467-9663.2009.00577.x

Lichter, D. T., Parisi, D., & Taquino, M. C. (2017). Together but apart: Do US whites live in racially diverse cities and neighborhoods? *Population and Development Review*, *43*(2), 229–255. https://doi.org/10.1111/padr.12068

Maré, D. C., Coleman, A., & Pinkerton, R. (2011). *Patterns of population location in Auckland* (SSRN Scholarly Paper No. ID 1860872). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=1860872

Maré, D. C., Pinkerton, R. M., Poot, J., & Coleman, A. (2012). Residential sorting across Auckland neighbourhoods. *New Zealand Population Review; Wellington*, *38*, 23–54.

Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces*, *67*(2), 281–315. https://doi.org/10.1093/sf/67.2.281

Nijkamp P, Poot J (2015) Cultural Diversity: a matter of measurement. In: Nijkamp P, Poot J and Bakens J (eds.) The economics of cultural diversity. Edward Elgar, Cheltenham, pp. 17-51.

Parry, K., & Eeden, A. van. (2015). Measuring racial residential segregation at different geographic scales in Cape Town and Johannesburg. *South African Geographical Journal*, *97*(1), 31–49. https://doi.org/10.1080/03736245.2014.924868

Reardon, S. F., & Firebaugh, G. (2002). Measures of multigroup segregation. *Sociological Methodology*, *32*(1), 33–67. https://doi.org/10.1111/1467-9531.00110

Spoonley, P. (2014). Superdiversity, social cohesion, and economic benefits. *IZA World of Labor*.

Statistics New Zealand. (2006). Ethnic groups in New Zealand. Retrieved February 18, 2017, from http://www.stats.govt.nz.

Statistics New Zealand. (2013a). *2013 Census definitions and forms. Available from www.stats.govt.nz.* Retrieved from www.stats.govt.nz.

Statistics New Zealand. (2013b). 2013 Census quickstats about culture and identity. Retrieved February 18, 2017, from http://www.stats.govt.nz/Census/2013-census/profile-and-summary-reports/quickstats-culture-identity.aspx

Statistics New Zealand. (2013c). Ethnicity. Retrieved May 1, 2017, from http://www.stats.govt.nz

Statistics New Zealand. (2015). *IDI data dictionary: 2013 census data*. Retrieved from www.stats.govt.nz.

Statistics New Zealand. (2017). 2001 Census of Population and Dwellings: change in ethnicity question. Retrieved November 9, 2017, from http://www.stats.govt.nz/Census/2001-census-data/change-in-ethnicity-question.aspx

Theil, H. (1972). *Statistical decomposition analysis: With applications in the social and administrative sciences*. North-Holland.

van der Pas, S., & Poot, J. (2011). Migration paradigm shifts and transformation of migrant communities: The case of dutch kiwis. *IDEAS Working Paper Series from RePEc; St. Louis*. Retrieved from https://search-proquest-com.ezproxy.waikato.ac.nz/docview/1698258824?rfr_id=info%3Axri%2Fsid%3Aprimo

Vertovec, S. (2019). Talking around super-diversity. *Ethnic and Racial Studies*, *42*(1), 125–139.

Voas, D., & Williamson, P. (2000). The scale of dissimilarity: Concepts, measurement and an application to socio-economic variation across england and wales. *Transactions of the Institute of British Geographers*, *25*(4), 465–481. https://doi.org/10.1111/j.0020-2754.2000.00465.x

White, M. J. (1983). The measurement of spatial segregation. *American Journal of Sociology*, *88*(5), 1008–1018. https://doi.org/10.1086/227768

White, M. J. (1986). Segregation and diversity measures in population distribution. *Population Index*, *52*(2), 198–221. https://doi.org/10.2307/3644339

Wong, D. W. S. (2002). Modeling local segregation: A spatial interaction approach. *Geographical and Environmental Modelling*, *6*(1), 81–97. https://doi.org/10.1080/13615930220127305

Wright, R., Ellis, M., Holloway, S. R., & Wong, S. (2014). Patterns of racial diversity and segregation in the United States: 1990–2010. *The Professional Geographer: The Journal of the Association of American Geographers*, *66*(2), 173–182. https://doi.org/10.1080/00330124.2012.735924

Appendix A

Table A1: Measures of Residential Sorting based on Actual Data: Auckland Region, 1991-2013

Year	**1991**				**1996**				**2001**				**2006**				**2013**			
Ethnicity	Group Size	ISeg	IIsol	EIS	Group Size	ISeg	IIsol	EIS	Group Size	ISeg	IIsol	EIS	Group Size	ISeg	IIsol	EIS	Group Size	ISeg	IIsol	EIS
New Zealand European	574,932	35.1	17.83	14.5	536,606	11.3	2.25	2	616,859	33.9	14.2	13.1	611,901	28.1	10.51	9.5	696,966	33.7	14.49	12.8
Other European	50,532	13.7	0.769	1.8	72,576	12.5	3.03	2.3	50,668	12.1	0.393	2	59,959	13.9	0.601	2.4	36,362	15	0.731	2.6
NZ Maori	85,926	33.9	7.09	9.7	105,213	31.9	4.52	7.8	127,704	29.9	5.61	7.5	137,304	29.2	5.29	7.1	142,767	27.3	3.94	6.1
Samoan	41,784	49.6	9.86	19.6	51,639	52.1	8.66	19.6	76,584	49.7	10.48	18.8	87,840	49.7	10.82	18.9	95,916	51.7	10.93	19.9
Cook Island Maori	17,466	49.9	5.13	17.5	21,234	51.2	4.17	17	31,077	48.1	4.95	15.6	34,371	48.4	5.08	15.5	36,546	53.3	5.88	18.9
Tongan	12,456	52.6	3.14	16.5	17,958	55.7	3.47	17.9	32,535	52.1	5.53	17.8	40,140	52	6.47	18.4	46,971	54.8	6.54	20.2
Niuean	9,354	50.3	2.29	15.2	11,466	53	1.84	15.2	16,038	48.9	2.23	13.9	17,667	48.6	2.08	13.4	18,555	53.4	2.17	15.9
Tokelauan	504	83.2	0.512	23.9	627	83.5	0.316	22.5	1,488	76.6	0.405	19.8	1,848	70.2	0.399	17.6	1,959	86.8	0.616	26.1
Fijian	1,506	50.3	0.299	11.4	3,174	39.1	0.212	7.5	4,155	45.6	0.359	10.2	5,847	38.6	0.335	7.8	8,493	48.8	0.51	11.5
Other Pacific Island	300	86.6	0.334	24.7	1,164	67.8	0.272	16.1	1,755	60.6	0.564	15.9	2,868	54.9	0.973	15.7	1,212	70.4	0.834	19.7
Southeast Asian	1,806	62.1	0.752	17.4	6,561	39.3	0.556	9.1	9,363	34.4	0.879	8.1	15,909	33.7	1.14	7.3	10,911	34.6	1.47	8.1
Chinese	9,738	29.3	0.794	6	23,505	30.8	1.01	6.3	3,8025	37	4.19	11	60,186	40.1	5.72	12	39,456	39.9	6.53	12.4
Indian	7,209	36.2	1.09	8.7	16,905	36.4	1.19	8.5	2,3484	36.2	2.37	9.6	39,262	38.4	3.99	10.9	34,064	41.5	6.72	13.6
Other Asian	231	89.7	0.313	26.6	2,240	48.3	0.271	11	10,086	40.9	1.33	10.9	19,105	39.9	1.97	10.1	12,335	37.9	2.02	9.6
Middle Eastern	282	85.4	0.255	23.1	1,194	56.6	0.138	11.3	3,624	42.4	0.452	9.4	6,897	40.5	0.963	9.9	3,759	47.1	1.26	13.1
Latin American/Hispanic	33	97.2	0.243	36.8	204	89.8	0.126	23.3	474	83.6	0.261	21.6	1,194	72.6	0.222	16.6	2,658	77.4	0.404	19.8
African	45	96.3	0.241	34.6	180	91.3	0.147	25	681	79.4	0.414	21.1	1,932	62	0.889	18.1	927	72.2	0.805	20.3
Others	108	99.4	0.143	40.5	198	96	0.109	30.5	279	99	0.139	40	100,110	19	1.61	3.3	15,639	20	0.321	3.1

Note

ISeg-Index of Segregation, IISol- Index of Isolation, EIS-Entropy Index of Segregation. We have multiplied the index values by 100 for easy interpretability

Table A2: Measures of Residential Sorting Based on Randomised Data

Auckland Region, 1991-2013

Year	1991				1996				2001				2006				2013			
Ethnicity	Group Size	ISeg	IIsol	EIS	Group Size	ISeg	IIsol	EIS	Group Size	ISeg	IIsol	EIS	Group Size	ISeg	IIsol	EIS	Group Size	ISeg	IIsol	EIS
New Zealand European	574,932	1.55	0.048	0.029	536,606	1.03	0.017	0.013	616,859	1.36	0.03	0.023	611,901	1.23	0.026	0.019	696,966	1.67	0.047	0.034
Other European	50,532	3.25	0.061	0.094	72,576	0.976	0.017	0.012	50,668	3.34	0.038	0.093	59,959	3.11	0.034	0.081	36,362	4	0.06	0.137
NZ Maori	85,926	2.47	0.057	0.061	105,213	2.31	0.024	0.048	127,704	2.26	0.035	0.05	137,304	2.2	0.032	0.047	142,767	3.28	0.058	0.099
Samoan	41,784	3.57	0.062	0.111	51,639	3.29	0.025	0.085	76,584	3.12	0.038	0.084	87,840	2.89	0.034	0.073	95,916	4.17	0.061	0.146
Cook Island Maori	17,466	5.61	0.064	0.23	21,234	5.21	0.026	0.181	31,077	4.89	0.039	0.174	34,371	4.56	0.036	0.152	36,546	6.53	0.063	0.303
Tongan	12,456	6.65	0.063	0.307	17,958	5.6	0.026	0.204	32,535	5.01	0.04	0.18	40,140	4.39	0.035	0.141	46,971	6.27	0.062	0.282
Niuean	9,354	7.64	0.066	0.388	11,466	7.06	0.026	0.3	16,038	6.75	0.04	0.296	17,667	6.4	0.036	0.265	18,555	9.27	0.064	0.54
Tokelauan	504	32.9	0.062	4.83	627	30.2	0.026	3.87	1,488	26.6	0.041	3.23	1,848	22.2	0.037	2.31	1,959	38	0.064	6.32
Fijian	1,506	19	0.064	1.86	3,174	13.4	0.026	0.918	4,155	15.1	0.04	1.18	5,847	12.3	0.036	0.802	8,493	16.4	0.063	1.45
Other Pacific Island	300	42.6	0.027	7.29	1,164	22.2	0.026	2.24	1,755	18.1	0.041	1.64	2,868	14.3	0.036	1.05	1,212	22	0.064	2.47
Southeast Asian	1,806	17.4	0.026	1.59	6,561	9.32	0.026	0.485	9,363	7.84	0.041	0.381	15,909	6.11	0.036	0.243	10,911	7.37	0.063	0.369
Chinese	9,738	7.46	0.026	0.369	23,505	4.92	0.026	0.165	3,8025	3.9	0.039	0.121	60,186	3.11	0.034	0.082	39,456	3.8	0.06	0.126
Indian	7,209	8.73	0.026	0.483	16,905	5.79	0.026	0.215	2,3484	4.98	0.04	0.18	39,262	3.87	0.035	0.116	34,064	4.13	0.06	0.143
Other Asian	231	48.4	0.026	8.68	2,240	16	0.026	1.25	10,086	7.63	0.041	0.364	19,105	5.54	0.035	0.207	12,335	6.87	0.062	0.326
Middle Eastern	282	43.4	0.025	7.53	1,194	21.7	0.026	2.15	3,624	12.7	0.041	0.875	6,897	9.17	0.037	0.484	3,759	12.5	0.063	0.896
Latin American/Hispanic	33	89.5	0.027	21.3	204	52.5	0.026	8.91	474	38.6	0.042	6.13	1,194	28.3	0.036	3.59	2,658	32.2	0.065	4.81
African	45	86	0.027	19.3	180	56.5	0.026	9.7	681	29.4	0.041	3.9	1,932	17.4	0.036	1.49	927	25.4	0.064	3.18
Others	108	92.2	0.028	27.4	198	84.8	0.027	17.7	279	97.2	0.042	29.4	100,110	2.52	0.033	0.058	15,639	8.9	0.063	0.502

Note

ISeg-Index of Segregation, IISol- Index of Isolation, EIS-Entropy Index of Segregation. We have multiplied the index values by 100 for easy interpretability

Table A3: Systematic Measures of Residential Sorting

Auckland Region, 1991-2013

Year	1991				1996				2001				2006				2013			
Ethnicity	Group Size	ISeg	IIsol	EIS	Group Size	ISeg	IIsol	EIS	Group Size	ISeg	IIsol	EIS	Group Size	ISeg	IIsol	EIS	Group Size	ISeg	IIsol	EIS
New Zealand European	574,932	34.1	17.8	14.5	536,606	10.4	2.23	1.99	616,859	33	14.18	13.1	611,901	27.2	10.51	9.48	696,966	32.6	14.46	12.8
Other European	50,532	10.8	0.723	1.71	72,576	11.6	3.01	2.29	50,668	9.06	0.354	1.91	59,959	11.1	0.565	2.32	36,362	11.5	0.671	2.47
NZ Maori	85,926	32.2	7.05	9.64	105,213	30.3	4.5	7.76	127,704	28.3	5.57	7.45	137,304	27.6	5.26	7.06	142,767	24.8	3.88	6.01
Samoan	41,784	47.7	9.82	19.5	51,639	50.5	8.63	19.5	76,584	48.1	10.45	18.7	87,840	48.2	10.79	18.8	95,916	49.6	10.88	19.8
Cook Island Maori	17,466	46.9	5.09	17.3	21,234	48.5	4.14	16.8	31,077	45.4	4.91	15.5	34,371	45.9	5.05	15.4	36,546	50	5.81	18.7
Tongan	12,456	49.2	3.09	16.2	17,958	53.1	3.44	17.7	32,535	49.6	5.5	17.7	40,140	49.8	6.44	18.3	46,971	51.8	6.48	20
Niuean	9,354	46.2	2.25	14.9	11,466	49.4	1.81	14.9	16,038	45.2	2.19	13.6	17,667	45.1	2.05	13.2	18,555	48.6	2.11	15.4
Tokelauan	504	75	0.463	20	627	76.4	0.289	19.4	1,488	68.1	0.364	17.1	1,848	61.7	0.363	15.6	1,959	78.7	0.552	21.1
Fijian	1,506	38.6	0.25	9.72	3,174	29.6	0.186	6.64	4,155	35.9	0.319	9.13	5,847	30	0.298	7.05	8,493	38.7	0.446	10.2
Other Pacific Island	300	76.7	0.284	18.8	1,164	58.6	0.246	14.2	1,755	51.9	0.523	14.5	2,868	47.4	0.937	14.8	1,212	62	0.768	17.7
Southeast Asian	1,806	54.1	0.705	16.1	6,561	33.1	0.529	8.66	9,363	28.8	0.84	7.75	15,909	29.4	1.11	7.07	10,911	29.4	1.41	7.76
Chinese	9,738	23.6	0.747	5.65	23,505	27.2	0.99	6.15	3,8025	34.4	4.15	10.9	60,186	38.2	5.68	11.9	39,456	37.5	6.48	12.3
Indian	7,209	30.1	1.04	8.26	16,905	32.5	1.17	8.3	2,3484	32.9	2.33	9.44	39,262	35.9	3.95	10.8	34,064	39	6.67	13.5
Other Asian	231	80	0.264	19.6	2,240	38.4	0.244	9.87	10,086	36	1.29	10.6	19,105	36.4	1.93	9.91	12,335	33.3	1.96	9.3
Middle Eastern	282	74.2	0.207	16.8	1,194	44.6	0.112	9.4	3,624	34	0.411	8.6	6,897	34.5	0.93	9.46	3,759	39.5	1.19	12.3
Latin American/Hispanic	33	73.4	0.195	19.7	204	78.5	0.1	15.8	474	73.3	0.22	16.5	1,194	61.8	0.186	13.5	2,658	66.7	0.341	15.7
African	45	73.6	0.192	18.9	180	80	0.121	16.9	681	70.8	0.373	17.9	1,932	54	0.852	16.9	927	62.7	0.742	17.7
Others	108	92.3	0.12	18	198	73.7	0.082	15.6	279	64.5	0.095	15	100,110	16.9	1.57	3.24	15,639	12.2	0.255	2.61

Note

ISSeg = Index of Systematic Segregation, ISIsol = Index of Systematic Isolation, EISS = Entropy Index of Systematic Segregation. We have multiplied the index values by 100 for easy interpretability

Figure A1: Scatterplot of *Index of Isolation* Values and Group Sizes
Based on Randomised Data, Auckland Region, 1991-2013