Ethnic Mobility and The Spatial Distribution of Ethnicity in Auckland

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Economics

at

The University of Waikato

by

MOHANA MONDAL

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2020

Abstract

Understanding a country's past, present and expected future population diversity at sub-national levels is important. Studying changing diversity in terms of the groups that constitute a population and how that varies within regions and between local areas assists in understanding socio-economic and demographic sub-national trends. It is important to project the probable future population diversity in regions for successful policy planning and implementation, group-specific investments in health, education, and community services, as well as the provision of non-government services.

This thesis is a compilation of four inter-related studies that examine the ethnic makeup of the Auckland region of New Zealand. Using New Zealand Census of Populations and Dwellings data (1991, 1996, 2001, 2006 and 2013), the first study identifies and empirically demonstrates the shortcomings in traditional measures commonly used for measuring residential sorting, and instead proposes an alternative preferred measure. Specifically, the study shows that the Entropy Index of Systematic Segregation is the measure of residential sorting that is least biased by group size.

Using the same data, the second study examines the long-term patterns in ethnic and economic residential sorting in Auckland at a fine geographic scale using disaggregated groups. The results show that residential sorting by ethnicity is much more prominent than sorting by economic factors. The results also show that, although residential sorting has been declining over time in Auckland, specific ethnic groups like the Chinese and Indians have become more residentially sorted over time. The New Zealand European, Other European, and New Zealand Māori groups were found to be the least residentially sorted, whereas small ethnic groups like the African, Latin American/Hispanic, Tokelauan, and 'Other Pacific Island' groups were the most residentially sorted, over the whole study period. The results also show that the dominant feature of residential sorting in Auckland is the sorting of subgroups (e.g. Chinese, Indian and South East Asian) within broad ethnic groups (e.g. Asian).

Using the New Zealand Linked Census (NZLC) data, the third study sheds light on the factors that predict the self-identified ethnic affiliation of adolescents in Auckland. The results show statistically significant relationships between the adolescents' ethnic identity and the ethnic identity assigned to them five to seven years previously by their parents. Additionally, the ethnic affiliation of adolescents is also associated with their age, sex, having been born in New Zealand, the ethnic makeup of the neighbourhood they live in, and their parents' ethnic identities. The results confirm patterns of complementarity between ethnicities and ethnic groups that are consistent with other research.

Finally, using NZLC data, the fourth study in this thesis describes the construction and calibration of a spatial microsimulation model, which can be used to project the expected future ethnic residential sorting and ethnic diversity in Auckland. Results show that our model is capable of reproducing the dynamics of residential sorting in Auckland with minimum error.

Notes on Publication

Chapter 2 has been accepted for publication:

Mondal, M., Cameron, M.P., & Poot, J. (forthcoming). Group-size bias in the measurement of residential sorting. In R. Patuelli & S. Suzuki (Eds.), *A Broad View of Regional Science - Essays in Honor of Peter Nijkamp*. Springer Nature.

Chapter 3 has been published:

Mondal, M., Cameron, M.P., & Poot, J. (2020). Cultural and economic residential sorting of Auckland's population, 1991-2013: An entropy approach. *Journal of Geographical Systems.* https://doi.org/10.1007/s10109-020-00327-1

Chapter 4 has been published as a University of Waikato Working Paper:

Mondal, M., Cameron, M.P., & Poot, J. (2020). *Determinants of ethnic identity among adolescents: Evidence from New Zealand.* (Working paper in Economics No.5/20). Hamilton, New Zealand: University of Waikato.

Chapter 5 has been published as a University of Waikato Working Paper:

Mondal, M., Cameron, M.P., & Poot, J. (2021). *Projecting the spatial distribution of ethnic groups in Auckland: Development of a spatial dynamic microsimulation model*. (Working paper in Economics No. 12/21). Hamilton, New Zealand: University of Waikato.

Acknowledgements

"Alone we can do so little; together we can do so much."

-Hellen Keller

This is the most important lesson that I take away from my PhD journey. It has taken help, cooperation and inspiration from a lot of people to complete my PhD and I would like to thank all of them.

I would like to begin by expressing my profound gratitude to my supervisors: Associate Professor Michael Cameron and Emeritus Professor Jacques Poot. I thank them for their patience, valuable suggestions, constructive criticisms and constant support throughout my PhD journey. They have played the role of mentors and guides in the best ever way possible for me to complete this thesis. I am indebted to them for all the knowledge and skills that I have developed during this journey.

I am extremely grateful to the University of Waikato for providing financial support for my PhD through the University of Waikato Doctoral Scholarship. I would also like to thank the 2014–2020 Capturing the Diversity Dividend of Aotearoa New Zealand (CaDDANZ) programme, funded by Ministry of Business, Innovation and Employment grant UOWX1404 for the generous funding received during the later stage of this PhD.

I would like to sincerely thank all the staffs and members of the Department of Economics, at the University of Waikato. I would specially like to thank Maria Fitzgerald and Denise Martin for their administrative support. I would like to express my gratitude to Renae Dixon who not only provided me administrative help, but has also been a constant emotional support throughout my PhD journey. I would also like to thank the staff of the Statistics New Zealand Micro-data Access team for their help with the data used in this thesis.

I would like to thank Geua Boe-Gibson and Dave Maré for their help in creating the maps in this thesis. I am extremely thankful to Robert Didham and Arthur Grimes for their detailed comments and suggestions which have helped to significantly improve of this work. I would like to extend my gratitude to Aldo Benini for his advice in the development of the microsimulation model in this thesis. I would also like to thank the participants at the NZAE (New Zealand Association of Economists) conference, held in Auckland in June 2018, and the

WEAI 94th Annual Conference, held in San Francisco, on June 2020 for sharing their valuable inputs.

Most importantly, I am indebted to my family: my father, Arindam Mondal; my mother: Sharmistha Mondal; and my elder sister, Meghna Bose for their endless support and constant encouragement. Being away from you all during this study was extremely difficult. You have always been there to support me through my tough times as well as cherish the good moments in my journey.

I would also like to thank my fiancé Abhiroop Nandi Ray and my best friend Bhagyashree Chowdhury for patiently and willingly listening to me talking about my research even when it was incoherent to them. You both have been always there to boost me up in my low times and encouraged me to keep going.

Finally, I would like to thank all my PhD colleagues and friends, especially Omoniyi (Niyi) Alimi, for their support, and help.

Disclaimer

The results in this paper are not official statistics. They have been created for research purposes from Census unit record data in the Statistics New Zealand Datalab. The opinions, findings, recommendations, and conclusions expressed in this paper are those of the authors, not Statistics New Zealand. Access to the anonymised data used in this study was provided by Statistics New Zealand under the security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation and the results in this paper have been confidentialised to protect these groups from identification and to keep their data safe. Careful consideration has been given to the privacy, security and confidentiality issues associated with using unit record census data.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Overview

The spatial distribution of a region's population in terms of the population's cultural, social and economic characteristics is a common topic of research. Ethnicity is one of the key variables which governments and private organisations use when conducting research related to diversity, equity of access, and socio-economic outcomes within a population. It is crucial to know about the past, present and projected future ethnic composition of a region for monitoring the demographic, social and economic progress of ethnic groups, for policy evaluation, and for ethnic-group-specific social service delivery (Goodyear 2009). Moreover, the ethnic composition of a region, along with its ethnicity-specific demographic and socio-economic characteristics, are determinants of health status and health care utilisation patterns (Ministry of Health 2001). It is therefore important to investigate the possible future spatial distribution of different ethnic groups for planning future public services, particularly in health, education and community services, as many of these are targeted at particular ethnic groups (Callister 2007). Understanding future ethnic diversity is also important for private and non-government organisations in order to understand the potential future demand for their services (Cameron and Poot 2019).

Studies on ethnicity include, but are not limited to, ethnic identity, ethnic residential sorting, and cultural/ethnic diversity. Ethnic identity is one's identity or sense of self as a member of one or more ethnic groups. Changing incentives and circumstances influence an individual's desire to belong to, or change to, a certain ethnic group (Phinney 2001). It has been recognised that individuals can have multiple ethnicities and might change their ethnic affiliation(s) over time (e.g., Carter et al. 2009; Simpson et al. 2016). This social phenomenon of individuals changing their ethnic identity or affiliation over time, is known as ethnic mobility. Ethnic mobility contributes to the changing ethnic makeup of places. Along with migration, fertility and mortality, ethnic mobility affects the size of the population of each ethnic group. With rising ethnic mobility, multiple ethnic affiliations can become more common. Ethnic mobility can lead to a loss or gain in the population size of a specific ethnic group (Khawaja et al. 2007). In New Zealand, Auckland is the most ethnically diverse city. The trend of people having multiple ethnic identities is impacting notably on Auckland and, given Auckland's large role in New Zealand's population system ultimately the country as a whole (Friesen 2008). Though ethnic mobility affects a large proportion of all ethnic groups in New Zealand (Didham 2016), studies on ethnic mobility in New Zealand are very limited and mainly focus on the individual's self-identification process. Moreover, for convenience, most research applications on ethnic identity have assumed that ethnic affiliation is constant over time (Carter et al. 2009).

Ethnic mobility is most prevalent among teenagers and young adults. For example, the age group in New Zealand that had changed ethnic identity the most between 2001 and 2006 were those aged between 5 and 14 years, followed by individuals belonging to the age group 15-24 (Statistics New Zealand 2009). Consequently, a full chapter (Chapter 4) of this thesis is devoted to determinants of ethnic identity among adolescents.

Based on language, ancestry, religion or customs, cultural diversity can exist even among people who share the same ethnic identity. It is also possible that people with different ethnic identities are sometimes very similar culturally (Maré and Poot 2019). The manifestation of cultural identity in terms of people's behaviour can depend on factors such as social capital, institutional quality and interactions between dissimilar people (Kemeny and Cooke 2017). Although cultural diversity may increase at local levels as a result of immigration, it is often observed that migrant groups tend to co-locate with their own group members (White and Glick 1999). Such sorting, segregation or clustering can reduce ethnic mobility. In contrast, the composition of international migration flows may increase the socio-economic diversity of a region when immigration overall is an amalgamation of high- and low-skilled migrant workers, which have different cultural backgrounds. The cultural diversity of many regions is expected to grow in the future, in part as a consequence of international migration (Poot and Pawar 2013). Greater population diversity might improve economic performance through innovation and the quality of life experienced by residents (Ottaviano and Peri 2006; Kemeny and Cooke 2018).

The degree to which different groups live separate from each other is popularly known as residential segregation or residential sorting[1] (Denton and Massey 1988; Johnston et al. 2007). In a summary of the literature on residential sorting, Denton and Massey (1988) 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)

[1] We prefer the term 'residential sorting', in order to encompass a range of measures of residential segregation including dissimilarity, isolation, and concentration (e.g. Massey and Denton 1988) and also to avoid any negative connotations associated with use of the word 'segregation'.

clustering. Each dimension brings out different features of the spatial distribution of social groups.

One of the major concerns in many countries is the impact of residential sorting on individual wellbeing and opportunities (e.g. Bennett 2011). Neighbourhood composition influences social and economic outcomes (Maré et al. 2012). In the literature, many geographical, historical, institutional, economic and behavioural factors have been identified as determinants of spatial sorting (e.g. Musterd 2005). Residential sorting can occur in terms of age, religion, ethnicity, race, income or other socio-economic characteristics like industry of work or occupation. All these characteristics are argued to be interrelated (Schelling 1971). For example, people decide on where to reside according to their preferences and constraints that may be impacted by cultural affiliation. They usually prefer to stay in close contact with people who they are familiar with or with whom they share similar characteristics (e.g. common ethnicity, language etc.). This results in groups of similar people clustering together. Another source of similarity influencing residential preferences and choices is that individuals with similar jobs are prone to have similar incomes and thus, their affordability for housing is also similar (Schelling 1971).

The socio-economic consequences of growing diversity and spatial sorting of the population are complex and can influence the wellbeing of individuals in positive or negative ways (Nijkamp et al., 2015; UNESCO). Previous studies of ethnic residential sorting (e.g. Forrest et al. 2006; Johnston et al. 2007; Lee et al. 2019; Hall et al. 2019) mostly show detrimental effects of growing diversity leading to increased spatial segregation. Pre-existing inequalities (e.g. earnings, wealth and poverty; see Grodsky and Pager 2001) may be intensified if poor neighbourhoods are concentrated with particular ethnic groups, making some groups more vulnerable to social problems (e.g. lower quality social institutions, increased crime, lower education and lower employment opportunities; see Massey and Denton 1993; Halpern-Felsher et al. 1997). On the other hand, growing cultural diversity has been shown to have positive impacts in terms of consumption patterns, decision making and innovation (Page 2007). More research is needed to better understand this growing diversity and its impacts, including in terms of spatial sorting, to maximise the benefits and adapt to the changes in diversity (Spoonley 2014).

1.2 Research Articles and Significance of the Research

The central objective of this thesis is to study the ethnic composition of Auckland, New Zealand. The main body of the thesis is comprised of four substantive chapters, which cover various research questions related to ethnicity in Auckland, New Zealand.

As a consequence of migrant settlement, international migration, growing ethnic diversity, population ageing, changing fertility and urban growth, and inter-ethnic marriage, the ethnic makeup of New Zealand is changing. This thesis is an effort to prepare New Zealand to respond and adapt to these changes and to maximise the benefits from an increasingly diverse population. The Auckland region contains 33.4 percent of the New Zealand population, making it the largest city in New Zealand (Statistics New Zealand 2018). The diversity in Auckland is mainly in terms of ethnicity, country of birth, socio-economic status, languages spoken, gender and age (Auckland Council 2018). Auckland, along with being the most diverse city in New Zealand, is also one of the most diverse cities in the world, with 41.6 percent of its population born overseas (Statistics New Zealand 2018), more than 200 ethnicities represented, and more than 160 languages spoken (Royal Society of New Zealand 2013). The ethnic composition in Auckland is predicted to evolve and change, with strong growth of the Asian and Pacific ethnic groups (Auckland Council 2018). The growth within the Māori and European categories is expected to be muted (Auckland Council 2018). The changing ethnic composition of Auckland is also a leading indicator of changes in ethnic composition in other parts of New Zealand. These changes are likely to have consequences for economic development, policy and national identity in Auckland, and these consequences are likely to spread to other parts of New Zealand (Friesen 2008). Given its high and growing diversity, Auckland was chosen as the area of focus for this thesis.

This thesis uses data obtained from the 1991, 1996, 2001, 2006 and 2013 New Zealand Census of Populations and Dwellings data. The New Zealand census collects data on individual characteristics which can be aggregated to different spatial scales. This thesis uses data aggregated to the area unit level.[2] Ethnicity is classified into a hierarchy of four levels (Levels

---

[2] Area units are non-administrative areas that are aggregations of meshblocks. In urban areas, an area unit is similar in size to a suburb or neighbourhood (Statistics New Zealand 2013). The area unit boundaries have changed over the years. Statistics New Zealand maintains a concordance file so that boundaries relating to earlier area unit patterns can also be generated. To ensure consistencies in area units used in this thesis, 2013 area unit boundaries have been used.

1 to 4)[3] according to the New Zealand Standard Classification of Ethnicity. The main (Level 1) ethnic groups defined in this classification are the European; Māori; Pacific People; Asian; Middle Eastern, Latin American and African (MELAA); and Others. These Level 1 ethnic groups are too broad to capture the considerable expected heterogeneity within each group. Thus, compared to the ethnic groups used in previous research in New Zealand (and comparable work elsewhere), this thesis considers ethnic groups at a finer scale (Level 2).[4]

In the Census, data on self-reported ethnic identification are collected and each individual can choose a single or multiple response.[5] Instead of using the more common approach of adopting prioritised ethnicity [6](e.g Johnston et al. 2005; Maré et al. 2012; Maré et al. 2016), the analyses in this thesis use every ethnicity that the individual reports, to avoid placing any ranking on the individual's preferences or choices, and also to avoid ignoring the diversity arising from multiple-ethnic affiliation. The analyses in this thesis explores the behaviours of almost all of the Level 2 ethnic groups.[7] The chosen level of the ethnic classification and the fine spatial scale in the analyses in the thesis are the maximum feasible levels of disaggregation for suitable derivation and interpretation of the results.

There have been previous studies on residential sorting by ethnicity/race, ancestry, education, income and/or occupation in many countries, particularly in the U.S. (Duncan and Duncan 1955; Florida and Mellander 2018; Lee et al. 2019; Hall et al. 2019), Canada (Balakrishnan et al. 2015; Fong and Hou 2009), Australia (Forrest et al. 2006; Johnston et al. 2007) and New Zealand (Johnston et al. 2002; 2005; 2011; Maré et al. 2016). 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*. These measures are

---

[3] See Table 2.1 of Chapter 2 for details.

[4] Statistics New Zealand classifies ethnicity into four levels (levels 1,2,3 and 4). Level 1 ethnic groups (New Zealand European or Other, Māori, Pacific, Asian, and Middle Eastern/Latin American/African) are the major ethnic group classifications. Level 2 ethnic groups consists of finer ethnic groups than Level 1. For example, Asian not further defined, South East Asian, Chinese, Indian and Other Asians are Level 2 ethnic groups under the Level 1 Asian category (Statistics New Zealand 2013).

[5] Up to three responses were recorded for each individual in 1991 and 1996 compared with up to six in the later Censuses. A very small fraction of individuals chooses more than one ethnicity. For example, in 2013, about 90 percent of the population self-identified with only one ethnicity and only 0.05 percent self-identified with belonging to four or more ethnicities (Statistics New Zealand 2015a). Thus, the change in number of responses is not of concern. Moreover, in Chapter 4, for the regressions we include census fixed effects to control for changes in the census ethnicity question.

[6] Prioritisation is a classification which assigns just one ethnicity to the person who has reported multiple ethnic responses (Didham 2005).

[7] Some Level 2 ethnic groups have been combined together due to their group size being very small.

simple to calculate and interpret and hence are used widely. But these aforementioned measures can suggest the presence of substantial residential sorting, even when there is none, such as when small groups are included in the calculations (Carrington and Troske 1997; Voas and Williamson 2000; Johnston et al. 2011). Moreover, the expected values of these measures are positive rather than zero under random sorting, resulting in these measures being biased by group size. Thus, the common practice of resorting to these indices to compare residential sorting of groups with different group sizes, both cross-sectionally and over time, is inappropriate. There has been to date relatively little systematic analysis of this issue.

Thus, the first article in the thesis (included as Chapter 2), titled **'Group-size bias in the measurement of residential sorting'**, empirically demonstrates the presence of group-size bias in the traditional measures of residential sorting and identifies a preferred index to measure residential sorting, which is least affected by group size variation. In this article, each measure of residential sorting is calculated using New Zealand census data for Auckland, from 1991 to 2013, and also the same measures are calculated based on data where individuals have been randomly spatially allocated. A random allocation should, in theory, lead to no sorting in expectation. The group-size bias in the traditional indices is demonstrated by plotting the relationship between group sizes and values of each measure of residential sorting (based on both actual and randomised data), regressing these index values on group sizes, calculating the measurement bias for each index, and also plotting the index values against group sizes.

Given the background demographic and socio-economic changes in Auckland in recent decades, the second article (included as Chapter 3 of the thesis), titled **'Cultural and economic residential sorting of Auckland's population, 1991-2013: An entropy approach'**, focuses on identifying the changes in residential sorting in Auckland from 1991 to 2013. This article specifically looks into whether residential sorting in Auckland has been declining over time, whether residential sorting by cultural (ethnicity) factors or socio-economic (education, occupation, income) factors is greater, and also whether residential sorting is more driven by sorting within or between the broad (Level 1) ethnic groups. Though this work can be considered as an extension of the previous studies done on residential sorting in New Zealand as well as elsewhere in the world, it has a number of novel aspects. First, this article is the first to use the entropy-based measure of spatial sorting and diversity (established in the previous chapter as the preferred measure) for New Zealand data. Second, this article captures spatial sorting in Auckland for a much longer period (covered by five successive censuses), and using a finer-grained classification of ethnic groups, in comparison to earlier research in New

Zealand (e.g. Johnston et al. 2005; Johnston et al. 2011; Maré et al. 2012). Another novel aspect of this article is the use of an overall socio-economic measure of sorting in Auckland by means of a combination of income, occupation, and qualification (following Florida and Mellander 2018), which has previously not been done for New Zealand data. This article also contributes to the very limited literature on residential sorting by factors other than ethnicity in New Zealand.

Despite the high ethnic mobility rate of the youth population in New Zealand (see Statistics New Zealand 2009), there have been very few studies focussing explicitly on this phenomenon. Moreover, most relevant earlier works (e.g. Kukutai 2007; 2008) describe the ethnic identification processes among children and focus on the impacts of having single/multiple ethnic affiliations and multiple-ethnic parents on a child's ethnic identity, and do not identify the causes associated with the changes in youth ethnic identity choices. Hence, the third article (included as Chapter 4 of the thesis), titled **'Determinants of ethnic identity among adolescents: Evidence from New Zealand'**, contributes to the limited literature on youth ethnic identity in New Zealand by focusing on identifying the predictors of self-declared ethnic identity choices among adolescents in Auckland, taking possibilities of multiple and changing ethnic affiliations into consideration. Using New Zealand Linked Census data for four inter-censal periods between 1991 and 2013, the same individuals, linked across two consecutive Censuses were identified, where in the first of the two Censuses their parents are likely to have recorded the adolescent's ethnicity, and in the second Census the adolescents are likely to have recorded their own ethnicity. Each and every ethnicity that an adolescent reported in the later census of the inter-censal period was included in the analysis and hence logistic regression analysis was used with the linked data pooled across all the inter-censal periods. This article identifies the major determinants of the first conscious ethnic affiliation choice of adolescents. These determinants include their sex, age, whether New Zealand-born, ethnicity stated at the first of the two censuses in the inter-censal period, parents' ethnicity, and the ethnic makeup of the neighbourhood.

Finally, the fourth article (included as Chapter 5 of the thesis), titled **'Projecting the spatial distribution of ethnic groups in Auckland: Development of a spatial dynamic microsimulation model'**, develops a spatial microsimulation model that can be used to project the future small-area ethnic diversity of the Auckland region at a fine spatial scale and maximum feasible disaggregation of ethnic groups. Using 1996-2001 NZLC data for Auckland, a spatial microsimulation model was developed that projects the ethnic population in Auckland in the year 2006. The actual 2006 census data was used to calibrate and validate the model for it to be used in projecting future ethnic diversity. Ethnic diversity projections require assessing diversities at the neighbourhood level, which is difficult due to the lack of relevant longitudinal data and under-developed small-area population projection methods (Cameron and Cochrane, 2017).

The microsimulation model developed in this article runs in five-year time steps, separately for children/adolescents and adults. The ethnic and location transitions of each individuals in the model are derived using logistic regressions. The model captures the possible multiple and changing ethnic identities to project the future ethnic diversity in Auckland. Moreover, the different geographic scales paint different dimensions of residential sorting (Reardon et al. 2009). Thus, analyses at different regional levels are necessary for a broader understanding of changing residential sorting patterns.

This microsimulation model is one of the first tools that is appropriate and adequate to project the future ethnic diversity at a small regional scale (area unit) with minimum error. In New Zealand, a range of government organisations produce population projections at national level (Statistics New Zealand 2014; 2015a) and sub-national level (Statistics New Zealand 2015b). The available official population projections cover only Level 1 ethnic groups (New Zealand European or Other, Māori, Pacific, Asian, and Middle Eastern/Latin American/African) and a limited number of Level 2 ethnic groups (Chinese, Indian, Samoan). One of the novel contributions of Article 4 is the use of disaggregated (Level 2) ethnic groups, which is important given the changing ethnic composition and the heterogeneity within the broad level 1 ethnic groups in New Zealand. Results show that our model is capable of projecting the future ethnic spatial distribution in Auckland with minimum error.

1.3 Thesis Outline

The remainder of the thesis is organised as follows. Chapter 2 answers the question of which measure of residential sorting is least biased by group size (Article 1). Chapter 3 looks into measuring ethnic and cultural residential sorting in Auckland by employing entropy measures of residential sorting and also exploits the decomposability characteristics of these measures (Article 2). Chapter 4 identifies the predictors of self-identified ethnic identities among adolescents in Auckland (Article 3). Chapter 5 develops and validates a spatial microsimulation model suitable to project the future spatial ethnic diversity and residential sorting in Auckland at a fine spatial scale (Article 4). Chapter 6 concludes the thesis.

References

Auckland Council. (2018). *Auckland plan 2050, evidence report*. Retrieved from https://www.aucklandcouncil.govt.nz/plans-projects-policies-reports-bylaws/our-plans-strategies/auckland-plan/about-the-auckland-plan/Evidence%20reports%20documents/evidence-report-developing-the-plan.pdf.

Balakrishnan, T.R., Maxim, P., Jurdi, R. (2005). Residential segregation and socio-economic integration of visible minorities in Canada. *Migration Letters, 2*, 126–143.

Bennett, P.R. (2011). The relationship between neighbourhood racial concentration and verbal ability: an investigation using the institutional resources model. *Social Science Research, 40,* 1124–1141.

Didham, R. (2005). *Understanding and working with ethnicity data*. Wellington, New Zealand: Statistics New Zealand.

Callister, P. (2007). *Special measures to reduce ethnic disadvantage in New Zealand – An examination of their role*. Wellington, New Zealand: Institute of Policy Studies.

Cameron, M.P. & Poot, J. (2019). Towards superdiverse Aotearoa: Dimensions of past and future ethnic diversity in New Zealand and its regions. *New Zealand Population Review, 45*, 18-45.

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

Carter, K. N., Hayward, M., Blakely, T., & Shaw, C. (2009). How much and for whom does self-identified ethnicity change over time in New Zealand? Results from a longitudinal study. *Social Policy Journal of New Zealand, 36,* 32–45.

Duncan, O.D., Duncan, B. (1955). Residential distribution and occupational stratification. *American Journal of Sociology, 60,* 493–503.

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.

Didham, R. (2016). Ethnic mobility in the New Zealand census, 1981-2013: A preliminary look. *New Zealand Population Review*, *42*, 27–42.

Florida, R., & Mellander, C. (2018). The geography of economic segregation. *Social Sciences, 7*(8), 123, 1-17.

Fong, E., & Hou, F. (2009). Residential patterns across generations of new immigrant groups. *Sociological Perspectives, 52*,409–428.

Forrest, J., Poulsen, M. & Johnston, R. (2006). Peoples and spaces in a multicultural nation: Cultural group segregation in metropolitan Australia. *Espace-Populations-Sociétés, 1*, 151–164.

Friesen, W. (2008). *Diverse Auckland: The face of New Zealand in the 21st century?* Auckland, New Zealand: University of Auckland, 1-20. Retrieved from http://hdl.handle.net/2292/8662

Goodyear, R. K. (2009). *The differences within, diversity in age structure between and within ethnic groups.* Wellington, New Zealand: Statistics New Zealand

Grodsky, E., Pager, D. (2001). The structure of disadvantage: individual and occupational determinants of the Black-White wage gap. *American Sociological Review*, *66*, 542–567.

Halpern-Felsher, B., Connell, J., Spencer, M., Aber, J.L., Duncan ,G., Clifford, E., Crichlow, W., Usinger,P., Cole, S., Allen, L., & Seidman, E. (1997). Neighborhood and family factors predicting educational risk and attainment in African American and White children and adolescents. In: J. B. Gunn, G. J. Duncan, J. L. Aber (Eds.) *Neighborhood poverty: Context and consequences for children* (pp. 146-173). New York, NY: Russell Sage Foundation.

Hall, M., Iceland, J., & Yi, Y. (2019). Racial separation at home and work: Segregation in residential and workplace settings. *Population Research and Policy Review, 38*, 671–694.

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*, 245–261.

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*, 115–129.

Johnston, R., Poulsen, M., & Forrest, J. (2007). The geography of ethnic residential segregation: a comparative study of five countries. *Annals of the American Association of Geographers, 97*, 713–738.

Johnston, R. J., Poulsen, M. F., & Forrest, J. (2011). Evaluating changing residential segregation in Auckland, New Zealand, using spatial statistics. *Tijdschrift Voor Economische En Sociale Geografie*: *Journal of Economic and Social Geography*, *102*(1), 1–23.

Kemeny, T., & Cooke, A. (2017). Urban immigrant diversity and inclusive institutions. *Economic Geography*, *93*(3), 267–291.

Kemeny, T., & Cooke, A. (2018). Spillovers from immigrant diversity in cities. *Journal of Economic Geography*, *18*(1), 213–245.

Khawaja, M., Boddington, B. & Didham, R. (2007). *Growing ethnic diversity in New Zealand and its implications for measuring differentials in fertility and mortality.* Wellington, New Zealand: Statistics New Zealand.

Lee, B.A., Farrell, C.R., Reardon, S.F., & Matthews, S.A. (2019). From census tracts to local environments: an egocentric approach to neighborhood racial change. *Spatial Demography*, *7*, 1–26.

Lee, J., & Bean, F. (2010). *The diversity paradox*. New York, NY: Russell Sage Foundation.

Massey, D.S., Denton, N.A. (1993). *American apartheid: Segregation and the making of the underclass*. Cambridge, MA: Harvard University Press.

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

Maré, D.C., Pinkerton, R.M., & Poot, J. (2016). Residential assimilation of immigrants: A cohort approach. *Migration Studies, 4*, 373–401.

Maré, D.C., & Poot, J. (2019). Commuting to diversity. *New Zealand Population Review*, *45*, 125–159.

Ministry of Health. (2001). *Implementing the New Zealand health strategy: The Minister of Health's first report on progress on the New Zealand health strategy*. Wellington, New Zealand: Author.

Musterd, S. (2005). Social and ethnic segregation in Europe: Levels, causes, and effects. *Journal of Urban Affairs*, *27*(3), 331-348.

Ottaviano, G. I. P., & Peri, G. (2006). The economic value of cultural diversity: Evidence from US cities. *Journal of Economic Geography*, *6*(1), 9–44.

Phinney, J. S. (2001). Ethnic identity, psychology of. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 4821–4824). Oxford, England: Pergamon. https://doi.org/10.1016/B0-08-043076-7/01721-6

Poot, J., & Pawar, S. (2013). Is demography destiny? Urban population change and economic vitality of future cities. *Journal of Urban Management*, *2*(1), 5-23.

Qian, Z. (2004). Options: Racial/ethnic identification of children of intermarried couples. *Social Science Quarterly*, *85*(3), 746–766.

Reardon, S.F., Farrell, C.R., Matthews, S.A., O'Sullivan, D., Bischoff, K., & Firebaugh G. (2009). Race and space in the 1990s: Changes in the geographic scale of racial residential segregation, 1990-2000. *Social Science Research*, *38*(1), 55-70

Royal Society of New Zealand. (2013). *Languages in Aotearoa New Zealand*. Retrieved from https://www.royalsociety.org.nz/assets/Uploads/Languages-in-Aotearoa-New-Zealand.pdf

Schelling, T.C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology, 1*, 143–186.

Simpson, L., Jivraj, S., & Warren, J. (2016). The stability of ethnic identity in England and Wales 2001–2011. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *179*(4), 1025–1049. https://doi.org/10.1111/rssa.12175

Spoonley, P. (2014). *Superdiversity, social cohesion, and economic benefits.* Retrieved from https://wol.iza.org/articles/superdiversity-social-cohesion-and-economic-benefits/long.

Statistics New Zealand. (2009). *Final report of a review of the official ethnicity standard*. Wellington, New Zealand: Author. Retrieved from http://archive.stats.govt.nz

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

Statistics New Zealand. (2014). *National Population Projections: 2014 to 2068*. Retrieved from http://www.stats.govt.nz

Statistics New Zealand. (2015a). *National Ethnic Population Projections: 2013(base)–2038*. Wellington, New Zealand: Author. Retrieved from http://archive.stats.govt.nz

Statistics New Zealand. (2015b). *Subnational Ethnic Population Projects: 2013-2038*. Retrieved from https://www. documentcloud.org/documents/2504441-sub-national-ethnicprojections-statistics-new.html

Statistics New Zealand. (2018). *2018 Census data allows users to dive deep into New Zealand's diversity*. Retrieved from https://www.stats.govt.nz/news/2018-census-data-allows-users-to-dive-deep-into-new-zealands-diversity

Statistics New Zealand (2020). *Subnational Population Estimates.* Retrieved from http://datainfoplus.stats.govt.nz

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

The United Nations Educational, Scientific and Cultural Organization (2009). *Investing in cultural diversity and intercultural dialogue.* Paris, France: Author.

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., & Glick, J.E. (1999). The impact of immigration on residential segregation. In F.D. Bean and S.Bell-Rose (Eds.) *Immigration and opportunity: Race, ethnicity and employment in the United States*. New York, NY: Russell Sage Foundation.

Chapter 2: Group-size bias in the measurement of residential sorting

2.1 Introduction

Residential segregation or sorting[8] 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 1997; Reardon and Firebaugh 2002; Fossett 2017), and extant studies mostly use the *Index of Dissimilarity* and/or the *Index of Isolation*. In this chapter, 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.[9] Hence the primary aim of this chapter is to show the sensitivity of popular measures of residential sorting to group size. An unbiased index of sorting would return an average value of approximately zero in the case where people are randomly distributed across areas. Thus, following Carrington and Troske (1997) we calculate modified versions of each of the selected sorting measures, which we refer to as the indices of systematic segregation. Each index of systematic segregation has an expected value of zero under random sorting. We find that the *Entropy Index of Systematic Segregation* measure is least affected by 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 we consider.

---

[8] We use 'residential sorting' as a term that encompasses a range of measures of residential segregation that include dissimilarity, isolation, and concentration (e.g. 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'.

[9] 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.

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 this chapter.

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. Although there is a relationship between income and the patterns and decision making processes of residential mobility (Rosenblatt and DeLuca 2012; Karina *et al.* 2014), Mondal *et al*. (2020) show that in Auckland residential sorting by ethnicity is much more prominent than sorting by income. Thus, we base our analysis on residential sorting by ethnicity. Throughout this chapter, by 'region' we mean the Auckland region, as defined by Statistics New Zealand. The Auckland region is made up of about 409 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 chapter refer to total responses, not total individuals. An ethnic 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 chapter proceeds as follows. In Section 2.2, we briefly discuss some relevant studies on popular measures of residential sorting. Section 2.3 describes the data and Section 2.4 details the methods. Section 2.5 presents and discusses the results, and Section 2.6 concludes.

2.2 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 (e.g. Johnston *et al.* 2002; 2005; 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) summarise the literature on residential sorting to that point in time, and conclude 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, viz. the principles of organisational equivalence, size invariance, transfers, and exchanges. By organisational equivalence, they mean that when an area 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, with the proportion of persons in the respective groups being greater in their original areas 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 the term 'segregation index' in the literature 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 area 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 area 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 twenty-five 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 also 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, i.e. 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 area unit population-weighted 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 chapter we compare the two 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.

2.3 Data

Auckland is the most ethnically diverse region in New Zealand. According to the 2013 Census, [10] the ethnic composition of its population at the time was: European (59.3 percent); Asian (23.1 percent); Pacific Islander (14.6 percent); Māori (10.7 percent); Middle Eastern, Latin American & African (MELAA, 1.9 percent); and Other Ethnicity (1.2 percent) (Statistics New Zealand 2013).[11] 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 information on each person present in New Zealand on census night. For usually-resident individuals the Census provides a range of information about each person, including location, age, sex, ethnicity, income level, occupation, education, marital status, etc., which can be aggregated to population statistics at the meshblock

[10] The most recent population census was held on March 6, 2018. At the time of collecting the data for this chapter, 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.

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

level.[12] The Auckland region is made up of 413 land-based area units,[13] 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, counts and calculations are based on data that have been suppressed for raw counts less than six, and otherwise randomly rounded to base three.[14]

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 Eastern, 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 2.1.[15]

---

[12] 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.

[13] In this chapter, we use 2013 area unit boundaries.

[14] 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.

[15] 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.

Table 2.1: 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 not further defined	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 not further defined	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 not further defined	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
Total responses all ethnic groups		4,450,350			4,542,633

Source: Statistics New Zealand (2013)

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.[16] 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), 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 European' 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).

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.

[16] 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).

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.

2.4 Methodology

As stated in the introduction, the aim of this chapter is to show the sensitivity of popular measures of residential sorting to group size. We achieve this aim using four steps.

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 2.2, applied to Census data for the Auckland region of New Zealand. High values of these indices represent more residential sorting by ethnicity. The values of these indices vary between 0 (when all areas have the same ethnic 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.[17] We then use scatter plots to display the relationship between group size and the value of each index.

[17] 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.

Table 2.2: Summary Measures of Residential Sorting

Index of Segregation	$ISeg_g = \frac{1}{2}\sum_{a=1}^{A}\left\|\frac{P_{ga}}{P_{g.}} - \frac{(P_{.a} - P_{ga})}{(P_{..} - P_{g.})}\right\|$
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	$EIS_g = \sum_{a=1}^{A}\frac{P_{.a}}{P_{..}}\left(1 - \frac{E_a}{\bar{E}}\right)$

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

$\bar{E} = -\frac{P_{g.}}{P_{..}}\, ln\left(\frac{P_{g.}}{P_{..}}\right) - (1 - \frac{P_{g.}}{P_{..}})\, ln\left(1 - \frac{P_{g.}}{P_{..}}\right)$

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).

Second, 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 and the probability of a person being a member of an ethnic group equal to the 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 bias, the expected value of a measure of sorting should be 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 area 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. To check the statistical significance of the effect of group size in relation to the different index values, we ran simple ordinary least squares (OLS) regression of each index value on group size (logarithmic scale), with census fixed effects added to the regression (accounting for growth in the Auckland population). However, the effect of group size identified using OLS could be spurious because of unobserved heterogeneity by ethnicity. To account for this and to again account for the change in average group size across censuses, we also ran regressions with group fixed effects and time fixed effects, using the fixed effect (FE) panel data estimator.

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.[18] 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: $ISIsol := \left(\left[\sum_{a=1}^{A} \pi_{ga} \frac{P_{ga}}{P_{.a}}\right] - \left(\sum_{a=1}^{A} \pi_{ga} \frac{P_{ga}}{P_{.a}}\right)_R\right) \Big/ \left(1 - \left(\sum_{a=1}^{A} \pi_{ga} \frac{P_{ga}}{P_{.a}}\right)_R\right)$. The subscript *R* refers to the average of values based on randomised allocations. We run again OLS and FE linear regression to identify the relationship between group size and the different measures of systematic residential sorting.

Finally, we define the bias 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).

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

2.5 Results and Discussion

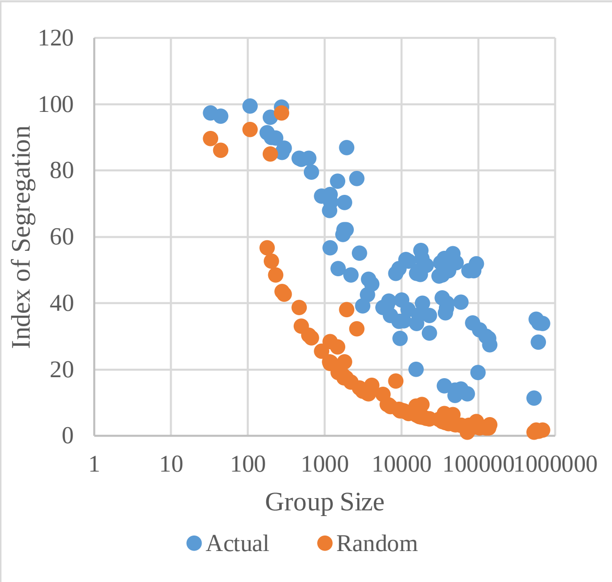
As stated in the introduction, the aim of this chapter is to show that the selected measures of residential sorting are sensitive to (and hence, biased by) group size and propose the best index among these to measure residential sorting. We calculate the values of the measures of residential sorting, 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.[19] As expected, we see that under random spatial allocation the values of the sorting indices are always less than the values based on actual data.

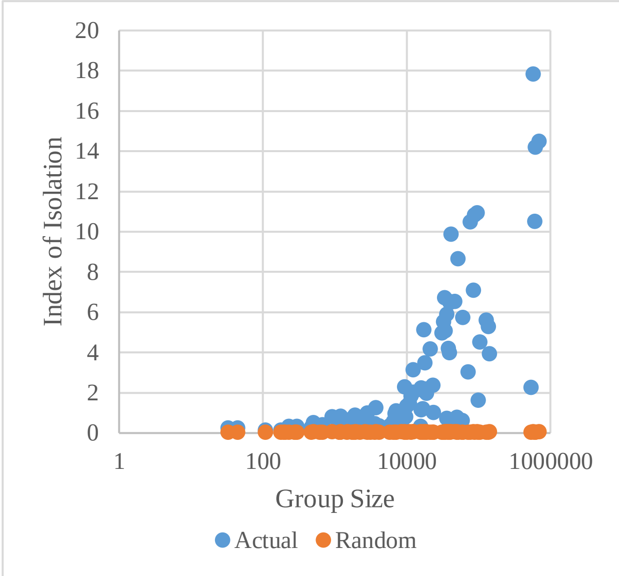
We now plot these index values based on actual data as well as the average values of sorting indices under random allocation, pooled across all five Censuses, against group size, in Figure 2.1. We use a logarithmic scale for group size. The panels in Figure 2.1 show that in the case of residential sorting 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.

[19] 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.

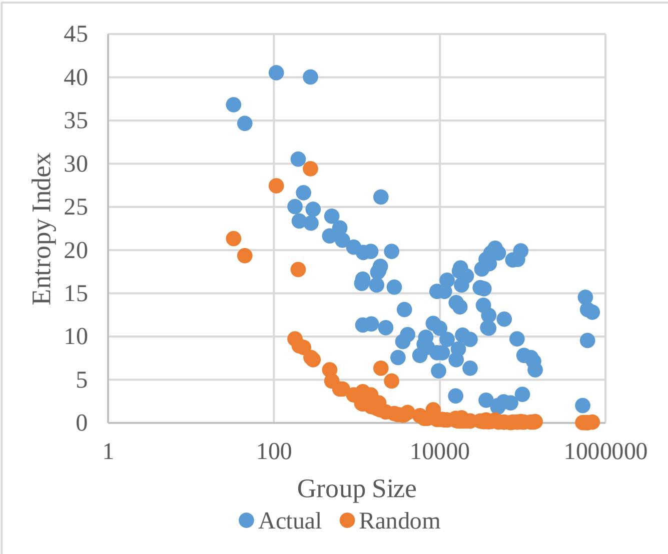
Figure 2.1: Scatterplots of index values and group sizes, based on randomised and actual data: Auckland Region, 1991-2013



2.1(a): *Index of Segregation* and group size



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



2.1(c): *Entropy Index of Segregation* and group size

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.[20] 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 positively correlated. This can be seen in Table 2.3 for our Auckland data. With sorting observed for 18 groups in 5 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*.

Table 2.3: 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

[20] 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. This group size effect has been investigated previously by Fossett (2017) in empirical terms with US data.

As noted in section 2.4, 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, using ordinary least square (OLS) estimation method and also with census and groups fixed effects, using the panel fixed effect estimation (FE) method. The results are shown in Table 2.4.

Table 2.4: Effect of Group Size on Sorting Indices

(a) Regression Results from Actual Data						
	OLS			**FE**		
	ISeg (1)	**IIsol (2)**	**EIS (3)**	**ISeg (4)**	**IIsol (5)**	**EIS (6)**
Log Group Size	-8.466***	1.252***	-2.366***	-13.063***	0.245	-5.534***
	(0.634)	(0.13)	(0.304)	(1.010)	(0.248)	(0.458)
R^2 within				0.78	0.17	0.75
R^2 between				0.70	0.67	0.34
R^2 overall	0.70	0.54	0.47	0.75	0.41	0.44
(b) Regression Results from Randomized Data						
Log Group Size	-8.000***	0.00049	-1.676***	-15.399***	0.005***	-4.141***
	(0.611)	(0.00043)	(0.201)	(0.867)	(0.001)	(0.314)
R^2 within				0.87	0.78	0.78
R^2 between				0.70	0.00	0.47
R^2 overall	0.71	0.68	0.51	0.68	0.45	0.47

Notes:

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

Standard errors in parentheses,* $p<0.10$, ** $p<0.05$, *** $p<0.01$

OLS = Ordinary Least Square regression with census fixed effects

FE = Fixed Effect regression with group and census fixed effects.

ISeg = *Index of Segregation*

IIsol = *Index of Isolation*

EIS = *Entropy Index of Segregation*

Columns (1) to (3) of Table 2.4 report the estimators from the OLS regressions, whereas columns (4) to (6) report the same for the FE regressions. Part (a) of the table reports the statistical significance of group size (logarithmic scale) in relation to the different index values based on the actual data. From columns (1) to (3) of part (a) of Table 2.4, 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 in absolute value than for the *Index of Segregation* (-8.47), even though they are similar measures. From columns (4) to (6), we see that group size is statistically significantly correlated with the *Index of Segregation* and the *Entropy Index of Segregation*, at the 1% level of significance. Again, we see that the coefficient for group size (-5.53) is much smaller in absolute value for the *Entropy Index of Segregation* than for the *Index of Segregation* (-13.06). We observe that the coefficient for group size for the *Index of Isolation* from both the OLS regression (1.25) and the FE regression (0.25) is smaller in absolute value 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, by running simple linear regression with census fixed effects, we observe that group size is statistically significantly correlated with both *Entropy Index of Segregation* and *Index of Segregation* (Table 2.4, part (b), Columns (1) and (3)). However, we observe that the absolute value of 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 the *Entropy Index of Segregation* is less affected by group size in the case of random sorting. We saw in Figure 2.1 that the *Index of Isolation* values after randomisation are almost zero and Column (2) in part (b) of Table 2.4 shows that there is no statistically significant relationship between the isolation measure and group size with randomised sorting. Results from the FE regression show that group size is statistically significantly correlated with all the measures, at the 1% level of significance (Table 2.4, part (b), Columns (4), (5) and (6)). Again, we find that the coefficient for group size for the *Entropy Index of Segregation* (-4.14) is much smaller in absolute value than that of the *Index of Segregation* (-15.4).

Following Carrington and Troske (1997), we next calculate the *Index of Systematic Segregation* for each index (Appendix Table A3)[21] and then check the statistical significance of the

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

relationship with group size (logarithmic scale) using simple linear regression with census fixed effects (Table 2.5, Columns (1), (2) and (3)) and fixed effect regression with group and census fixed effects (Table 2.5, Columns (4), (5) and (6)). From columns (1) to (3), we see that all three of the indices of systematic segregation are sensitive to group size, with the effect being statistically significant at the 1 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. From columns (4), (5) and (6) we observe that the *Index of Systematic Segregation* and the *Entropy Index of Systematic Segregation* are sensitive to group size, at 1% level of significance. Again, the coefficient of log group size for the *Index of Systematic Segregation* (-10.01) is much more negative than that of the *Entropy Index of Systematic Segregation* (-2.11).

Table 2.5: Effect of Group Size on Indices of Systematic Segregation

	OLS			**FE**		
	ISSeg (1)	**ISIsol (2)**	**EISS (3)**	**ISSeg (4)**	**ISIsol (5)**	**EISS (6)**
Log Group Size	-6.432*** (0.648)	1.254*** (0.130)	-0.980*** (0.243)	-10.015*** (1.107)	0.244 (0.249)	-2.112*** (0.329)
R^2 within				0.64	0.17	0.50
R^2 between				0.56	0.67	0.13
R^2 overall	0.57	0.54	0.20	0.64	0.34	0.19

Notes:

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

Standard errors in parentheses,* $p<0.10$, ** $p<0.05$, *** $p<0.01$

OLS = Ordinary Least Square regression with census fixed effects

FE = Fixed Effect regression with group and census fixed effects

ISSeg = *Index of Systematic Segregation*

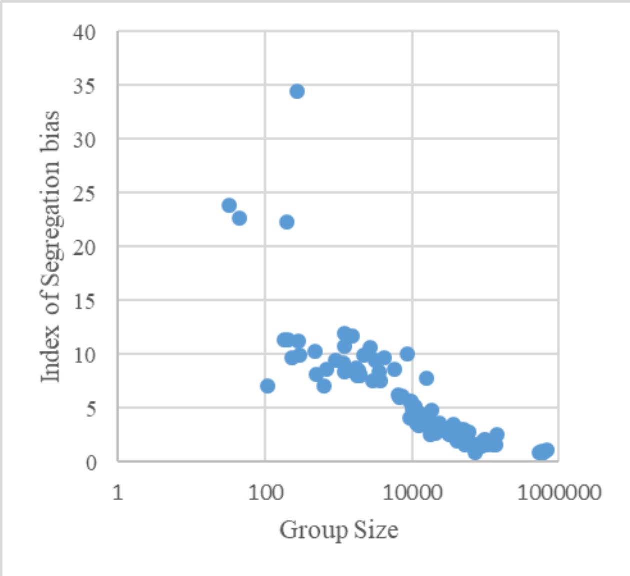
ISIsol = *Index of Systematic Isolation*

EISS = *Entropy Index of Systematic Segregation*

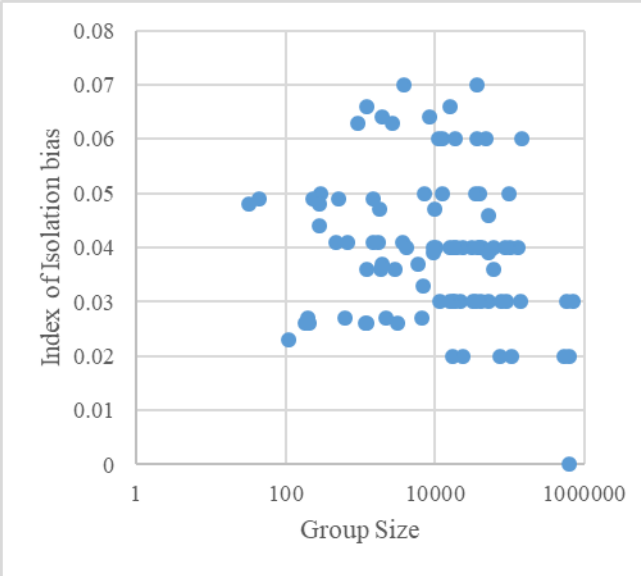
Comparing the values from Table 2.4, part (a), columns (1) to (6), with those of Table 2.5, we conclude that the *Entropy Index of Systematic Segregation* is the best measure, as the coefficient of group size for this measure both in the case of OLS (-0.98) and FE (-2.11) is much smaller in absolute terms than in the case of the *Index of Systematic Segregation* based on actual data.

Finally, we calculate the bias values for each of the three original indices and plot them against group sizes (on a logarithmic scale) in Figure 2.2. 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 that group size has a far less notable effect on the bias defined as the difference between the *Index of Isolation* and the Maré *et al.* (2012) modification of this original index. Recall that under random sorting the values of the *Index of Isolation* itself appear to be almost zero irrespective of group size (see Figure 2.1, Panel (b)).

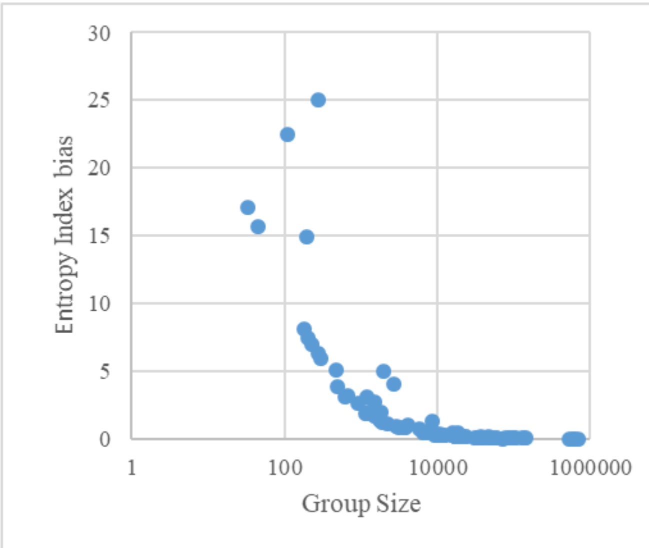
Figure 2.2: Scatter plot of index bias and group size



2.2(a): Relationship between *Index of Segregation* bias and group size



2.2(b): Relationship between *Index of Isolation* bias and group size



2.2(c): Relationship between *Entropy Index of Segregation* bias and group size

We run a simple OLS regression, with census fixed effects (Table 2.6, Columns (1), (2) and (3)), and also FE regression with group and census fixed effects (Table 2.6, Columns (4), (5) and (6)) to see the relationship between the index bias and the group size (on a logarithmic scale). From the OLS estimation results, we find that group size is negatively related to the index bias values, with statistical significance at the 1% level in all three cases. Moreover, we observe that the coefficient for the *Entropy Index of Segregation* (-1.39) is somewhat smaller in absolute terms than the coefficient for the *Index of Segregation* (-2.03), while the effect of group size on bias is very little indeed for the Index of Isolation.[22] While the decline in the bias with group size is somewhat greater with the *Index of Segregation* than with the *Entropy Index of Segregation*, the bias of the latter is generally smaller. We observe this from our results from the FE regression. From columns (4) and (6) we see that, the decline in bias for *Entropy Index of Segregation* (-3.42) with increase in group size (for a given group) is greater than that for the *Index of Segregation* (-3.05).

Table 2.6: Effect of Group Size on Sorting Index Bias (Difference between Original Measures and Systematic Indices)

	OLS			FE		
	ISeg-ISSeg (1)	**IIsol-ISIsol (2)**	**EIS-EISS (3)**	**ISeg-ISSeg (4)**	**IIsol-ISIsol (5)**	**EIS-EISS (6)**
Log Group Size	-2.034*** (0.163)	-0.002*** (0.00033)	-1.387*** (0.167)	-3.048*** (0.427)	0.001* (0.001)	-3.422*** (0.259)
R^2 within				0.50	0.85	0.78
R^2 between				0.82	0.55	0.46
R^2 overall	0.67	0.78	0.50	0.66	0.57	0.47

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

Standard errors in parentheses,* $p<0.10$, ** $p<0.05$, *** $p<0.01$

OLS = Ordinary Least Square regression with census fixed effects

FE = Fixed Effect regression with group and census fixed effects.

ISeg = *Index of Segregation*; ISSeg = *Index of Systematic Segregation*

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

EIS = *Entropy Index of Segregation*; EISS = *Entropy Index of Systematic Segregation*

[22] Because regression coefficients are linearly related to the dependent variable, the coefficients in Table 2.6 can of course also be obtained by subtracting the coefficients in Table 5 from the corresponding columns in Table 2.4. However, Table 2.6 also reports the R^2 (within, between and overall) and the correct standard errors of the regressions of bias on group size.

Overall, our results show that all sorting measures considered are sensitive to group size. However, we find that the *Entropy Index of Systematic Segregation* is much less affected by group size than the *Index of Systematic Segregation*. Moreover, the *Entropy Index of Systematic Segregation* is an unbiased index because it has an expected value zero with randomisation (Carrington and Troske (1997).

2.6 Conclusion

The aim of this chapter 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 empirically that this bias is affected by group size. 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 effect of group size on the index value is the least with the *Entropy Index of Systematic Segregation*.

We interpret the observed empirical relationship between the *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*.[23]

Our chapter 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

[23] 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.

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 the results of these studies.

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.

Denton, N. A., & Massey, D. S. (1988). Residential segregation of blacks, hispanics, and asians by socioeconomic status and generation. *Social Science Quarterly*, *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, Cham.

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.

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

James, D. R., & Taeuber, K. E. (1985). Measures of segregation. *Sociological Methodology*, *15*, 1–32.

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.

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.

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.

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.

Karina, S., Jack, B., & Clara, H. M. (2014). Ethnicity, education and income, and residential mobility between neighbourhoods. *Journal of Ethnic and Migration Studies*, *40*(4), 512-527.

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.

Maré, D. C., Coleman, A., & Pinkerton, R. (2011). *Patterns of population location in Auckland*. (Motu Working Paper 11–06). Wellington: Motu Economic and Public Policy Research.

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

Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces*, *67*(2), 281–315.

Mondal, M., Cameron, M.P., & Poot, J. (2020). Cultural and economic residential sorting of Auckland's population, 1991-2013: An entropy approach. *Journal of Geographical Systems.* https://doi.org/10.1007/s10109-020-00327-1

Nijkamp, P. & Poot, J. (2015) Cultural diversity: a matter of measurement. In: P. Nijkamp, J. Poot & J. 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.

Reardon, S. F., & Firebaugh, G. (2002). Measures of multigroup segregation. *Sociological Methodology*, *32*(1), 33–67.

Rosenblatt, P., & DeLuca, S. (2012). "We don't live outside, we live in here": Neighborhood and residential mobility decisions among low-income families. *City and Community*, 11, 254–284.

Spoonley, P. (2014). Superdiversity, social cohesion, and economic benefits. *IZA World of Labor*. Retrieved April 17, 2020, from https://wol.iza.org/articles/superdiversity-social-cohesion-and-economic-benefits.

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.* 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.

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*. May 1, 2017, from http://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.

Theil, H. (1972). *Statistical Decomposition Analysis: With Applications in the Social and Administrative Sciences*. North-Holland Pub. Co.

van der Pas, S., & Poot, J. (2011). Migration paradigm shifts and transformation of migrant communities: The case of Dutch Kiwis. *CReAM Discussion Paper No. 12/11.*London: Centre for Research and Analysis of Migration, University College London.

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.

White, M. J. (1983). The measurement of spatial segregation. *American Journal of Sociology*, *88*(5), 1008–1018.

White, M. J. (1986). Segregation and diversity measures in population distribution. *Population Index*, *52*(2), 198–221.

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

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*, *66*(2), 173–182.

Chapter Appendix

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

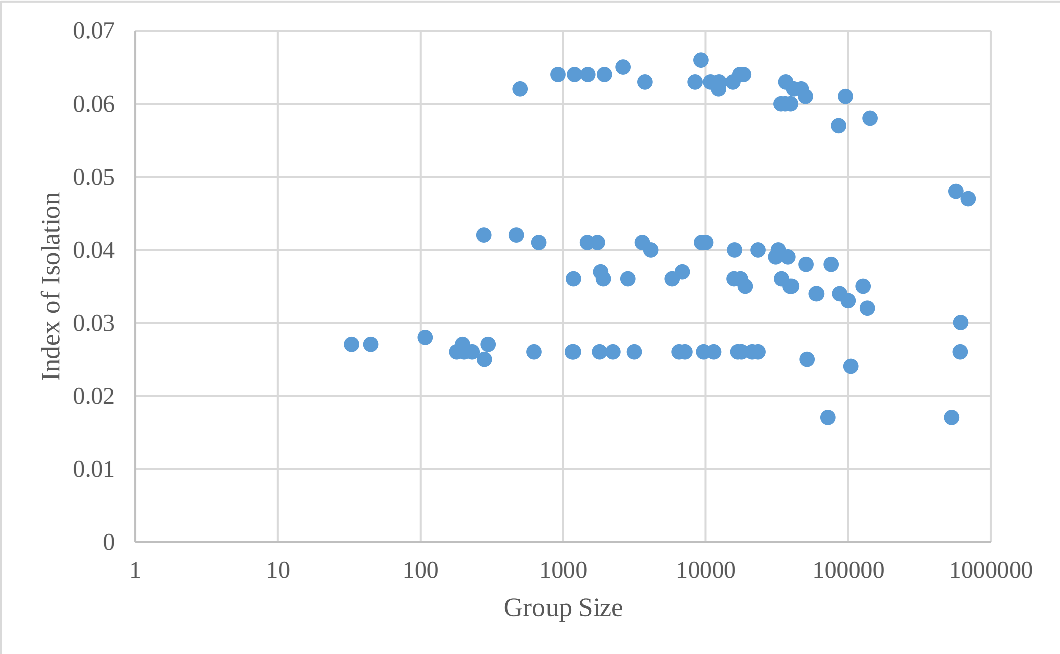
Note: ISeg = *Index of Segregation*, IISol = *Index of Isolation*, EIS = *Entropy Index of Segregation*. We have multiplied the index values by 100 for easy interpretability

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



Chapter 3: Cultural and economic residential sorting of Auckland's population, 1991-2013: An entropy approach

3.1 Introduction

A ubiquitous and persistent phenomenon around the world is that the spatial distribution of a city's population is, in terms of its cultural and socio-economic characteristics, not random but systematic and clustered. Such residential segregation, also referred to more broadly as spatial sorting, can be thought of as the degree to which groups live away from each other (Denton and Massey 1988; Johnston et al. 2007). Spatial sorting has many geographical, historical, institutional, economic and behavioural determinants (e.g. Musterd 2005). Residential sorting can occur in terms of age, language, religion, ethnicity, race and income, or other socio-economic characteristics like industry of work, or occupation.

Schelling (1971) argued that all of the characteristics that may exhibit residential segregation are interrelated. People locate according to their preferences and constraints, and individuals like to stay in close contact with people with whom they share similar characteristics. Networks are often driven by common ethnicity or language use, as such networks facilitate communication and trust. This leads people of the same cultural identity to cluster together. Moreover, house prices and rents are spatially highly correlated, leading to clearly defined low cost and high cost housing areas. Consequently, people may be found to live near others with a similar income, as their capacities to afford housing are then similar. Industry and occupation are, besides age and education, also important predictors of income. People with similar jobs tend to have similar incomes, generating another source of similarity of residential preferences and choices (Schelling 1971). Understanding and measuring existing residential sorting patterns is crucial for forecasting future housing demands, local transport, and infrastructural and communal facilities, as well as services such as education and health.

Neighbourhood composition influences social and economic outcomes (Maré et al. 2012). The repercussions of residential segregation for individual well-being and opportunities (e.g. Bennett 2011) are a major concern in many countries. If particular socio-economic groups are concentrated in particular neighbourhoods, this may exacerbate existing inequalities in terms of earnings, wealth and poverty (Grodsky and Pager 2001). Racially concentrated poor neighbourhoods may be more susceptible to social problems like lower quality social institutions, increased crime, low property values, lower education levels, and lower employment opportunities (Halpern-Felsher et al. 1997; Massey and Denton 1993).

One important, and related, trend in recent decades is the strong growth in international migration which has been making cities more culturally diverse and is expected to continue to do so in the future (Poot and Pawar 2013). The migrant flows' mixture of temporary and permanent highly skilled 'talent' and lower skilled workers may increase socio-economic diversity of a city, in addition to cultural diversity. Immigration may increase diversity at local levels, but counteracting this is the tendency of migrant groups to cluster as well (White and Glick 1999).

A particularly interesting case is that of Auckland, the largest city of New Zealand, which has become one of the most diverse cities in the world, with more than 40 percent of its population born abroad, more than 200 ethnicities represented, and 160 languages spoken. Much of this diversity is due to immigration since the 1990s, but this has been superimposed on historical diversity resulting from a strong presence of the indigenous Māori population, many of whom were attracted to the city from their iwi (tribal) areas for employment (e.g. Pool, 1991). Auckland is now highly diverse in terms of ethnicity, country of birth, socio-economic status, gender, and age (Auckland Council 2018).

Consequently, we focus in this paper on the cultural and economic diversity of Auckland. We measure cultural diversity by ethnicity. Ethnicity is an integral expression of an individual's culture (Betancourt and López 1993). In the New Zealand Census, the ethnicity of an individual is defined as including any ethnic group that the individual identifies with (Statistics New Zealand 2013a). New Zealand residents can affiliate themselves with multiple ethnicities in the Census and some other collections of official data (Kukutai 2008). The extent to which individuals have been identifying with multiple ethnic groups has been increasing. Moreover, resulting from large increases in migration flows since the 1990s – with recruitment based on job skills, financial assets and family ties – and the abolition of a governmental preference for traditional source countries (the United Kingdom and some other European countries), there has been a rise in the number of distinct ethnic identities in New Zealand (New Zealand Ministry for Culture and Heritage 2015). Hence the ethnic composition of New Zealand is changing, with the Māori, Pacific and Asian ethnic group proportions growing faster than the European proportion (Statistics New Zealand 2004). The population of New Zealand has also a high rate of residential mobility, as well as increasing inter-ethnic marriage and cohabitation (Statistics New Zealand 2007). To maximise the benefits and adapt to changes associated with such an increasingly diverse population, more research is needed to better understand this growing diversity and its impacts (Spoonley 2014).

Table 3.1 shows the growth and changing ethnic mix of Auckland's population between 1991 and 2013.[24] Over this period, Auckland's population grew from 0.9 million in 1991 to 1.4 million in 2013, and accounts for about one third of New Zealand's population. The ratio of the number of ethnicities declared (total responses) to the population increased between 1991 and 2013 from 1.05 to 1.11, which is indicative of growth in people identifying with more than one ethnicity over this period. It should be noted, however, that the number of individuals without a stated or imputed ethnicity increased from one percent to six percent of the population. European ethnicity decreased from 72 percent of total responses in 1991 to 54 percent in 2013. If we define 'superdiversity' as the case in which no single major ethnic group represents a majority in the population, it is clear that Auckland is close to becoming superdiverse (see also Cameron and Poot 2019).

[24] Data from the 2018 Census of Population and Dwellings were not yet available at the time of writing of this paper.

Table 3.1: Ethnic composition (level 1, total responses) of the Auckland population, 1991-2013

	1991	1996	2001	2006	2013
European	72.40	67.08	62.81	51.17	53.52
Māori	10.55	11.30	10.61	10.03	9.68
Pacific peoples	11.39	11.79	12.85	13.00	13.22
Asian	5.36	9.15	12.60	17.12	20.83
Middle Eastern, Latin American, African (MELAA)	0.28	0.66	1.11	1.36	1.69
Other	0.01	0.02	0.02	7.32	1.06
Total	100.00	100.00	100.00	100.00	100.00
Total responses	981,786	1,118,595	1,203,612	1,368,354	1,474,848
Total people with stated ethnicity	933,729	1,012,212	1,101,594	1,239,054	1,331,427
Average stated number of ethnicities per person	1.05	1.11[a]	1.09	1.10	1.11
Total people without stated ethnicity	10,047	56,436	57,297	65,907	84,123
Total people, Auckland	943,777	1,068,645	1,158,891	1,304,958	1,415,550

[a] Note that this average stated number of ethnicities per person in 1996 is not directly comparable to that for 1991 and 2001, due to the increase in the number of multiple ethnicity responses in 2006 partially resulting from a change in the information provided for the ethnicity question. Adjusting for that, the average stated number of ethnicities per person in 1996 may be estimated to be about 1.07.

Source: Statistics New Zealand (2019a; 2019b)

Those who report that they identify with Māori ethnicity represent a fairly stable ten percent of total responses. During the 19th century colonialization period, this indigenous population lost much of their lands and resources. They also tended to live in poorer and more crowded houses than Pākehā.[25] As noted above, many Māori migrated after the Second World War to the cities for work. Postwar industrialisation and import substitution policies led to very low unemployment and a high demand for labour. Since the 1950s, Pacific people were also encouraged to migrate to New Zealand’s cities, particularly Auckland, to meet the growing demand for labour. When economic conditions deteriorated in the 1970s, restrictions on Pacific migration were increased. A points system for immigration introduced in the 1990s also favoured skills over family ties. Some Pacific migration nonetheless continued. Over the 1991-2013 period, the proportion of responses identifying with a Pacific ethnicity increased from eleven percent to thirteen percent.

From the late 1980s and the removal of the ‘traditional source country’ criterion, migrants from non-traditional source countries began migrating to New Zealand in larger numbers, especially from Asia. In 1991 only five percent of Auckland's ethnicity responses identified with an Asian ethnicity, but the proportion increased sharply to about 21 percent in 2013. Though the Asian population has increased in every region in New Zealand, the largest increase has been observed in Auckland (Statistics New Zealand 2019b). The largest two Asian population sub-groups in 2013 were Chinese and Indian (Statistics New Zealand 2019b). Besides employment-related migration, another cause of the growth in the Asian population is a large influx of international students undertaking tertiary studies, some of whom are settling in Auckland afterwards.

Responses of ethnicities from the Middle East, Latin America, and Africa (MELAA) and ‘Other’ make up a very small but growing percentage of total responses, up to 2.7 percent in 2013. The large percentage of ‘Other’ in 2006 is an anomaly due the introduction on the census form of a separate ethnicity of ‘New Zealander’, which was highly publicised and politicised in the media at the time and was mostly selected by New Zealand Europeans. The category was kept in the 2013 census but the number selecting it at that time had dropped by 85 percent compared with 2006.

The growing ethnic diversity of Auckland’s population is clearly impacting on the patterns of segregation and spatial sorting that we will analyse in this paper. In the remainder of the paper,

---

[25] Pākehā are non-Māori, usually of European ethnic origin or background.

we prefer to use the terms 'residential sorting' or 'spatial sorting' where possible, to encompass a range of spatial population distribution phenomena that include segregation, isolation, and concentration. Our preferred terms are not only broader than the conventional term of spatial segregation, but also carries none of the negative connotations associated with the latter.

Spatial sorting can create a vicious cycle of disadvantages – a lack of secure and well-paid employment in one's neighbourhood, or at commuting distance, leads to low income, which in turn leads to low quality housing. Low quality housing makes it hard to maintain good health. Low income can create barriers to access to good education, which leads to low future employment opportunities for children, which reinforces income inequality across generations (Dalziel 2013). This makes it important to understand how sorting patterns by economic variables are related to sorting patterns by cultural variables (such as ethnicity).

Income inequality in New Zealand rose rapidly during the 1980s and early 1990s, and this increase was more rapid than in other developed countries (Alimi et al., 2018). Additionally, income inequality increased particularly fast in Auckland (Alimi et al., 2016). While inequality has been fluctuating since the beginning of the 21st century, the Global Financial Crisis triggered a further increase. Socio-economic inequality intersects with ethnicity. In 2013, the average income of Māori was 78.9 percent of that of non-Māori. One-third of Māori aged over 15 had no school qualifications, and only six percent of Māori and two percent of Pacific people held a bachelor's degree. Though there have been improvements in socio-economic indicators (life expectancy, education, employment and income) over time, there has been a relative decline in the number of Māori employed in skilled occupations. Pacific people are also a relatively large proportion of the unemployed, lower-skilled and low-income workers in Auckland, and have substantially lower incomes than other ethnic groups (Auckland Council 2018). Māori and Pacific peoples live disproportionately in low-income households due to a complex set of circumstances, economic transformations and a succession of past policies, since colonial times for the former, and since the 1970s for the latter.

Given this background to the demographic and socio-economic changes in Auckland in recent decades, in this paper we focus on identifying the changes in residential sorting over time. Specifically, the purpose of our paper is to address the following research questions:

i. Has residential sorting been declining over time in Auckland?

ii. Is residential sorting by cultural factors greater than residential sorting by economic factors in Auckland?

iii. Is residential sorting mostly driven by sorting *between* broad groups, or *within* broad groups (i.e. by sorting between sub-groups)?

While this is not the first paper to consider these, or related, research questions, there are several novel aspects to our analysis. First, we use entropy as the mathematical principle for measuring both spatial sorting and diversity. While entropy is not an uncommon approach to diversity and sorting in the literature, our paper is to our knowledge the first contribution using entropy in the New Zealand context. One of the main advantages of entropy measures is their property that an aggregate index can be decomposed into the weighted sum of within-group and between-group measures (Theil 1972). We use this property to see how sensitive the residential sorting index values are to the level of aggregation in our data, and to answer our third research question.

The second contribution of this paper is that we consider spatial sorting in Auckland over a fairly long period of nearly a quarter century (1991-2013), while earlier work has tended to capture shorter periods. Third, while earlier work has addressed the impact of varying granularity of the spatial data (i.e. the definition and size of areas), we are able to quantify the effect of changing the granularity of the classification. We do this for a cultural variable (ethnicity) and an economic variable (occupation).

Regarding the first research question, Manley et al. (2015) found that, at a micro-scale, ethnic residential sorting in Auckland declined from 2001-2013. Related to the third research question, Manley et al. (2019) found that the intensity of segregation for larger ethnic groups in Auckland remained static over the 2001 to 2013 period, but reduced drastically for smaller ethnic groups. Here we revisit these trends over the longer period 1991-2013. A longer time frame is important given the radical economic reforms that took place in New Zealand during the decade following 1984 (Evans et al., 1996).

Regarding the second research question, Maré et al. (2012) found stronger residential sorting by ethnicity than by income or qualification in Auckland, but using data for 2006 only. This New Zealand finding is consistent with U.S. evidence of greater segregation by ethnicity than by social class measured by education or occupation or income (Farley 1977; Sims 1999). Here we revisit whether residential sorting in Auckland is greater by ethnicity than economic factors (income, qualification and occupation) when we use our dataset for the 1991-2013 period.

Regarding the third question, past New Zealand studies (Johnston et al. 2008; Maré et al. 2011) have already found that similar groups (i.e. sub-groups belonging to a larger ethnic group) tend to co-locate. That suggests a high degree of sorting of ethnic sub-groups within high-level ethnic groups. Decomposing multi-ethnic segregation in Auckland at multiple spatial scales was recently undertaken by Manley et al. (2019), following Lichter et al. (2015) who used the Theil Index to decompose metropolitan segregation in the U.S. into its within- and between-place components from 1990 to 2010. Fowler et al. (2016) undertook a similar kind of study to evaluate the roles of area types in ethno-racial change. Our study complements these earlier works, by considering within-and-between ethnic group and occupational group components of sorting rather than spatial components.

The remainder of the paper proceeds as follows. In Section 3.2, we discuss relevant studies on residential sorting, with a particular focus on North American, Australian, and New Zealand research. Section 3.3 describes the data, and Section 3.4 details the methods. Section 3.5 presents and discusses the results, and Section 3.6 concludes.

3.2 Background Literature

Of all countries in which there has been research on residential sorting by ethnicity/race, education, income and/or occupation, the largest number of studies have been conducted for the U.S. Recent reviews that refer to key contributions to this vast literature can be found in Lee et al. (2019) and Hall et al. (2019). In one of the earliest such studies, Duncan and Duncan (1955) found that the most segregated occupational groups were the ones with the highest and the lowest rankings in terms of socioeconomic status. Farley (1977) measured the degree of socioeconomic and residential segregation in central cities and densely populated suburban areas and found that minority individuals in the U.S. tended to cluster with other minority group members. Simkus (1978) found that gross occupational residential segregation in urbanized areas increased slightly during the 1950s but, taking race into consideration, levels of racial residential segregation between White residents and non-White residents in the lowest occupation groups in 1960 were low. Massey (1979) used 1970 Census data and found that segregation of the Spanish-American and White populations declined with increases in socio-economic status. Denton and Massey (1988) used data from the 1980 U.S. Census to look into patterns of residential segregation by socioeconomic status. They showed that the Black population were strongly segregated from the ‘Anglo’ population irrespective of their occupation, educational qualification, or income. Ellis et al. (2004) found ethnic minority groups to be more segregated in the labour market than in the housing markets, and that more intergroup contact takes place during work hours than in the home environment, which results in less workplace segregation. Johnston et al. (2004) demonstrated that the interurban variations in segregation levels between U.S. Metropolitan Statistical Areas are strongly related to urban size, ethnic diversity and relative size of the individual minority groups.

Overall, studies in the U.S. (e.g., Domina 2006; Duncan and Duncan 1955; Farley 1977; Fischer 2003) demonstrate substantial residential segregation based on ethnicity and socioeconomic variables. Education, occupation and income make up an individual’s social status together with ethnicity (Weeden and Grunsky 2005) and these dimensions are related and jointly reinforcing. Florida and Mellander (2018) therefore compared cultural with occupational, income and educational segregation as well as a combined measure of overall economic segregation. They emphasise that income is a consequence of education and occupation and, thus, to understand economic sorting, the latter factors should be considered as well. They applied measures of sorting to the different economic variables, and formed an Overall Economic Sorting Index by averaging the sorting index values for the individual economic variables. They found that economic segregation is associated with more highly educated, larger and denser metro regions. They also found that economic segregation is related to ethnicity, mode of transport and income inequality.

There is also a substantial literature on residential sorting outside of the U.S. For New Zealand research, Canadian studies are also particularly relevant. Balakrishnan et al. (2005) conducted a comparative study on residential segregation across major CMAs (Census Metropolitan Areas) in Canada using 2001 census data. They found considerable variation in segregation levels across these CMAs. They did not find any systematic relationship between residential segregation and socio-economic achievements (education, occupation and income). Walks and Bourne (2006) used 1991 and 2001 Canadian census data and found that Toronto, Vancouver, Montréal and Winnipeg were the most residentially segregated CMAs in Canada. They also found that the Black population and the Latin American population show patterns of high residential segregation, as they are less economically successful than the other ethnic groups. Fong and Hou (2009) looked into residential patterns of three minority groups (South Asian, Chinese and Black populations) in the four largest metropolitan areas of Canada (Calgary, Montreal, Toronto, and Vancouver) using 2001 census data. They found that these minority groups show patterns of residential integration over generations.

In Australia, studies of residential sorting are based on ancestry data, as the Australian census does not ask any direct question related to an individual's ethnic identification. Instead, respondents can state up to two ancestries and, for the foreign born, country of birth is also known. Forrest et al. (2006) found spatial desegregation of non-host ancestral groups and Aboriginal people in metropolitan regions of Australia, using data from the 2001 Census. Their results suggest that the presence of ethnic clusters is a temporary phenomenon in Australia. Johnston et al. (2007) used 2001 Australian census data to describe levels of segregation in Australia and to analyse the factors affecting the levels of segregation. They found that residential segregation was most prominent in larger cities and where the minority ethnic groups formed a large proportion of the total population. Johnston et al. (2016) used 2011 Australian census data and analysed residential segregation of 42 ancestral groups in Sydney. They found that segregation is more prominent among smaller ancestral groups, the most recently arrived, and individuals who are culturally different from the host society. For all ancestral groups, segregation was greater at the macro (regional) and micro (neighbourhood) level than at the intermediate meso (suburban district) levels.

For New Zealand, most studies have focused on ethnic residential sorting using data from the population census. In contrast with our paper, which covers the 1991-2013 period, there have been few previous studies concerned with longer-term trends in residential sorting. Moreover, previous studies of residential segregation in New Zealand have mainly looked at a limited number of ethnic groups, or groups by country of origin or birth (e.g. Maré et al., 2016). Johnston et al. (2002) showed the presence of prominent residential concentration patterns among Polynesians (that is, Pacific Peoples plus Māori). Johnston et al. (2005) analysed variations in the degree of residential segregation of the Māori population across the urban areas of New Zealand from 1991 to 2001. They found that the degree of segregation for this ethnic group varies according to the relative group size within each urban area. Johnston et al. (2008) showed that, in 2006, the Pacific Islander group was the most residentially segregated in Auckland. Johnston et al. (2011) used New Zealand Census data from 1991 to 2006 and found that, in comparison to Māori, Pacific Peoples were more likely to cluster in areas where their co-ethnics dominated.

Few studies in New Zealand have looked at residential sorting by characteristics other than ethnicity. Like Johnston et al. (2008), Maré et al. (2011) found that the greatest residential sorting in Auckland is by Pacific Peoples, but also by people with university degrees. In another paper, Maré and Coleman (2011) found that ‘own-group’ attraction was a much stronger determinant of residential sorting than urban amenities. Maré et al. (2012) found that the Pacific Islanders, people with higher university degrees and with higher levels of education, higher income, and the elderly, exhibited the greatest levels of residential sorting. Finally, Maré et al. (2016) studied the residential assimilation of immigrants after their arrival in Auckland, using census data from 1996-2006. The groups included in the study were limited to immigrants from the United Kingdom, China, India, South Africa, and the Republic of Korea. They found distinct patterns of residential assimilation for most of the immigrant groups. They also found that the longer that immigrants from each group had spent in the host country, the more their residential concentration declined. Manley et al. (2015) looked at changing ethnic residential sorting among the main four broad ethnic groups (European, Māori, Asian and Pacific Peoples) in Auckland for the period from 2001 to 2013. They found that at each of three geographical scales (macro (localities); meso (area unit); micro (meshblock)), Pacific Peoples were the most and Europeans were the least residentially segregated. They also found that a decline in residential sorting at the micro (meshblock) level could be observed for Māori, Asian and Pacific Peoples.

As noted in the introduction, our paper contributes to the growing literature on residential sorting in New Zealand. We use a finer-grained categorisation of ethnic groups than used in previous research in New Zealand to better capture the heterogeneity within the broad ethnic groups. Unlike previous research in New Zealand, we also look into long-term trends (close to a quarter century) of residential sorting and we use entropy as a mathematical principle for measuring sorting and diversity. Additionally, we measure overall economic sorting in Auckland by means of a combination of income, occupation and qualification (following Florida and Mellander 2018). Finally, we also consider how much between-group and within-group sorting contributes to the overall level of sorting, which has not been previously done in New Zealand (or elsewhere, to our knowledge).

3.3 Data

We obtained population data from the 1991, 1996, 2001, 2006, and 2013 New Zealand Census of Population and Dwellings for the Auckland metropolitan region of New Zealand. The New Zealand Census of Population and Dwellings is usually conducted every five years (the 2011 census was delayed until 2013 due to a large earthquake in Christchurch) and collects a range of socio-demographic information on each member of the New Zealand population present and normally resident in New Zealand on census night. The census data on each individual include characteristics such as location of usual residence, age, sex, ethnicity, income level, occupation, education, and marital status. These microdata can be aggregated to population statistics at various spatial levels. For the purpose of the present paper, each measure of residential sorting (described below) was calculated based on data aggregated to the area unit level for individuals aged 22 years and above.[26] The Auckland region is made up of 413 area units. Their median area is 169 hectares (1.3 km by 1.3 km). Four area units had no usually resident population in any of the censuses and were therefore dropped, leaving 409 for the analysis.

In accordance with the strict confidentiality rules laid down by Statistic New Zealand, the summary statistics, counts and calculations are based on data that have been suppressed for raw counts less than six and otherwise randomly rounded to base three.[27]

An ethnic group consists of people who generally have any of the following: common proper name of the group, common elements of culture, similar interests, feelings and actions, or share a common ancestral as well as geographic origin (Statistics New Zealand 2013a). A person's ethnicity is the ethnic group or groups that that person identifies with or feels a sense of belonging to. It is a measure of cultural affiliation (in contrast to race, ancestry, country of birth, or citizenship). Ethnicity is self-perceived, and a person can belong to more than one ethnic group. New Zealand residents can change their ethnic affiliation for statistical purposes at any time.

According to the New Zealand Standard Classification of Ethnicity, ethnicity is classified in a hierarchy of four levels. 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

[26] Area units are non-administrative areas that are aggregations of meshblocks. In urban areas, an area unit is similar in size to a suburb or neighbourhood (Statistics New Zealand 2013a). We use 2013 area unit boundaries.

[27] Counts that are already a multiple of three are left unchanged, and all other counts are rounded randomly either up or down to be a multiple of three.

Zealand 2015). The main (Level 1) ethnic groups defined in the Census are: European, Māori, Pacific Peoples, Asian, Middle Eastern, Latin American & African (MELAA) and Others. Given the considerable heterogeneity expected within each of these broad ethnic groups, we also use data on Level 2 ethnic groups.[28] In our analysis, we proportionally distributed the population counts of the 'not further defined' category for three Level 2 ethnic groups into the corresponding Level 2 groups within the same Level 1 ethnic group.[29]

Two issues affect the comparability of ethnicity data in New Zealand over time. First, the format and wording of the Census ethnicity question changed twice between 1991 and 2001. In 1991 and 2001, the question was almost the same, but both differed substantively from the question in 1996.[30] Thus, comparability across Censuses is likely to be affected. Substantial changes include increased multiple responses in 1996 and 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). These inconsistencies apply particularly to the 'European' ethnic groups (including 'New Zealand European') and the 'Māori' ethnic group. In the 1996 data, the count for 'Other Europeans' was much higher than in the 1991 or 2001 data. The count for the 'New Zealand European' category decreased in 1996, which 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). For example, van der Pas and Poot (2011) noted that in the 1996 Census, almost 48,000 people identified themselves as Dutch, compared with just 27,866 in 2001 and 29,000 in 2006.

Second, the treatment of responses of 'New Zealander' to the Census ethnicity question has changed over time. In 2001, those who considered themselves simply to be a 'New Zealander' were likely to have been counted in the New Zealand European category. However, in 2006

[28] Refer to Appendix Table A1 for the Level 1 and Level 2 classification of ethnicities in New Zealand.

[29] We ran the analysis also with 'not further defined' dropped, and again with 'not further defined' as a separate category. The differences in results with those reported in this paper are minimal, but available upon request to interested readers.

[30] Specifically, 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 Pākehā' 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). Also, the 2001 question used the word 'Māori' rather than 'NZ Māori' (Statistics New Zealand 2017).

New Zealander was explicitly included as a new category and this change received much publicity in the media. This was no longer a prominent issue by 2013 and the increase in counts for the New Zealand European category from 2006 to 2013 is therefore partly attributable to fewer people identifying themselves as 'New Zealander' by 2013.

We use three different variables in our analysis of economic residential sorting (viz. educational attainment, occupation, and income). For educational attainment, we use the variable 'Highest Qualification' for all years from 1996 onwards.[31] The classifications under this category for 1996 and 2001 are different from those for 2006 and 2013.[32] Due to unavailability of data on the same variable for 1991, we used 'Highest Secondary School Qualification' for 1991.[33] This issue affects our results over time somewhat, but is not expected to have impacted on our conclusions.

In the Census, an 'occupation' is defined as a set of jobs that require an individual (including the self-employed) to perform identical sets of tasks (Statistics New Zealand 2013a). We use the New Zealand Standard Classification of Occupations (NZSCO99),[34] which is a five level hierarchical classification with nine broad major groups (Statistics New Zealand 2015). As in the case of ethnicity, we use both Level 1 and Level 2 occupation levels. From 1991 to 2013, reporting and classification of occupations in the New Zealand Census of Population and Dwellings has changed (Hancock 2015). Since 1996, the group 'Armed forces' has been included under 'Personal and Protective Service Workers'. Therefore, we combined these groups for the calculations in 1991 as well.

---

[31] Highest qualification is derived for people aged 15 years and over, and combines highest secondary school qualification and post-school qualification to obtain a single highest qualification by category of attainment (Statistics New Zealand 2015).

[32] For highest qualification, 2013 and 2006 Census data has limited comparability with 2001 Census data due to the progressive introduction of the National Certificate of Educational Achievement (NCEA) from 2002. NCEA is now the main qualification for secondary school students (Statistics New Zealand 2013a).

[33] This is the highest secondary school qualification gained by category of attainment, and was collected for people aged 15 years and over (Statistics New Zealand 2015).

[34] The Australian and New Zealand Standard Classification of Occupations (ANZSCO) was only introduced in 2006.

Finally, we also use data on total personal income.[35] The number of income intervals and the bounds have changed over the years due to inflation and real income growth.[36] For simplicity, we have not attempted to adjust the data to a common set of intervals. This might affect the year-wise comparability of the sorting values; however it is unlikely to substantially impact the conclusions drawn from the analysis.

For all variables, we aimed to keep the number of groups similar, for better comparability in sorting, as residential sorting measures are sensitive to the number of groups (Mondal et al. forthcoming). For example, 18 groups were used in the analysis of ethnic sorting in 2018 and 16 income groups. In any case, where we measure diversity in an area unit we use the Evenness Index, which corrects for the number of categories in the classification. This is elaborated in the next section.

3.4 Methodology

There are many different measures that can be used as indicators of residential sorting (see e.g. Massey and Denton 1988; Nijkamp and Poot 2015; Reardon and Firebaugh 2002). We choose entropy-based measures of residential sorting and diversity, following the seminal contribution by Theil and Finezza (1971). Entropy measures are conceptually and mathematically attractive and provide the only multigroup index than can be decomposed into a sum of *between*- and *within*-group components (Reardon and Firebaugh 2002). Additionally, we provide the first application of this approach with New Zealand data.

From information theory, we define the (Shannon) entropy of a system X, ($H(X)$), with possible outcomes $x_1, x_2, \ldots x_N$ and $p(x_i)$ the probability of state x_i occurring, as:

$$H(X) = -\sum_{i=1}^{N} p\,(x_i) \ln p(x_i) \quad (1)$$

[35] In the Census, total personal income is collected for people aged 15 years and over, who usually live in New Zealand and are present on census night (including those who state not receiving any income). Total personal income is the before-tax income for the respondent, and is collected as an income range rather than an actual dollar income (Statistics New Zealand 2015).

[36] The detailed year-wise income bands are shown in Appendix Table A4.

Interpreting the fraction of a population that belongs to a certain group as the probability of a randomly selected person belonging to that group, we can define the diversity (entropy) index (E_a) of the population in area a in terms of a given classification as:

$$E_a = -\sum_{g=1}^{G} \frac{P_{ga}}{P_a} \ln \frac{P_{ga}}{P_a} \tag{2}$$

in which P_{ga} is the number of people from group g (= 1, 2, …, G) located in area unit a (= 1, 2, …, A), and P_a is the total number of people in area unit a. We will additionally denote P_g as the number of members of group g in Auckland and P to be the total number of people in Auckland. The minimum of the diversity index is reached when only one of the groups is present, in which case $E_a = 0$.[37] Maximum diversity occurs when all groups are equally represented in area unit a, in which case $E_a = \ln(G)$. Because we are considering classifications that have different numbers of categories, it is convenient to normalise the entropy diversity index to an evenness index I_a that varies between zero and one in all cases (e.g., Nijkamp and Poot 2015):

$$I_a = -\frac{\sum_{g=1}^{G} \frac{P_{ga}}{P_a} \ln \frac{P_{ga}}{P_a}}{\ln(G)} \tag{3}$$

To investigate the geographical differences in diversity across Auckland area units, we calculated the evenness index of each area unit in Auckland for each of the four classifications and use choropleth maps to show the spatial distribution of this diversity measure across Auckland. Following Florida and Mellander (2018), we also averaged the area unit diversities (with equal weights) for the three economic (income, qualification and occupation) variables in each census to create an overall economic diversity measure which can be compared with the cultural diversity measure based on ethnicity.

[37] We define 0*ln(0) = $\lim_{q\to 0}[\,q(\ln(q)] = 0$ to allow calculation of D_a even in the case of there being groups who have zero members in any area at some point in time.

Spatial sorting can be defined as the average extent to which diversity of an area unit differs from that of the city as a whole. Hence if we compare group g with all other groups combined, the entropy of area a (E_{ga}) becomes:

$$E_{ga} = -\frac{P_{ga}}{P_a}\ln\left(\frac{P_{ga}}{P_a}\right) - (1 - \frac{P_{ga}}{P_a})\, ln\left(1 - \frac{P_{ga}}{P_a}\right) \quad (4)$$

while for the city as a whole ($\bar{E}_g$) it is:

$$\bar{E}_g = -\frac{P_g}{P}\, ln\left(\frac{P_g}{P}\right) - (1 - \frac{P_g}{P})\, ln\left(1 - \frac{P_g}{P}\right) \quad (5)$$

A natural measure of spatial sorting / segregation of group g (EIS_g) is then (see e.g. Iceland et al. 2002):

$$EIS_g = \sum_{a=1}^{A}\frac{P_a}{P}\left(1 - \frac{E_{ga}}{\bar{E}_g}\right) \quad (6)$$

which is simply the area-population weighted average of one minus the relative entropy of the areas $\left(\frac{E_{ga}}{\bar{E}_g}\right)$ with respect to group g. This index varies between zero (when the group is distributed proportionally to the total population in all area units) to one (when all areas in which group g is represented contain no other group).

When the composition of a city's population in terms of groups according to a classification (ethnicity, occupation, etc.) changes, it is useful to have an overall measure of residential sorting for the city that accounts for whether segregated groups are becoming more or less important. This overall measure is *Theil's Multi-group Segregation Index H* (Theil 1972; Theil and Finezza 1971; White 1986). To calculate this index we first measure the city-wide entropy (diversity) *(E)* of the classification, i.e. $E = -\sum_{g=1}^{G}\frac{P_g}{P}\ln\frac{P_g}{P}$. We also calculate the ratio $r_{ga} =$

$\frac{P_{ga}}{P_a} / \frac{P_g}{P}$ which measures the extent to which a group is overrepresented ($r_{ga} > 1$) or underrepresented ($r_{ga} < 1$) in area a. Theil's multi-group spatial sorting index (H) is now calculated as follows (e.g. Reardon and Firebaugh 2002, Table 1):

$$H = \frac{1}{E} \sum_{g=1}^{G} \frac{P_g}{P} \sum_{a=1}^{A} \frac{P_a}{P} r_{ga} \ln r_{ga} \tag{7}$$

Essentially, H measures the group-population weighted average of the extent to which the spatial distribution of a group differs from the spatial distribution of the entire population. H varies between zero and one. The index is zero when all areas have the same population composition. The index is one if there is, for each group in the classification, no area in which more than that one group is represented. An alternative way of calculating an overall city index of residential sorting (H^*) is to simply take the group-population weighted average of EIS_g, i.e. to calculate:

$$H^* = \sum_{g=1}^{G} \frac{P_g}{P} EIS_g \tag{8}$$

This calculation gives approximately the same value as *H*, but is easier to interpret. We calculate this measure to investigate whether residential sorting is been changing over time in Auckland. We also use the H^* values calculated for each classification to compare residential sorting by cultural and economic factors in Auckland. Yet another way of calculating *H* is to exploit the property that it measures the relative extent to which the diversity of city as a whole differs from the population-weighted average of the area units' diversity (Theil and Finezza, 1971; White 1986). The diversity of the city is given by *E* as defined previously. We can then calculate *H* also as follows:

$$H = \frac{E - \sum_{a=1}^{A} \frac{P_a}{P} E_a}{E} \tag{9}$$

Finally, following Reardon et al. (2000), we consider the impact of multi-level classification on *Theil's Multi-group Segregation Index H*. Considering different levels of aggregation, we decompose the index values into between-group and within-group components and show how sensitive the sorting index is to the level of aggregation in the classifications. In our case, we consider a classification with two levels (coarse – single digit – and more refined – double digit) for both ethnicity and occupation, as only these two measures have multiple levels of classification that allow for this decomposition.

Specifically, consider that *g* = 1, 2,…*G* indexes the most detailed classification and that *n* = 1, 2, …*N* is an aggregation of these groups into a smaller number of broader groups (i.e. *N* << *G*). Theil's Multi-group Sorting Index values (*H)* can be decomposed into between-group and within-group components for ethnicity and occupation using the following formula:

$$H = \frac{E_N}{E} H_N + \sum_{n=1}^{N} \frac{P_n E_n}{PE} H_n \qquad (10)$$

(see Reardon et al. 2000;Reardon and Firebaugh 2002). Here, *H* is the Theil Index calculated over all groups in the city (Level 2), H_N is the Theil Index calculated among the Level 1 groups, and H_n is the Theil Index calculated *within* each of the Level 1 groups (i.e. *between* the Level 2 groups). E_N is entropy among the supergroups (Level 1), E_n is the entropy *within* Level 1 group *n*, and E is the entropy of the population as a whole (i.e., Level 2). P and P_n are respectively the size of population as a whole and the population size of Level 1 group *n*.

3.5 Results

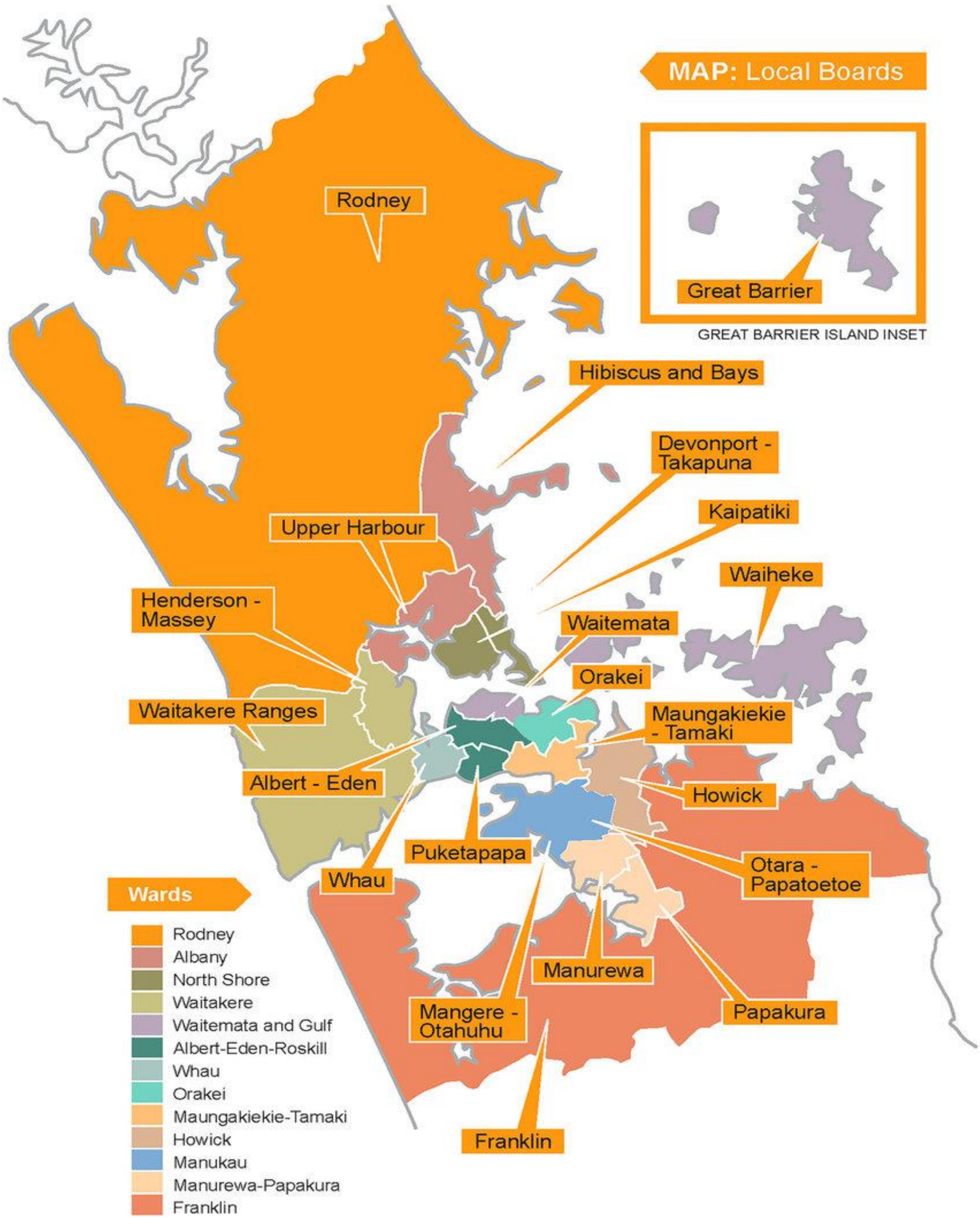
Figure 3.1 shows choropleth maps of the evenness scores of area units in Auckland for each of the variables in 2013. Lower values represent lower levels of diversity and are signalled by lighter colours on the map. For reference, Panel (a) in Figure 3.1 shows a map of the 13 wards (which elect the Auckland Council mayor and 20 Councillors) as well as the 21 local boards (that are concerned with local issues) that make up the Auckland area (Fathimath 2017). The Central Business District is in the Waitamata and Gulf ward. Panel (b) shows that ethnic diversity varies widely across the city. Generally, the central urban area exhibits much greater diversity than the rural fringes. Ethnic diversity is also much greater south of the city centre than north of the city centre and harbour bridge. Central Auckland has two large tertiary institutions along with many language schools and other training institutions, which attract students from overseas, and contribute to high ethnic diversity in central Auckland. Moreover, the two largest contributors to the Skilled Migrant visa category are India and China (Ministry of Business, Innovation and Employment 2016), and their residential location is also relatively clustered (Maré et al. 2016).

The evenness scores for the economic variables are displayed in Panels (c)-(e) in Figure 3.1. These use the same legend as Panel (b). In terms of diversity across qualification groups, there is less diversity in south Auckland and in those parts of the city centre where students dominate. Qualifications and ethnicity evenness values are actually negatively correlated (r = -0.35). In contrast, evenness scores of ethnicity are positively correlated (r = 0.37) with those of occupation shown in Figure 3.1 (d). On the whole, the map for occupations shows less spatial contrast than for education and a lower average evenness score across area units. The lower occupational evenness score in the CBD reflects the dominance of the services sector there. The map for income (Figure 3.1(e)) shows a spatially dominant high level of evenness. This simply reflects a fairly even distribution everywhere of the population across the income categories [38] but it is not necessarily indicative of low income inequality. We do observe lower evenness where students live and in the south of Auckland. Income and ethnicity evenness values are negatively correlated (r = -0.20). In general, it is clear that there are more spatial differences in Auckland in terms of cultural diversity than in terms of any of the economic variables.

[38] The income categories are listed in Appendix Table A4.

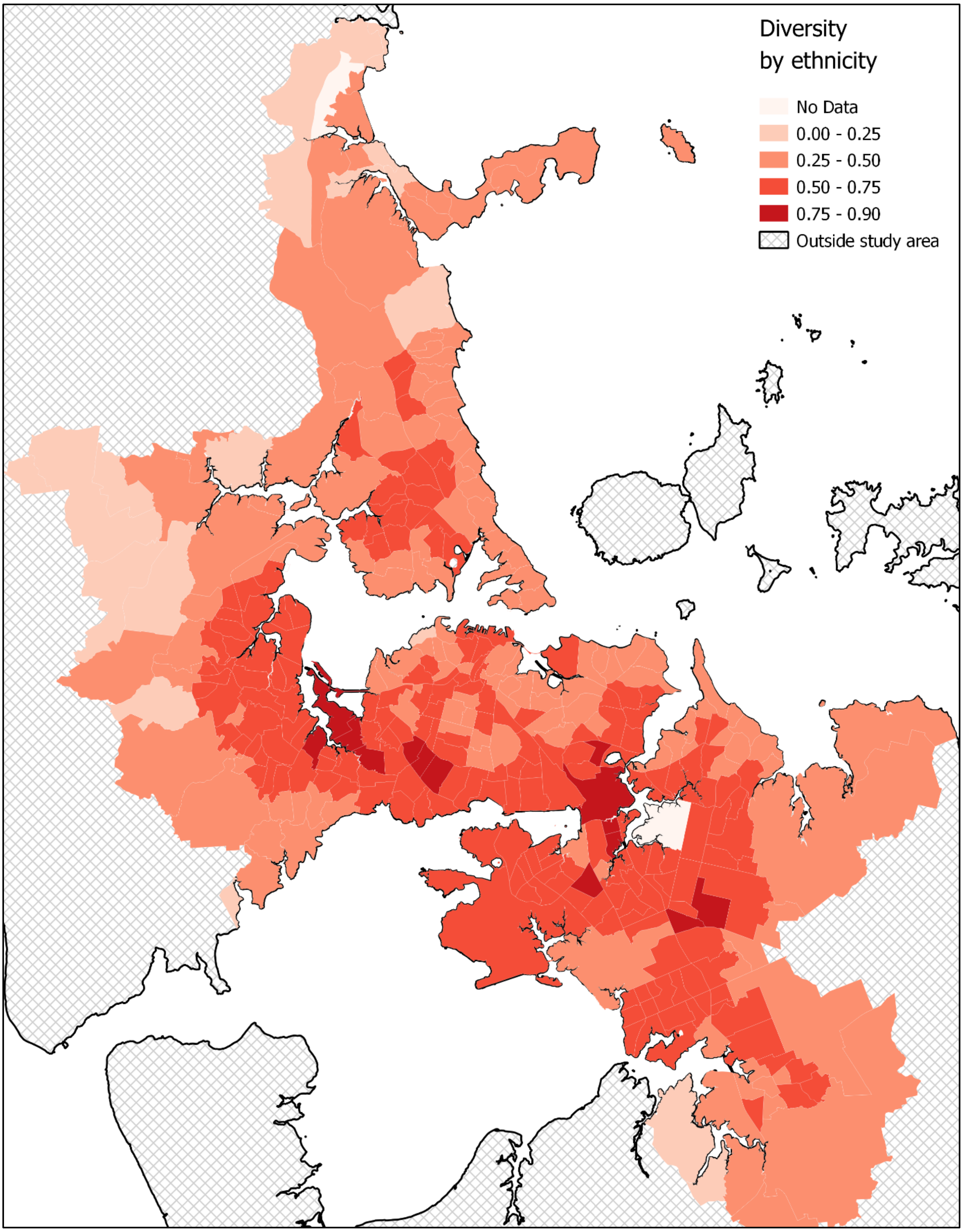
Figure 3.1: Diversity in Auckland by Cultural and Economic Variables, 2013

(a) Auckland Wards and Local Boards

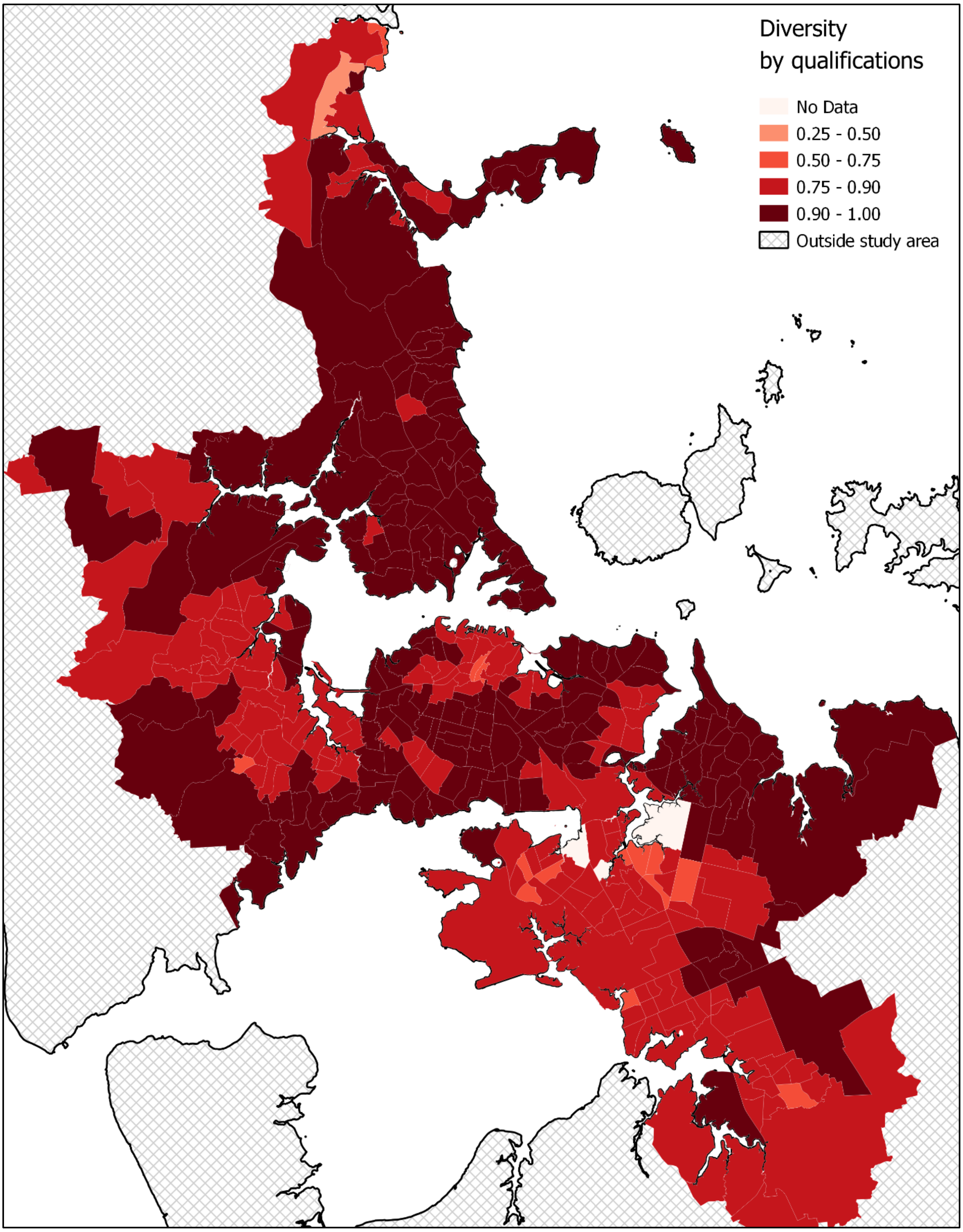


Source: Fathimath (2017)

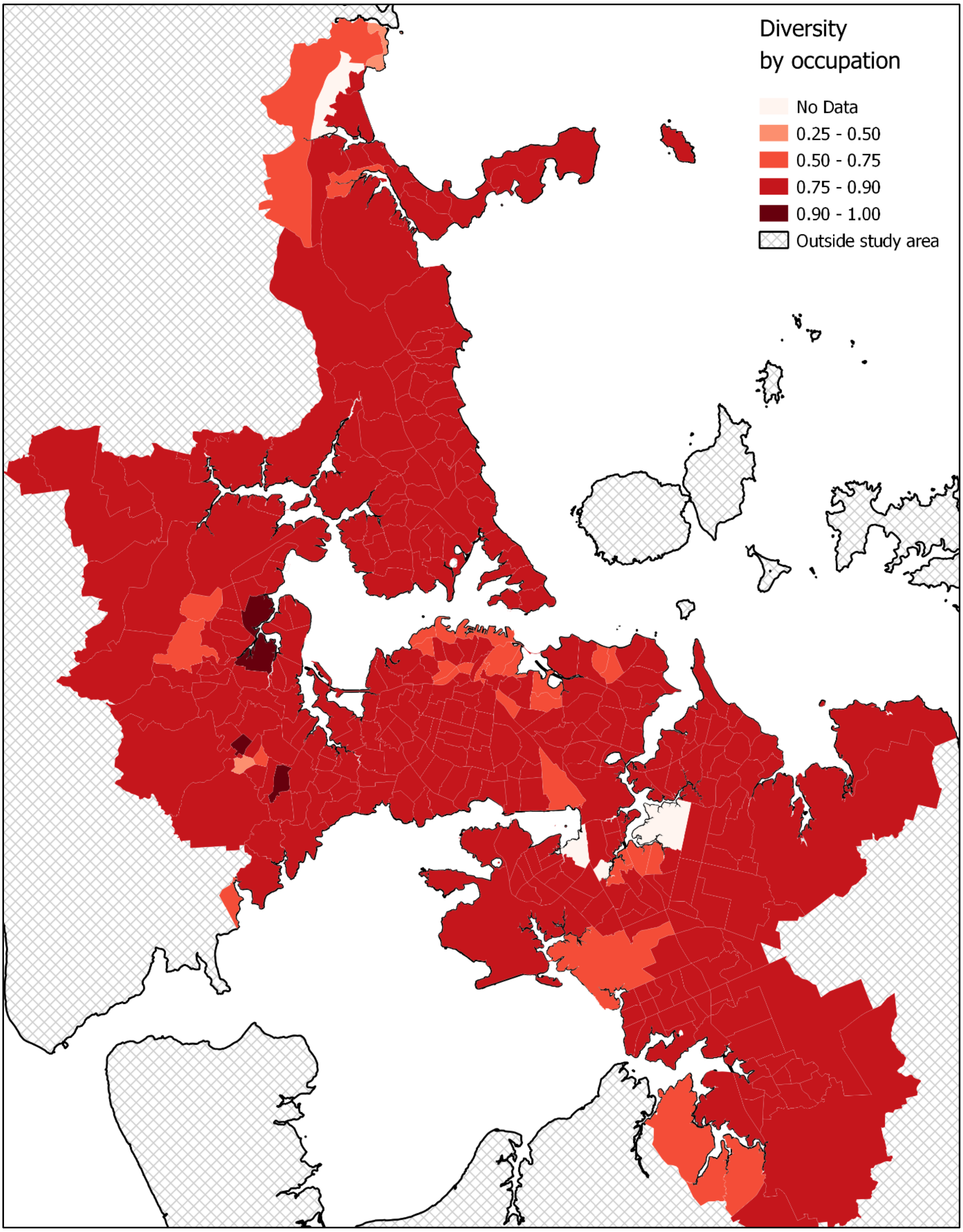
(b) Evenness Index - Ethnicity



(c) Evenness Index – Qualification



(d) Evenness Index - Occupation



(e) Evenness Index - Income

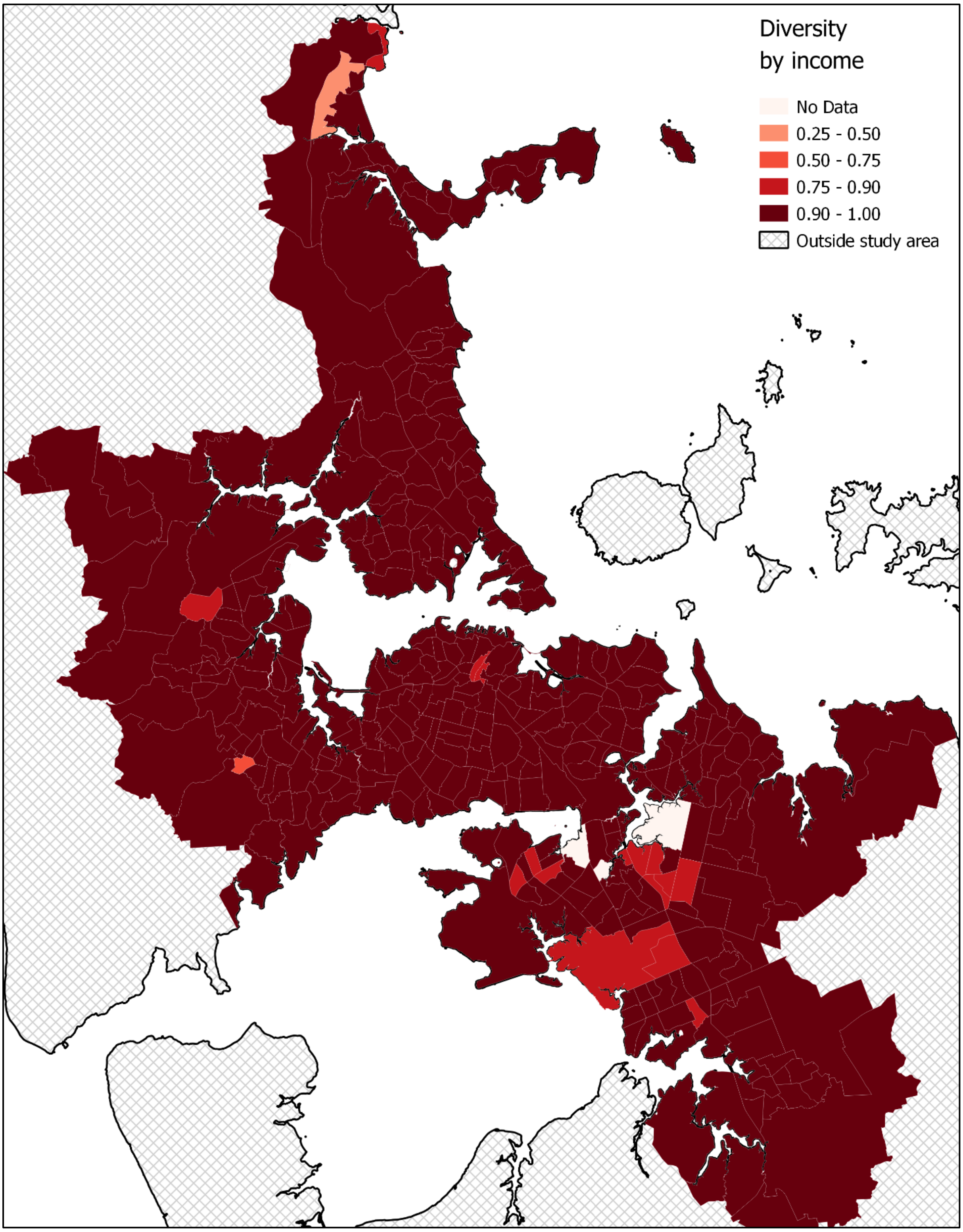
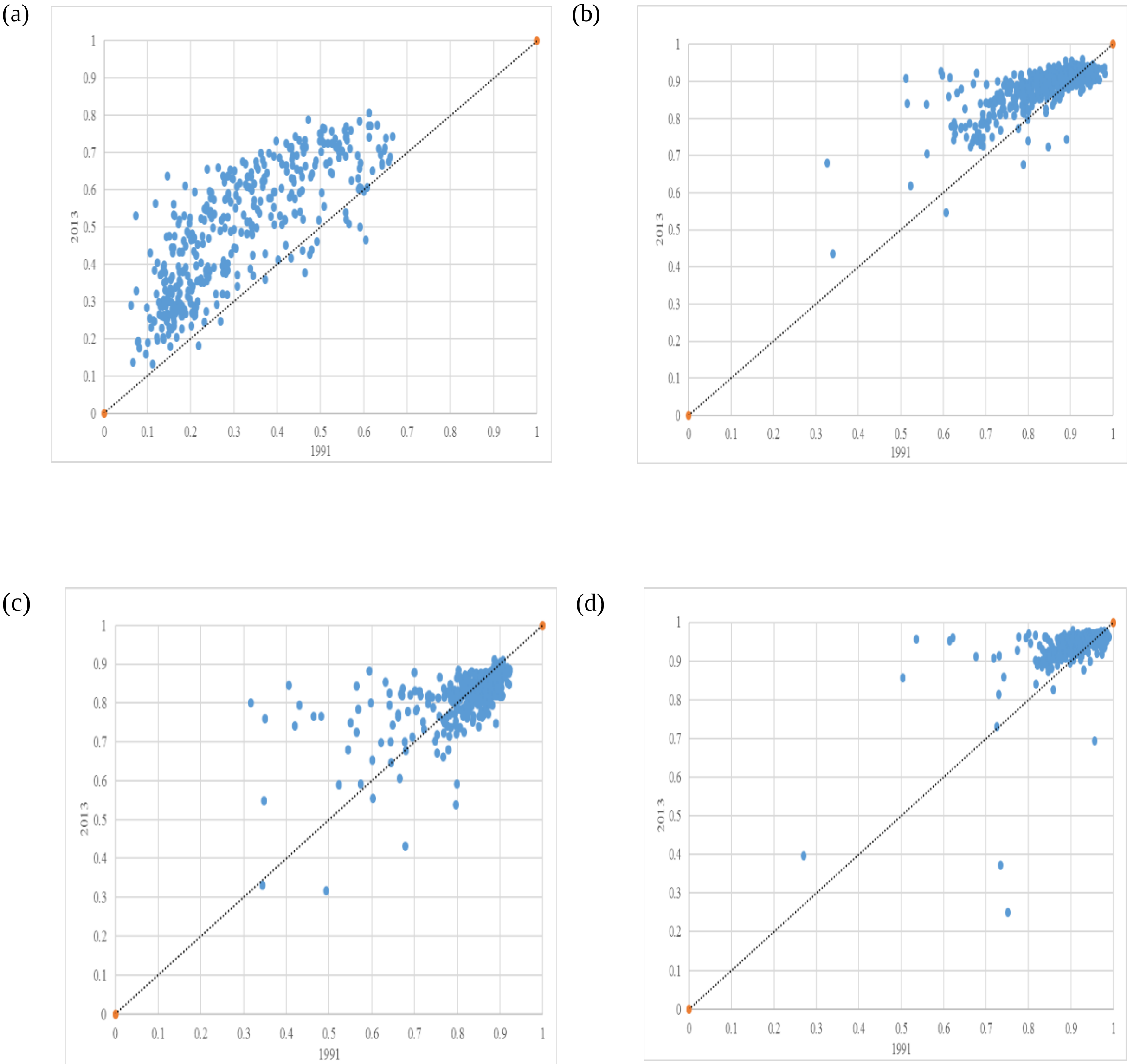


Figure 3.2 shows the relationship between 1991 and 2013 values of the evenness measure of diversity in each area unit of Auckland for each of the variables, where each dot represents one area unit. In the Figures, almost all observations for all the variables, except for occupation, lie above the 45-degree line. This means that for most area units in Auckland, diversity has increased between 1991 and 2013, except for occupation (Panel (c)). For occupation, area units appear roughly equally split between those that had increasing diversity and those that had decreasing diversity.

Figure 3.2: Cultural and Economic Diversity of Auckland Area Units: A comparison between 1991 and 2013



Notes: (a) Ethnicity; (b) Qualification (c); Occupation; (d) Income. Diversity is measured by means of the Evenness Index.

Table 3.2 reports the Auckland-wide evenness indexes over the period 1991 to 2013. This shows that Auckland has generally become more economically and culturally diverse. For all of the variables except occupation, diversity has increased over this period.[39] In the case of occupation, the downward trend in evenness can be attributed to the growing dominance of services and related occupations in New Zealand (Statistics New Zealand 2017).

Table 3.2: Auckland-wide levels of the evenness index of group counts within the classifications, 1991-2013

Year	Ethnicity	Qualification	Occupation	Income	Economic
1991	0.402	0.878	0.896	0.949	0.908
1996	0.465	0.848	0.893	0.922	0.888
2001	0.505	0.884	0.900	0.949	0.911
2006	0.647	0.894	0.892	0.969	0.918
2013	0.584	0.922	0.871	0.970	0.921

Note: We calculate Auckland-wide Evenness Indexes (I) for each classification in each census years, $I = E / \ln(G)$, where G is the number of groups in a classification. See also Section 3.3.

Sorting by ethnicity

Table 3.3 shows the Entropy Index of Sorting values for ethnicity in Auckland in 1991, 1996, 2001, 2006 and 2013. We observe that for all Level 2 ethnic groups within the Pacific Island broad ethnic group, along with the Chinese and Indian ethnic groups, there has been an increase in spatial sorting. These groups appear to show Schelling-type behaviour in that they appear to be increasingly seeking to live with their co-ethnics (Schelling 1971). In Auckland, groups of Chinese are clustered in the wealthier suburbs, but most are concentrated in middle-priced suburbs. The Indian population is also observed to have major concentrations in these areas. A large number of Asian students are concentrated in central Auckland, which is near the largest tertiary institutions (Ho 2015). Friesen (2008) also found a significant level of clustering among the Asian population in Auckland. A ‘zone of familiarity’, including provision of ethnic goods and services and employment in ethnic businesses run by co-ethnics may contribute to this outcome. Poulsen et al. (2004) found that despite policies promoting multiculturalism in New Zealand, many among the Chinese or Indian ethnic groups choose to maximise their economic

[39] This number has declined from 2006 to 2013 because the number of people calling themselves New Zealander declined for New Zealand as a whole from 430,000 in 2006 to just under 66,000 in 2013 (Statistics New Zealand 2013a).

success by being involved in small businesses serving their own community and thus reside in neighbourhoods with a larger proportion of their ethnicity. In contrast to the Chinese, Indian and Pacific groups, Table 3.3 shows that for the New Zealand European, South-East Asian, and all of the Level 2 ethnic groups within the MELAA broad ethnic group, residential sorting has declined over time.

Table 3.3: Entropy Index of Sorting (EIS) by ethnicity: Auckland, 1991-2013

	1991		1996		2001		2006		2013	
Ethnicity	Share of Total Responses (%)	EIS	Share of Total Responses (%)	EIS	Share of Total Responses (%)	EIS	Share of Total Responses (%)	EIS	Share of Total Responses (%)	EIS
New Zealand European	70.6	0.145	35.8	0.020	62.6	0.131	49.2	0.095	56.8	0.128
Other European	6.2	0.018	47.6	0.023	5.14	0.020	5.42	0.024	5.78	0.026
NZ Maori	10.5	0.097	6.90	0.078	11.0	0.075	10.7	0.071	8.43	0.061
Samoan	5.13	0.196	3.39	0.196	5.96	0.188	6.3	0.189	5.39	0.199
Cook Island Maori	2.15	0.175	1.39	0.170	2.45	0.156	2.56	0.155	2.16	0.189
Tongan	1.53	0.165	1.18	0.179	2.36	0.178	2.77	0.184	2.35	0.202
Niuean	1.15	0.152	0.75	0.152	1.30	0.139	1.31	0.134	1.08	0.159
Tokelauan	0.06	0.239	0.04	0.225	0.085	0.198	0.107	0.176	0.07	0.261
Fijian	0.19	0.114	0.208	0.075	0.259	0.102	0.352	0.078	0.35	0.115
Other Pacific Island	0.04	0.247	0.076	0.161	0.178	0.159	0.259	0.157	0.19	0.197
Southeast Asian	0.22	0.174	0.43	0.091	0.95	0.081	1.44	0.073	1.73	0.081
Chinese	1.2	0.060	1.54	0.063	3.86	0.110	5.44	0.120	6.27	0.124
Indian	0.89	0.087	1.11	0.085	2.38	0.096	3.55	0.109	5.41	0.136
Other Asian	0.03	0.266	0.15	0.110	1.02	0.109	1.73	0.101	1.96	0.096
Middle Eastern	0.04	0.231	0.078	0.113	0.368	0.094	0.623	0.099	0.60	0.131
Latin American/Hispanic	0.004	0.368	0.013	0.233	0.039	0.216	0.066	0.166	0.09	0.198
African	0.006	0.346	0.012	0.250	0.069	0.211	0.175	0.181	0.15	0.203
Others	<0.001	0.405	0.003	0.305	<0.001	0.399	8.08	0.033	1.2	0.031

Sorting by other variables

In line with the geographically-based results reported in Figure 3.1, residential sorting by qualification, occupation, and income is much less apparent than for ethnicity.[40] We find that the greatest residential sorting is exhibited by people with high education and high income.[41] These results are consistent with previous research (Maré et al. 2011; Maré et al. 2012). Maré et al. (2011) found prominent patterns of concentration of residents with high income in specific regions of Auckland, but less distinct patterns for the low and middle income groups.

It can be easily shown by regression that there is a small negative effect of group size on the level of sorting. With respect to occupation, this can be easily illustrated by the 'Legislators and Administrators' group, which had the highest level of sorting in 1991 (*EIS* = 0.361), but only accounted for 0.013% of the Auckland labour force. The next three groups with the highest levels of residential sorting in 1991 (with their 1991 labour force share in parentheses) are: 'Salespersons, Demonstrators & Models' (5.65%), 'Drivers and Mobile Machinery Operators' (2.9%) and 'Other Craft & Related Trades Workers' (1.62%). All of these experienced a notable decline in residential sorting between 1991 and 2013. At the other end of the scale, the four occupational groups with the lowest residential sorting in 1991 were: 'Office Clerks' (12.7%), 'Other Professionals' (3.54%), 'Personal & Protective Services Workers' (7.1%) and 'Other Associate Professionals' (7.86%). The groups with the next lowest EIS, 'Market Orientated Agricultural & Fishery Workers' (three percent) experienced a huge 1991-2013 increase in sorting (from 0.016 to 0.171), indicative of the expansion of residential land at the cost of land used for market gardening.

In terms of overall residential sorting, Theil's Multi-group Segregation Index (*H*) (see Table 3.4) shows a decline in ethnic residential sorting between 1991 and 2013 (the low values in 1996 and 2006 are partially due to the census question issues discussed earlier). Table 3.4 also shows that the residential sorting by income has remained constant over time at this very broad level. However, this does not imply that there are no change in the distribution of 'poor' and 'rich' areas. The index does not inform on levels of income, but simply on the spread across the census questionnaire income intervals. We observe[42] a notable increase in residential

[40] See Appendix Tables A2, A3 and A4 respectively.
[41] See Appendix Tables A2 and A4.
[42] See Appendix Table A4.

sorting of those in the 'open-ended' highest income category, with EIS increasing from 0.119 in 1991 to 0.135 in 2013, suggesting that the rich are less evenly spread spatially than they used to be.

Residential sorting by occupation shows a downward trend from 1991 to 2006, with a slight increase subsequently.[43] This might be due to a number of factors. The female labour force participation rate has increased in New Zealand (from 54.3 percent in 1991 to 64.5 percent in 2006) (Statistics New Zealand 2017). While there is gender segregation in employment by occupation, occupational segregation has declined and there has been a structural transformation in employment towards employment in services. Consequently, whereas there were historically 'blue collar' (male employment dominated) and 'white collar' area units, that distinction has become less over time (e.g. van Mourik et al., 1989) – leading to lower spatial sorting by occupation. For qualification, the residential sorting trend is also downward (the 1991 value is not directly comparable due to a changing classification).[44]

Table 3.4: Theil's Multi-Group Index of residential sorting by ethnicity and socioeconomic variables: Auckland, 1991-2013

Year	Ethnicity	Qualification	Occupation	Income	Economic
1991	0.135	0.028	0.035	0.015	0.026
1996	0.039	0.036	0.031	0.015	0.028
2001	0.122	0.029	0.027	0.015	0.023
2006	0.096	0.028	0.024	0.015	0.023
2013	0.122	0.025	0.025	0.015	0.022

Note: The Economic Index of residential sorting is the simple average of the Theil Multi-group Index for qualification, occupation, and income.

Comparing the Theil Multi-Group Index across the four chosen characteristics - ethnicity, qualification, occupation and income - we see that the greatest degree of residential sorting occurs by ethnicity. Among the economic variables, residential sorting is greatest by occupation. Again, the lack of residential sorting by income might seem surprising. However, previous research for New Zealand (Maré et al. 2011; Maré et al. 2012) has also found low residential sorting by income. New Zealand has certainly had historically low levels of spatial

[43] See Appendix Table A3

[44] See Appendix Table A2.

income inequality. Moreover, the use of personal income instead of household income could also have contributed to low measured sorting by income. Maré et al. (2012) calculated income sorting by personal income as well as household income and found that the sorting values were slightly higher for household income.

Taking the average of the Theil Multi-Group Indexes for the economic variables, following Florida and Mellander (2018), we see from the final column of Table 3.4 that, firstly, spatial sorting in Auckland is less in economic terms than in cultural terms and, secondly, that ethnic and economic spatial sorting levels are less in 2013 than in 1991.

Finally, we show how sensitive the Theil's multi-group measure of sorting is to the level of aggregation, by decomposing the *H* values into between-group and within-group components.[45] We do this for our cultural variable (ethnicity) and one economic variable (occupation). The results for ethnicity are reported in Table 3.5(a), and the corresponding analysis for occupation is reported in Table 3.5(b). The 'Theil at level 2' column of Table 3.5(a) repeats the index values already reported in the 'Ethnicity' column of Table 3.4. The second and third columns decompose the first column into a share of sorting that occurs between Level 1 ethnic groups, and a share that occurs within Level 1 groups (i.e. between Level 2 groups within the same Level 1 group), as shown in Eq. (9). The fourth and fifth columns show the percentage shares of between- and within-group sorting. In terms of shares, the results imply that the co-location of Level 1 ethnic groups (e.g. Pacific Islanders) has been decreasing over time, but that sorting *between* Level 2 ethnic sub-groups (e.g. Samoan, Cook Island Māori, Tongan, etc.) within their Level 1 groups has increased in importance. In other words, Level 2 ethnic groups are increasingly sorting away from other Level 2 groups within the same Level 1 broad ethnic group. For instance, there are fewer suburbs that are generic Pacific Island communities, with Samoan, Tongan and other Pacific groups increasingly located separately from each other.

[45] The data used for the Level 1 calculation has been constructed from the Level 2 data sheets (using a bottom-up approach), so that the total population count at both levels are the same.

Table 3.5: Decomposition of Theil's Multi-Group Index of residential sorting by ethnicity and occupation: Auckland, 1991-2013

(a) Decomposition of Theil's Multi-Group Index of residential sorting- Ethnicity

	Theil at level 2	Between level 1 groups	Within level 1 groups	Between Level 1 groups proportion (%)	Within Level 1 groups proportion (%)
1991	0.135	0.054	0.080	40.3	59.7
1996	0.039	0.031	0.008	78.9	21.1
2001	0.122	0.066	0.056	53.8	46.2
2006	0.096	0.038	0.057	40.1	59.9
2013	0.122	0.044	0.078	36.0	64.0

(b) Decomposition of Theil's Multi-Group Index of residential sorting - Occupation

	Theil at level 2	Between Level 1 groups	Within Level 1 groups	Between Level 1 groups proportion (%)	Within Level 1 groups proportion (%)
1991	0.035	0.012	0.023	33.2	66.8
1996	0.031	0.011	0.020	35.8	64.2
2001	0.027	0.011	0.016	40.7	59.3
2006	0.024	0.010	0.015	40.6	59.4
2013	0.025	0.007	0.017	29.0	71.0

Table 3.5(b) repeats the analysis for occupation. In this case, the sorting is higher within Level 1 groups (i.e. between Level 2 groups within each Level 2 group) in all years, and segregation has generally been declining between Level 1 groups and between Level 2 groups within a Level 1 group. The notable exception is the increase in the 'within Level 1' component between 2006 and 2013, leading also to an increase in the Level 2 share over that period. This suggests that while there was a trend of enclaves of people of similar occupations within a higher level occupational grouping co-locating less, this trend reversed after 2006. Analysis with 2018 data may reveal whether this reversal is one-off or indicative of longer-term underlying phenomena that lead again to co-location.

3.6 Conclusion

We applied entropy-based measures of residential sorting and diversity to census data for Auckland over the period from 1991 to 2013. We find that, broadly speaking, residential sorting by ethnicity, qualification and occupation declined over this period, whereas sorting by income remained fairly constant. Calculations with the Theil Multi-group Index reinforced that both cultural and economic residential sorting in Auckland declined over this period.

One of the research questions in this paper was to identify whether individuals exhibit the greatest level of residential sorting by their cultural or by economic characteristics. We considered ethnicity as our cultural variable. We formed our economic index of residential sorting as a combination of income, qualification and occupation, which – as stated by Florida and Mellander (2018) – captures the mutually reinforcing aspects of income, qualification and occupational sorting in a better way than they do individually. We find that residential sorting is greater by cultural factors (ethnicity) than by economic factors (income, qualification and occupation), separately as well as combined.

This result might seem surprising, given that we can imagine enclaves of privilege or relative deprivation. Why then, do the data not support this? Part of the reason is likely to be our chosen level of geographical aggregation. In urban areas, an area unit is approximately the size of a suburb, with an average population of about 2000. If we were to complete our analysis at a lower level of geographical aggregation (e.g. meshblocks, which are roughly neighbourhoods or city blocks), we might observe more residential sorting by these other characteristics. However, small cell sizes would become problematic when conducting this analysis across many groups and many small geographical areas, leading to a greater degree of necessary suppression of data counts (Statistics New Zealand requires this due to concerns about confidentiality of data). This explains why previous analyses that have used meshblock-level data (e.g. Maré et al. 2011), have used more aggregated ethnic or other groups. Our analyses should be seen as complementary to that earlier work. Moreover, the lack of prominent pattern of income sorting might also be due to the use of total personal income and not household income, which might play a role in household location decisions (Maré et al. 2011).

From the decomposition results, we find that individuals are increasingly tending to co-locate more according to their finer ethnic groups than their broad ethnic groups. The finer ethnic groups are not co-locating together with other groups within the same broad ethnic group, i.e. there is spatial heterogeneity of the finer ethnic groups. For example, the Tokelauans and the Niueans co-locate more with their own-group members now, but they do not tend to co-locate with other groups under the broad Pacific group. This can create both problems as well as opportunities for public services (Caldwell et al. 2017). Thus, it is becoming increasingly important to look at residential sorting at a much finer scale.

Our findings contribute to the extant literature on residential sorting in a number of ways. First, our interpretation is based on the results from entropy-based measures, which is new in New

Zealand. We strongly recommend the use of entropy-based measures in future research, as along with many desirable properties, they are least sensitive to group-size (Mondal et. al forthcoming). Second, this is among the first studies to consider residential sorting within and between broad ethnic groups. This is important because the broad (Level 1) ethnic groups are very heterogeneous and may not represent the characteristics and choices of their component (Level 2) groups. For instance, the ‘Asian’ broad ethnic group includes diverse Level 2 groups such as Southeast Asian, Chinese and Indian ethnicities. An argument could be made that even the Level 2 groups are too heterogeneous (e.g. Southeast Asian), and that Level three groups (Thai, Vietnamese, etc.) would be an improvement. We leave that as an exercise for future research. Previous studies in New Zealand have found that the Pacific group tends to co-locate with its own group members the most (Johnston et al. 2011; Maré et al. 2012). However, using finer-grained (Level 2) ethnic groups we observe that although the Level 2 ethnic groups under the broad Pacific group are also highly residentially segregated, residential sorting is also relatively high among those in the MELAA group. That the conclusions change depending on the level of analysis demonstrates the importance of considering the appropriate level of ethnic aggregation. Finally, this paper is one of only a few that include occupation in studying residential sorting in New Zealand.

This study can be extended in a number of ways. In addition to using even more finer-grained ethnic groups, more complex patterns and trends in residential sorting can be identified by combining cultural and socio-economic variables through cross-tabulated groups (e.g. ethnicity-income, ethnicity-qualification etc.). Though we find a less pronounced pattern of residential sorting by occupation, education and income than for ethnicity, further investigation of other socio-economic variables, as well as of other cultural variables like language and religion, offers also fruitful avenues for future research. Additionally, when looking at residential sorting by occupation, we only looked at individuals who are employed, and not at those who are not in the work force because they are unemployed, fulltime carers or retired. Given the ageing of the population, the study of residential (re)location of older couples and individuals is of growing importance. Moreover, rather than taking a descriptive approach there is also much scope for in-depth regression modelling of residential location, as previously explored by Maré and Coleman (2011). Finally, the consequences of current and future trends in residential sorting for individual wellbeing and local social capital are also a demanding but important topic for further investigation

References

Alimi O, Maré DC, Poot J (2016) Income inequality in New Zealand regions. In: P. Spoonley (ed) Rebooting the Regions. Massey University Press, Auckland, pp 177–212.

Alimi OB, Maré DC, Poot J (2018) More pensioners, less income inequality? The impact of changing age composition on inequality in big cities and elsewhere. N Z Popul Rev 44: 49-84.

Auckland Council. (2018) Auckland plan 2050, evidence report. https://www.aucklandcouncil.govt.nz/plans-projects-policies-reports-bylaws/our-plans-strategies/auckland-plan/about-the-auckland-plan/Evidence%20reports%20documents/evidence-report-developing-the-plan.pdf. Accessed 8 August 2019.

Balakrishnan TR, Maxim P, Jurdi R. (2005) Residential segregation and socio-economic integration of visible minorities in Canada. Migr Lett 2:126–143.

Bennett PR (2011) The relationship between neighbourhood racial concentration and verbal ability: an investigation using the institutional resources model. Soc Sci Res 40: 1124–1141.

Betancourt H, López SR (1993) The study of culture, ethnicity, and race in American psychology. Am Psychol 48: 629–637.

Caldwell JT, Ford CL, Wallace SP, Wang MC, Takahashi LM (2017) Racial and ethnic residential segregation and access to health care in rural areas. Health Place 43: 104–112.

Cameron MP, Poot J (2019) Towards Superdiverse Aotearoa: Dimensions of past and future ethnic diversity in New Zealand and its regions. N Z Popul Rev 45: 18-45.

Dalziel P (2013) Education and skills. In: Rashbrooke M (ed) Inequality: a New Zealand Crisis. Bridget Williams Books, Wellington, pp 184-194.

Denton NA, Massey DS (1988) Residential segregation of Blacks, Hispanics and Asians by socioeconomic status and generation. Soc Sci Quart 69: 797–817.

Domina T (2006) Brain drain and brain gain: rising educational segregation in the United States, 1940–2000. City Community 5:387–407.

Duncan OD, Duncan B (1955) Residential distribution and occupational stratification. Am J Sociol 60: 493–503.

Ellis M, Wright R, Parks V (2004). Work together, live apart? Geographies of racial and ethnic segregation at home and at work. Ann Assoc Am Geogr 94:620–37.

Evans L, Grimes A, Wilkinson B, Teece D (1996) Economic reform in New Zealand 1984–95: The pursuit of efficiency. J Econ Lit 34(4): 1856–1902.

Farley R (1977) Residential segregation in urbanized areas of the United States in 1970: an analysis of social class and racial differences. Demography 14: 497–518.

Fathimath A (2017): Impact of municipal amalgamation on stakeholder collaboration: the case of Auckland, New Zealand, Kōtuitui N Z J Soc Sci Online. https ://doi.org/10.1080/1177083x.2017.1367313

Fischer MJ (2003) The relative importance of income and race in determining residential outcomes in U.S. urban areas, 1970-2000. Urban Aff Rev 38: 669–696.

Florida R, Mellander C (2018) The geography of economic segregation. Soc Sci 7: 123.

Fong E, Hou F (2009) Residential patterns across generations of new immigrant groups. Sociol Perspect 52:409–428.

Forrest J, Poulsen M, Johnston R (2006) Peoples and spaces in a multicultural nation: cultural group segregation in metropolitan Australia. Espace-Populations-Sociétés 1:151–164.

Fowler CS, Lee BA, Matthews SA (2016) The contributions of places to metropolitan ethnoracial diversity and segregation: Decomposing change across space and time. Demography 53: 1955–1977.

Friesen W (2008) Diverse Auckland: the face of New Zealand in the 21st Century. Asia New Zealand Foundation, Wellington.

https://www.asianz.org.nz/assets/Uploads/4820390270/Diverse-Auckland-the-face-of-New-Zealand-in-the-21st-Century.pdf. Accessed 17 January 2019

Grodsky E, Pager D (2001) The structure of disadvantage: individual and occupational determinants of the Black-White wage gap. Am Sociol Rev 66: 542–567.

Hall M, Iceland J, Yi Y (2019) Racial Separation at Home and Work: Segregation in Residential and Workplace Settings. Popul Res Policy Rev 38:671–694.

Halpern-Felsher B, Connell J, Spencer M, Aber JL, Duncan G, Clifford E, Crichlow W, Usinger P, Cole S, Allen L, Seidman E (1997) Neighborhood and family factors predicting educational risk and attainment in African American and White children and adolescents. In: Brooks-Gunn J, Duncan G, Aber JL (eds) Neighborhood Poverty: Context and Consequences for Children. Russell Sage Foundation, New York, pp. 146-173.

Hancock A (2015) Where have they gone? changes in occupations using 1991-2013 New Zealand census data. Statistics New Zealand, Wellington. https://ojs.victoria.ac.nz/LEW/article/view/2217. Accessed 12 January 2019

Ho E (2015) The changing face of Asian peoples in New Zealand. N Z Popul Rev 41: 95–118.

Iceland J, Weinberg DH, Steinmetz E (2002) Racial and Ethnic Residential Segregation in the United States 1980-2000, appendix B. US Census Bureau, Washington DC.

Johnston R, Poulsen M, Forrest J (2002) Rethinking the analysis of ethnic residential patterns: segregation, isolation, or concentration thresholds in Auckland, New Zealand? Geogr Anal 34: 245–261.

Johnston R, Poulsen M, Forrest J (2004) The comparative study of ethnic residential segregation in the USA, 1980–2000. Tijdschr Econ Soc Ge 95: 550–569.

Johnston R, Poulsen M, Forrest J (2005) Ethnic residential segregation across an urban system: the Maori in New Zealand, 1991–2001. Prof Geogr 57: 115–129.

Johnston R., Poulsen M, Forrest J (2007) The geography of ethnic residential segregation: a comparative study of five countries. Ann Assoc Am Geogr 97: 713–738.

Johnston R, Poulsen M, Forrest J (2008) Asians, Pacific Islanders, and ethnoburbs in Auckland, New Zealand. Geogr Rev 98: 214–241.

Johnston R, Poulsen M, Forrest J (2011) Evaluating changing residential segregation in Auckland, New Zealand, using spatial statistics. Tijdschr Econ Soc Ge 102: 1–23.

Johnston R, Forrest J, Jones K, Manley D (2016) The scale of segregation: ancestral groups in Sydney, 2011. Urban Geogr 37(7): 985-1008.

Kukutai T (2008) Ethnic self-prioritisation of dual and multi-ethnic youth in New Zealand. Statistics New Zealand, Wellington

Lee BA, Farrell CR, Reardon SF, Matthews SA (2019) From census tracts to local environments: an egocentric approach to neighborhood racial change. Spat Demography 7:1–26.

Lichter D T, Parisi D, Taquino M C (2015). Toward a new macro-segregation? Decomposing segregation within and between metropolitan cities and suburbs. Am Sociol Rev 80:843–873.

Manley D, Johnston R, Jones K, Owen D (2015). Macro, meso and microscale segregation: Modeling changing ethnic residential patterns in Auckland, New Zealand, 2001–2013. Ann Assoc Am Geogr 105: 951–967.

Manley D, Johnston R, Jones K (2019) Decomposing multi-level ethnic segregation in Auckland, New Zealand, 2001–2013: Segregation intensity for multiple groups at multiple scales. Tijdschr Econ Soc Ge 110(3): 319–338.

Maré DC, Coleman A (2011) Estimating the determinants of population location in Auckland. Motu Economic and Public Policy Research, Wellington. https://papers.ssrn.com/abstract=1861531. Accessed 17 January 2019

Maré DC, Coleman A, Pinkerton R (2011) Patterns of population location in Auckland. Motu Economic and Public Policy Research, Wellington. https://papers.ssrn.com/abstract=1860872. Accessed 1 December 2018

Maré DC, Pinkerton RM, Poot J (2016) Residential assimilation of immigrants: a cohort approach. Migr Stud 4: 373–401.

Maré DC, Pinkerton RM, Poot J, Coleman A (2012) Residential sorting across Auckland neighbourhoods. N Z Popul Rev 38: 23–54.

Massey DS (1979) Effects of socioeconomic factors on the residential segregation of blacks and Spanish Americans in U.S. urbanized areas. Am Sociol Rev 44: 1015–1022.

Massey DS, Denton NA (1993) American Apartheid: Segregation and the Making of the Underclass. Harvard University Press, Cambridge.

Massey DS, Denton NA (1988) The dimensions of racial segregation. Soc Forces 67: 281-315.

Ministry of Business, Innovation and Employment (2016). Migration and labour force trends: Auckland overview 2015. New Zealand Immigration, Wellington. https://www.mbie.govt.nz/info-services/immigration/migration-research-and-evaluation/migration-and-labour-force-trends/document-and-image-library/INZ_Migrant.pdf. Accessed August 17 2018

Mondal, M., Cameron, M.P., & Poot, J. (forthcoming). Group-size bias in the measurement of residential sorting. In R. Patuelli & S. Suzuki (Eds.), *A Broad View of Regional Science - Essays in Honor of Peter Nijkamp.* Springer Nature.

Musterd S (2005) Social and ethnic segregation in Europe: levels, causes, and effects. J Urb Aff, 27(3): 331-348.

New Zealand Ministry for Culture and Heritage (2015). Te Ara the Encyclopedia of New Zealand. https://teara.govt.nz/en/population-change/print. Accessed 7 August 2019.

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.

Pool I (1991) Te Iwi Maori. Oxford University Press, Oxford.

Poot J, Pawar S (2013) Is demography destiny? Urban population change and economic vitality of future cities. J Urb Manag 2(1): 5-23.

Poulsen M, Johnston J, Forrest J (2004) Is Sydney a divided city ethnically? Aust Geogr Stud 42:356–77.

Reardon SF, Firebaugh G (2002) Measures of multigroup segregation. Sociol Methodol 32: 33–67.

Reardon SF, Yun JT, Eitle TM (2000) The changing structure of school segregation: measurement and evidence of multiracial metropolitan-area school segregation, 1989-1995. Demography 37: 351–364.

Schelling TC (1971) Dynamic models of segregation. J Math Sociol 1: 143–186.

Sharkey P (2013) Stuck in place: urban neighborhoods and the end of progress toward racial equality. University of Chicago Press, Chicago.

Sims, M. (1999) High-Status Residential Segregation among Racial and Ethnic Groups in Five Metro Areas, 1980–1990. Soc Sci Quart 80:556–73.

Simkus AA (1978) Residential segregation by occupation and race in ten urbanized areas, 1950-1970. Am Sociol Rev 43: 81–93.

Spoonley P (2014) Superdiversity, social cohesion, and economic benefits. IZA World of Labor. https://wol.iza.org/articles/superdiversity-social-cohesion-and-economic-benefits/long. Accessed November 15 2018

Statistics New Zealand (2004). Ethnic population projections: issues and trends. 9–14. http://archive.stats.govt.nz/browse_for_stats/population/estimates_and_projections/ethnic-pop-projections-issues-and-trends.aspx. Accessed 8 August 2019.

Statistics New Zealand (2006) Ethnic groups in New Zealand. http://archive.stats.govt.nz/Census/2013-census/profile-and-summary-reports/quickstats-culture- identity/ethnic-groups-NZ.aspx. Accessed December 1 2018

Statistics New Zealand (2007) Internal migration. http://archive.stats.govt.nz/browse_for_stats/population/Migration/internal-migration.aspx. Accessed on 15 December 2018

Statistics New Zealand (2013a) 2013 Census definitions and forms.

http://archive.stats.govt.nz/Census/2013-census/info-about-2013-census-data/2013-census-definitions-forms/definitions.aspx. Accessed January 12 2019

Statistics New Zealand (2013b) 2013 Census information by variable. http://archive.stats.govt.nz/Census/2013-census/info-about-2013-census-data/information-by-variable.aspx. Accessed January 1 2019

Statistics New Zealand (2015) IDI data dictionary: 2013 census data.

http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/idi-data/2013-census-data.aspx. Accessed January 2 2019

Statistics New Zealand (2017a) 2001 Census of population and dwellings: change in ethnicity question. http://www.stats.govt.nz/Census/2001-census-data/change-in-ethnicity-question.aspx. Accessed January 11 2019

Statistics New Zealand (2017b) GDP rises on strength in services.

https://www.stats.govt.nz/news/gdp-rises-on-strength-in-services. Accessed November 6 2018

Statistics New Zealand (2017c) Labour force participation rate. http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/nz-social-indicators/Home/Labour%20market/lab-force-particip.aspx#anchor16. Accessed October 17 2018

Statistics New Zealand (2019a) Ethnic Group (Grouped Total Responses) and Maori Descent by Sex, for the Census Usually Resident Population Count, 1991, 1996, 2001 and 2006. http://nzdotstat.stats.govt.nz/wbos/Index.aspx?DataSetCode=TABLECODE231. Accessed 7 August 2019.

Statistics New Zealand. (2019b). Ethnic group (total responses) by age group and sex, for the census usually resident population count, 2001, 2006, and 2013 Censuses. http://nzdotstat.stats.govt.nz/wbos/Index.aspx?DataSetCode=TABLECODE8021. Accessed 7 August 2019.

Theil H (1972) Statistical Decomposition Analysis: with applications in the social and administrative sciences. North-Holland, Amsterdam.

Theil H, Finezza AJ (1971) A note on the measurement of racial integration of schools by means of informational concepts. J MathSociol 1: 187-194.

van der Pas S, Poot J (2011) Migration paradigm shifts and transformation of migrant communities: the case of Dutch Kiwis. CReAM. http://cream-migration.org/publ_uploads/CDP_12_11.pdf. Accessed September 21 2018

van Mourik A, Poot J, Siegers JJ (1989) Trends in occupational segregation of women and men in New Zealand: some new evidence. N Z Econ Pap 23: 29-50.

Walks RA, Bourne LS (2006) Ghettos in Canada's cities? Racial segregation, ethnic enclaves and poverty concentration in Canadian urban areas. Can Geographer/Le Géogr Can 50:273–297.

Weeden KA, Grunsky DB (2005) The case for a new class map. Am J Sociol 111: 141–212.

White MJ (1986) Segregation and diversity measures in population distribution. Popul Index 52: 198-221.

White MJ, Glick JE (1999) The impact of immigration on residential segregation. In: Bean FD and Bell-Rose S (eds) Immigration and Opportunity: Race, Ethnicity and Employment in the United States. Russell Sage.

Chapter Appendix

Table A1: Ethnic group classification in New Zealand

Ethnic group code (Level 1)	Ethnic group code description (Level 1)	Ethnic group code (Level 2)	Ethnic group code description (Level 2)
01	European	10	European not further
		11	NZ European
		12	Other European
02	Māori	21	NZ Māori
03	Pacific Peoples	30	Pacific Island not further
		31	Samoan
		32	Cook Island Māori
		33	Tongan
		34	Niuean
		35	Tokelauan
		36	Fijian
		37	Other Pacific Island
04	Asian	40	Asian not further defined
		41	Southeast Asian
		42	Chinese
		43	Indian
		44	Other Asian
05	MELAA	51	Middle eastern
		52	Latin American/Hispanic
		53	African
06	Other	61	Other ethnicity

Note: There are 21 ethnic groups at Level 2. In the empirical analysis, population counts of the three 'not further defined' categories at Level 2 have been proportionally distributed among the corresponding Level 2 groups within the same Level 1 ethnic group. Consequently, the analysis at Level 2 is based on 18 groups.

Source: Statistics New Zealand (2013a)

Table A2: Entropy Index of Sorting (EIS) by qualification: Auckland, 1991-2013

1991			1996			2001			2006			2013		
Qualification	Percentage of population	EIS	Qualification	Percentage of population	EIS	Qualification	Percentage of population	EIS	Qualification	Percentage of population	EIS	Qualification	Percentage of population	EIS
No School Qualification	38.4	0.04	No Qualification	35.6	0.0609	No Qualification	24.4	0.055	No Qualification	22.5	0.057	No Qualification	17.01	0.0516
Overseas Qualification	4.34	0.025	Overseas School Qualification	2.89	0.0115	Overseas Secondary School Qualification	7.89	0.016	Overseas Secondary School	7.33	0.017	Overseas Secondary School Qualification	8.92	0.0198
School Certificate	21.8	0.006	School Certificate Qualification	13.2	0.0065	Fifth Form Qualification	15.4	0.008	Level 1 Certificate	13.2	0.009	Level 1 Certificate	11.49	0.0132
Sixth Form Certificate, University Entrance	17.6	0.019	Sixth Form Qualification	10.7	0.0054	Sixth Form Qualification	12.1	0.003	Level 2 Certificate	11.1	0.003	Level 2 Certificate	10.46	0.005
Higher School Certificate, Higher Leaving Certificate	6.61	0.021	Higher School Qualification	6.08	0.0244	Higher School Qualification	7.2	0.019	Level 3 Certificate	9.67	0.01	Level 3 Certificate	10.61	0.0102
University Bursary, Scholarship	6.56	0.065				Other NZ Secondary School Qualification	0.02	0.283	Level 4 Certificate	9.1	0.012	Level 4 Certificate	8.419	0.0138
Other School Qualification	4.67	0.023	Basic Vocational Qualification	3.81	0.0054	Basic Vocational Qualification	4.18	0.003	Level 5 Diploma	4.05	0.007	Level 5 Diploma	4.423	0.005
			Skilled Vocational Qualification	6.82	0.0154	Skilled Vocational Qualification	4.93	0.014	Level 6 Diploma	5.38	0.015	Level 6 Diploma	4.984	0.0123
			Intermediate Vocational Qualification	1.63	0.0204	Intermediate Vocational Qualification	2.07	0.008						

			Advanced Vocational Qualification	9.05	0.023	Advanced Vocational Qualification	8.65	0.017						
			Bachelor Degree	7.15	0.0748	Bachelor Degree	9.4	0.063	Bachelor Degree and Level 7 Qualification	13	0.053	Bachelor Degree and Level 7 Qualification	16.4	0.0379
			Higher Degree	2.99	0.0781	Higher Degree	3.73	0.067	Post-graduate and Honours Degrees	1.79	0.05	Post-graduate and Honours Degrees	2.93	0.0361
									Masters Degree	2.37	0.054	Masters Degrees	3.56	0.0441
									Doctorate Degree	0.497	0.101	Doctorate Degree	0.747	0.0971

Table A3: Entropy Index of Sorting (EIS) by occupation: Auckland, 1991-2013

	1991		1996		2001		2006		2013	
Occupation	Percentage of Population	EIS	Percentage of Population	EIS	Percentage of Population	EIS	Percentage of Population	EIS	Percentage of Population	EIS
Legislators & Administrators	0.013	0.361	0.064	0.242	0.959	0.0817	1.29	0.0597	1.21	0.0736
Corporate Managers	14.8	0.033	14.7	0.032	15.1	0.0249	17.0	0.0227	18.6	0.0182
Physical, Mathematical & Engineering Science Professionals	1.83	0.036	2.26	0.025	3.20	0.0231	3.69	0.0213	4.43	0.0215
Life Science & Health Professionals	2.82	0.021	2.57	0.034	2.71	0.0289	2.90	0.0253	3.49	0.0278
Teaching Professionals	3.90	0.048	3.93	0.017	4.45	0.0129	4.74	0.0092	5.74	0.0078
Other Professionals	3.54	0.013	3.76	0.049	5.22	0.0402	5.65	0.0355	6.83	0.0304
Physical Science & Engineering Associate Professionals	3.66	0.044	3.13	0.014	2.97	0.0140	2.87	0.0105	2.88	0.0128
Life Science & Health Associate Professionals	0.93	0.022	0.84	0.038	0.875	0.0351	0.85	0.0311	0.92	0.0428
Other Associate Professionals	7.86	0.006	9.28	0.016	9.72	0.0093	11.1	0.0055	11.8	0.0044
Office Clerks	12.70	0.013	12.3	0.006	11.4	0.0053	9.51	0.0047	9.07	0.0062
Customer Service Clerks	4.63	0.028	4.73	0.010	4.44	0.0094	3.74	0.0093	3.51	0.0129
Personal & Protective Services Workers	7.10	0.008	7.91	0.016	7.81	0.0122	7.67	0.0108	7.77	0.0110

Salespersons, Demonstrators & Models	5.65	0.246	6.21	0.007	5.94	0.0069	5.39	0.0061	5.14	0.0083
Market Orientated Agricultural & Fishery Workers	3.00	0.016	3.10	0.201	2.64	0.1539	2.22	0.136	1.92	0.171
Building Trade Workers	5.36	0.030	4.82	0.016	4.69	0.0174	4.67	0.0178	3.70	0.0230
Metal & Machinery Trades Workers	3.88	0.039	3.07	0.029	2.67	0.0307	2.36	0.0322	2.01	0.0475
Precision Trade Workers	1.11	0.033	1.04	0.040	0.841	0.0418	0.80	0.0376	0.41	0.105
Other Craft & Related Trades Workers	1.62	0.120	1.50	0.031	1.20	0.0339	0.99	0.0326	0.67	0.0769
Industrial Plant Operators	0.75	0.069	0.747	0.105	0.604	0.1055	0.53	0.1233	0.32	0.193
Stationary Machine Operators & Assemblers	4.55	0.044	3.96	0.065	3.85	0.0698	3.14	0.0691	2.23	0.0738
Drivers and Mobile Machinery Operators	2.90	0.142	2.66	0.043	2.6	0.0554	2.55	0.0556	1.97	0.0621
Building & Related Workers	0.26	0.120	0.316	0.125	0.409	0.0997	0.44	0.0912	0.20	0.189
Labourers & Related Elementary Service Workers	7.20	0.048	7.10	0.047	5.67	0.044	5.95	0.0482	5.20	0.0452

Table A4: Entropy Index of Sorting (EIS) by income: Auckland, 1991-2013

1991			1996			2001			2006			2013		
Income in NZD	Percentage of Population	EIS	Income in NZD	Percentage of Population	EIS	Income in NZD	Percentage of Population	EIS	Income in NZD	Percentage of Population	EIS	Income in NZD	Percentage of Population	EIS
Nil income /Loss	4.99	0.014	Loss	0.185	0.014	Loss	0.738	0.036	Loss	0.628	0.033	Loss	0.457	0.058
1-2,500	5.17	0.012	Zero Income	4.96	0.012	Zero Income	4.85	0.012	Zero Income	5.60	0.011	Zero income	8.59	0.014
2,501 - 5,000	3.31	0.007	1 - 5,000	9.61	0.007	1 - 5,000	8.67	0.007	1 - 5,000	7.51	0.007	1-5,000	6.34	0.008
5,001 - 7,500	6.63	0.021	5,001 - 10,000	14.1	0.021	5,001 - 10,000	11.0	0.011	5,001 - 10,000	7.50	0.009	5,001-10,000	5.25	0.009
7,501 - 10,000	13.5	0.018	10,001 - 15,000	14.6	0.018	10,001 - 15,000	13.0	0.013	10,001 - 15,000	10.5	0.011	10,001-15,000	7.99	0.009
10,001 - 15,000	14.6	0.009	15,001 - 20,000	9.24	0.009	15,001 - 20,000	8.46	0.006	15,001 - 20,000	7.85	0.008	15,001-20,000	8.46	0.011
15,001 - 20,000	11.1	0.005	20,001 - 25,000	8.83	0.005	20,001 - 25,000	7.69	0.006	20,001 - 25,000	6.77	0.006	20,001-25,000	6.55	0.008
20,001 - 25,000	9.87	0.004	25,001 - 30,000	9.66	0.004	25,001 - 30,000	8.53	0.005	25,001 - 30,000	6.95	0.008	25,001-30,000	5.56	0.005
25,001 - 30,000	9.33	0.006	30,001 - 40,000	12.6	0.006	30,001 - 40,000	13.7	0.005	30,001 - 35,000	6.84	0.006	30,001-35,000	5.15	0.005
30,001 - 40,000	10.6	0.014	40,001 - 50,000	6.74	0.014	40,001 - 50,000	8.60	0.010	35,001 - 40,000	7.44	0.004	35,001-40,000	5.7	0.006
40,001 - 50,000	5.14	0.023	50,001 - 70,000	5.10	0.023	50,001 - 70,000	8.00	0.025	40,001 - 50,000	10.2	0.005	40,001-50,000	9.27	0.004

50,001 - 700,00	3.41	0.048	70,001 - 100,00	2.27	0.048	70,001 - 100,000	3.52	0.052	50,001 - 70,000	11.5	0.015	50,001- 60,000	7.73	0.005
70,001 or More	2.36	0.119	100,001 or More	2.06	0.119	100,001 or More	3.24	0.119	70,001 - 100,000	5.59	0.034	60,001- 70,000	5.99	0.007
									100,001 or More	5.05	0.110	70,001- 100,000	8.97	0.020
												100,001- 150,000	4.73	0.049
												150,001 or More	3.27	0.135

Note: NZD refers to New Zealand Dollar

Chapter 4: Determinants of Ethnic Identity among Adolescents: Evidence from New Zealand

4.1 Introduction

Ethnic mobility is defined as the social phenomenon whereby people change their ethnic identity/affiliation over time. This switching of ethnic identity can be triggered by changing incentives and circumstances that impact on the desire of an individual to belong to a specific ethnic group. Ethnic mobility plays an important role in social change but, due to scarcity of appropriate longitudinal data, the literature on (dynamic) ethnic mobility is much less extensive than the literature on (static) ethnic identity. Though the literature finds that individuals can affiliate themselves with more than one ethnicity and may change their ethnic affiliation over time (for example Carter *et al.* 2009, Simpson *et al.* 2016), there are relatively few studies to date looking into the factors associated with these changes. In most research applications it is assumed for convenience that ethnic affiliation is constant over time (Carter *et al.* 2009).

The extant literature on ethnicity looks mainly at the predictors of ethnic identity (for example Nelsen 1990, Phinney and Chavira 1992, Qian 2004, Casey and Dustmann 2010, Lee and Bean 2010) and less so at *changes* in ethnic identity. More research is needed in order to assess the prevalence of ethnic mobility, as well as to answer questions regarding the fluidity and causes of ethnic mobility (Brown *et al.* 2010). Extant studies on inter-ethnic mobility in New Zealand have been based on the longitudinal Survey of Family, Income and Employment (SoFIE) (see Carter *et al.* 2009) or on data that links individuals between censuses (Statistics New Zealand 2009, Didham 2016). In a report on inter-censal ethnic mobility, Statistics New Zealand (2009) identified significant changes in ethnic identity of individuals between 2001 and 2006. They also reported that as people age, their ethnic mobility declines. According to this report, the age group that had changed their ethnic identity the most were those aged between 5 and 14 years, followed by individuals belonging to the age group 15-24.

According to Phinney (1989, 1990, 1992), the formation of ethnic identity of an individual has three stages: (1) a stage of unexamined ethnic identity; (2) an exploration period and (3) ethnic identity achievement. Unexamined ethnic identity refers to the stage where individuals have not previously been exposed to issues of ethnic identity. In this stage, mostly in childhood, individuals are influenced by the dominant culture and ethnic attitudes from their parents or other adults. In the second stage, as they transition into adolescence, an individual explores and becomes aware of their own ethnicity through their experiences. This stage involves acquiring ethnic knowledge, for example by reading, mixing with people, visits to ethnic museums, or active participation in cultural events. This stage might result in complete absorption of one's own culture as well as dismissal of the values of the dominant culture. This stage is followed by the development of individual ethnic identity, where an individual accepts and internalizes their own ethnicity. Umaña -Taylor *et al.* (2009) add stages of resolution and affirmation before the stage of ethnic identity achievement among adolescents. According to them, the period of exploration results in adolescents feeling more independent in terms of decision making, which promotes resolution of ethnic identity as they are actively deciding what ethnicity means to them. They also add that exploration and resolution of ethnic identity is facilitated by an adolescent's cognitive transitions. However, affirmation (that is, positive/negative feelings about one's ethnicity) is a social-context-driven process that depends largely on others' perception of one's ethnic group, and can be achieved at a young age prior to adolescence.

Though the international literature highlights the importance of studying the ethnic-identity transition of young adults (Brown *et al.* 2010), this age group has not yet been looked at explicitly in studies on ethnic identity or ethnic mobility in New Zealand. In this paper, we therefore focus our attention on this period of substantial transition for all people: the transition to adulthood. This offers a unique perspective, as adolescents are given the opportunity to define their own ethnic affiliation/s, often for the first time. Specifically, we look at adolescents aged 13 to 17 years in one Census, and capture their ethnicity (and thereby any ethnic mobility) in the following Census by means of linked longitudinal census data. Observing persistence or change in these adolescents' ethnic identity, we look at the extent to which their current ethnic identity is associated with their ethnic identity in the preceding Census, along with other factors.

We use Phinney's model of ethnic identity formation among children as the conceptual framework for our study. Following Phinney (1989, 1990), ethnic identity of the adolescent in the previous Census resembles the adolescents' stage of unexamined ethnic identity. The exploration period is the period between Censuses.46 The adolescents' ethnic identity in the following Census is the stage of ethnic identity achievement. For example, an adolescent present in the census in both 1991 and 1996, may be assumed to still be in the stage of unexamined ethnic identity in 1991, then be in an exploration period between 1991 and 1996, and achieve his/her ethnic identity by 1996. We recognise that this characterisation of the stages

[46] The inter-censal periods considered in our work are 1991-1996, 1996-2001, 2001-2006, and 2006-2013.

of ethnic identity will not represent the experience of all, or necessarily even a majority, of the adolescents in our study. However, it is a useful theoretical construct to guide our empirical work.

The paper proceeds as follows. In Section 4.2, we discuss relevant studies on the ethnic identity of children and adolescents, with a particular focus on New Zealand research. Section 4.3 describes the data and Section 4.4 details the method. Section 4.5 presents and discusses the results, and Section 4.6 concludes.

4.2 Background

Most studies related to the ethnic identity of children in the U.S. have been done using broad ethnic categories, and do not capture the significant heterogeneity present within these broad categories. These studies also focus mainly on minority ethnic groups only. Rumbaut (1994) identified the main predictors of ethnic identity of the children as their sex, nativity (U.S born or foreign born), parents' nativity and professional status (dummy variables for parents in high-status professions), language and racial-ethnic discrimination (being treated less fairly than other groups based on race and ethnicity). Eschbach and Gómez (1998) examined the determinants of changing ethnic identification in a representative national sample of Hispanic high school students. They looked at students who changed their ethnic affiliation between the first and second waves of the High School and Beyond surveys in 1980 and 1982. They found that use of Spanish language and the Hispanic composition of the school were strongly related to switching to non-Hispanic ethnic identity. Ethnic group concentration was found to be significantly and negatively related to identity switching. Sex and family income were found to have little impact on the odds of switching identity.

In Canada, Frideres and Goldenberg (1982) found that, in terms of one's ethnic identity formation as well as ethnic identity change, family, sex and occupation play a significant role, family being the most important. According to them, one's ethnic identity develops as a response to structural conditions in the society, and also to adapt to the conditions related to the contest for scarce desirable goods. They also add that in Canada, a systematic relationship exists between ethnic affiliation and occupation, education and income. Tsang *et al.* (2003) conducted qualitative research examining the concept of ethnic identity through the experiences of satellite children[47] in Canada. They found that ethnic identity choices/responses were strongly influenced by whether the child intended to return to their country of origin. Moreover, language barriers, cultural barriers, perception of one's ethnic identity, and acceptance from the host (Canadian) society, also influenced the ethnic identity choices of these children.

In Australia, Rosenthal and Hrynevich (2007) studied the developmental changes in the nature of ethnic identity in younger and older adolescents of minority non-Anglo groups (of Greek and Italian origin) as well as the dominant Anglo society. Language, religion, social activities, maintenance of cultural traditions, family life, perception about their own ethnic group as well as attachment towards their own ethnic group were found to be associated with a child's ethnic identity.

The ethnic identity of children can be influenced by the ethnic traits of their parents (Casey and Dustmann 2010), although research on which parent affects ethnic identity of the children to the greater extent shows varied results. Among the relevant studies in the U.S., De Snyder *et al.* (1982) found that female children of Mexican-American couples tended towards identifying themselves as Mexicans more than did male children. Stephan and Stephan (1989) identified that the ethnic identity of children of intermarried couples is affected more by the minority parent. Nelsen (1990) showed the mother to be the most influential, whereas Waters (1989) stated that due to the common use of the father's surname, the father is more influential in the ethnic identity formation of children of an intermarried couple. Saenz *et al.* (1995) identified that factors associated with the ethnic identity of the children of Asian-Anglo intermarried couples occur at three levels: (1) child (age, sex, generation, Asian language); (2) parent (ethnic group, Asian parent's sex, Asian parent's education) and (3) ethnic community (group size, ethnic heterogeneity, socio-economic status). Children with an Asian father had a greater tendency towards identifying themselves as Asian than those whose mother was Asian. This is similar to the theoretical literature on the ethnic identity formation of children of intermarried couples (Hwang and Murdock 1991). Lee and Bean (2010), using 2010 U.S. Census data, found that children of multi-racial parents often exhibit a single ethnicity, usually identifying themselves as belonging to the majority group, due to greater social acceptability and better

[47] Children of ethnically Chinese immigrants to North America, mainly from Hong Kong and Taiwan, whose parents (mostly fathers) have returned to their country of origin after immigration, to pursue economic advantages, leaving the mother and the child to try and settle in the new country (Man 1994, Tsang *et al.* 2003).

opportunities. The ethnic identity of immigrant children was determined by their exposure to own-ethnic group members as well as non-group members, their family ties and their parents, exposure to natives, fluency in the minority parent's language and familiarity with the host country culture.

Ethnic mobility and identity may be affected by peer groups. In an exploratory study, Phinney and Chavira (1992) examined the changes in ethnic identity that occur in young adolescents within a sample of eighteen adolescents from three ethnic groups (Asian American, Black, and Hispanic). It was found that the ethnic identity of children was influenced by the ethnic group they belong to and the peer group they interact with. In a qualitative study in Europe, Tizard and Phoenix (1995) interviewed children with mixed parentage (one white parent, and one African or African-Caribbean parent) who were living in London. The authors found that school, social class and peer groups influenced the children's ethnic identity much more than the racial characteristics of their parents.

In another qualitative study, Kickett-Tucker (2009) found strong sense of self, Aboriginal culture, family, friends and Aboriginal language to be the important contributors of ethnic identity of both children and youths of urban Perth in Western Australia. Mowen and Stansfield (2016) observed prominent shifts in the racial identity of the immigrant children in San Diego and Miami from 1991 to 2013. Peer influence, and stress regarding social as well as educational performance and the need to maintain a positive dignity influenced the identities of these children. Moreover, they found a clear relation between shifts in racial identity of the immigrant children and their attachment with family and the values they assign to their self-worth and self-esteem. In a qualitative study in Australia, Kickett-Tucker (2008) found that peer interactions through school sport settings provided opportunities to Indigenous (Aboriginal) students to affirm their racial identity and self-esteem in a positive way. Aboriginal students participating in sports interacted with their own group members and hence collectively identified and expressed themselves positively as belonging to an Aboriginal group.

Ethnic community can also influence ethnic mobility. Saenz *et al.* (1995) found that children living in a neighbourhood containing more people belonging to their Asian parent's ethnic group, were more likely to identify themselves as Asian, while children living in heterogeneous areas were more likely to identify themselves as Anglo. Fitzpatrick and Hwang (1992) established strong support for the relationship between social structure, especially group size and heterogeneity, and intergroup relations in the formation of ethnic identity. Qian (2004), using 1990 U.S. Census data, found that children's identification varied by minority concentration in their neighbourhood. Children of couples in which the minority spouse had part white ancestry tended more towards being identified as white.

In New Zealand, studies on ethnic mobility are very limited and mainly focus on an individual's self-identification process. In a cross-sectional study of inter-censal change, Coope and Piesse (2000) found there was considerable mobility between ethnic groups. For example, they found a 23 percent inflow and 6 percent outflow for the Māori ethnic group between 1991 and 1996. Possible reasons for ethnic mobility in New Zealand include changes in the ethnicity question between censuses, changes in the socio-political environment, changes in the political structure (Carter *et al.* 2009), ethnogenesis[48] (Kukutai and Didham 2009) and increases in intermarriage (Callister *et al.* 2005, Howard and Didham 2005, Kukutai 2007. Callister *et al.* 2008).

Didham (2016) used New Zealand Linked Census (NZLC) data from 1981 to 2013 to investigate ethnic mobility (for Level 1 ethnic groups)[49] in New Zealand. He considered both inflow and outflow of individuals and found that ethnic mobility affects a large proportion of all ethnic groups. He also found that ethnic mobility affects age groups in a different manner, as individuals move through educational, employment, partnering and peer-group changes throughout their life course.

In New Zealand, studies on ethnic mobility of the youth population are particularly limited. Kukutai (2007) showed that European mothers in European-Māori couples identify their child as Māori as often as Māori mothers do. Her finding challenges the assumption that minority ethnic identity is transmitted by minority parents only. Māori ethnic identity was found to be transmitted in a less predictable manner across generations as the parental union becomes more European (one partner identifying as both Māori and European and the other as European). Moreover, Kukutai (2007) stated that a child's Māori identity is consistent with patriarchal rules, with Māori paternity being more influential than Māori maternity in designing the child's Māori identity.

Kukutai (2008) then observed the responses to ethnic group and main ethnic group questions included in the first wave of the longitudinal Youth Connectedness survey in 2006. She focused

---

[48] The formation of new ethnic categories within a larger community, such as 'New Zealander'.

[49] European, Māori, Pacific, Asian, Middle Eastern/Latin American/African (MELAA), and Other.

on whether the ethnic identity response changed according to contextual factors, and found that affiliation changes might occur when children reach an age when they define their own ethnicity rather than having a parent do it for them.

These New Zealand studies (Kukutai 2007, 2008) describe ethnic identification patterns but do not identify the primary causes of changes. Carter *et al.* (2009) examined changes in self-identified ethnicity among New Zealand adults (aged over 15), from 2002 to 2005. They looked at the proportion of people that changed their self-identified ethnicity over the first three years of SoFIE. They found that the biggest predictor of an individual's ethnicity at wave 2 was dependent on the individual's ethnicity at wave 1. Hence, self-declared ethnic identity is a social process with strong temporal persistence. They found that adults who changed their ethnic identity were more likely to be younger, overseas born, belong to deprived groups,[50] and have poorer health.

In contrast with earlier work on New Zealand, in this paper we focus exclusively on young adolescents in New Zealand. We also use more disaggregated ethnic groups than earlier studies, which helps us better capture the heterogeneity within the broad ethnic groups used in previous studies (Mondal *et al*. forthcoming). In New Zealand, previous research on ethnic identity among children has mainly focused on the influence of having single/multiple ethnicities and multi-ethnic parents on ethnic identity of children (Kukutai 2007, 2008). Instead, our research looks into the predictors of self-declared ethnic identity choices made by adolescents residing in Auckland.

4.3 Data

Auckland is the most ethnically diverse region in New Zealand. Auckland is also relatively youthful: 35.9 percent of residents are aged under 25 years, compared with 34.2 percent for all of New Zealand (Statistics New Zealand 2013a). In Auckland, European (59.3 percent), Asian (23.1 percent), Pacific Peoples (14.6 percent), Māori (10.7 percent), MELAA[51] (1.9 percent)

---

[50] Measured by the New Zealand Deprivation Index (NZDep2001), which provides a neighbourhood-level (approximately 100 people) deprivation score (Salmond and Crampton 2002). The Index is a tool for measuring socio-economic position and health/social outcomes based on eight questions regarding income, home ownership, support, employment, qualifications, living space, communication and transport.

[51] Middle Eastern/Latin American/African.

and Other (1.2 percent) are the major ethnic groups (Statistics New Zealand 2013a).[52] Auckland accounts for one-third of the New Zealand population, and has the largest population of the 16 regions of New Zealand. Because of its ethnic diversity and relative youthfulness, we chose Auckland as our area of focus.

We recognise that there is a key period during which an adolescent transitions from their ethnicity being recorded in official data by their parents, to their ethnicity being recorded by themselves. We infer that the most likely time for this transition is in late adolescence, and look at the ethnic affiliation of individuals (when they were a child) in one Census, in which their ethnicity was likely to have been recorded by their parent/s, followed by a Census in which their identity was likely to have been recorded by the individual themselves (once they have attained greater independence). Specifically, we take individuals aged between 18-22 in the current census who have been linked in the NZLC to the same individuals who were aged between 13 and 17 in the previous census.[53]

Data for this analysis were obtained from the 1991, 1996, 2001, 2006 and 2013 New Zealand Census of Population and Dwellings data for the Auckland region of New Zealand.[54] The New Zealand Census of Population and Dwellings is usually conducted every five years, and collects a range of socio-demographic information on individuals present in New Zealand on census night who are usually resident in New Zealand. These census data include information about individual characteristics like usually-resident location, sex, age, ethnicity, education, occupation, marital status, and income level. These individual data can be aggregated to population statistics at several geographical scales. Successive censuses have been linked to create longitudinal datasets, which enable us to investigate changes in population characteristics, including ethnic identity, across time.[55] The link rate for individuals from the

---

52 Percentages do not sum to 100 percent, as people can report more than one ethnicity.

53 Due to the seven-year gap between the 2006 and 2013 censuses, for the 2006-2013 linked Census we include individuals aged between 11 and 17 in 2006, who are aged between 18-24 in 2013.

54 The most recent population census was held on March 6, 2018. Early results from this census have been released since late 2019, but the 2018 data required for this paper were not available at the time the research was conducted.

55 The NZLC (New Zealand Linked Census) links adjacent censuses in pairs, so that the seven censuses from 1981 to 2013 are linked into six pairs (Didham 2016).

1991 to 1996 Census was 72 percent, 1996 Census to 2001 Census was 69.5 percent and from the 2001 to 2006 Census was 70.3 percent (Statistics New Zealand 2014).[56]

Our analysis is based on data aggregated to the area unit level.[57] The Auckland region is made up of 413 land-based area units, of which 409 had a non-zero usually resident population throughout the period from 1991-2013. Area units with no usually resident population were excluded from the analysis. The 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 Statistic New Zealand, all counts, including numbers of observations in regression models, have been randomly rounded to base three.[58]

Ethnicity is the ethnic group or groups a person identifies with or has a sense of belonging to (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 2013b). The Level 1 and Level 2 classification of ethnicity are shown in Table 4.1. We consider ethnic groups at a finer scale (Level 2) than those used in previous research in New Zealand (and comparable work elsewhere). This is important because the Asian and Pacific broad ethnic groups in particular mask substantial heterogeneity in characteristics. In contrast, past studies in other countries have focused on a small number of groups in their studies of ethnic identification (for example, only Hispanic, Asian-Anglo or Asian-American; see Phinney and

---

56 A census pair '*t*,*t*-1' refers to a pair of censuses where individual records in census (*t*) are linked to those of the previous census (*t*-1). For example, if we are looking at linking records from the 1996 Census to those from the 1991 Census, we will refer to this as the 1991–1996 census pair. Though the terms 'matching' and 'linking' are used interchangeably, a 'link' refers to a record pair where the connection has been assessed as probable.
A 'match' refers to a record pair where the connection is true. The matching process comprises of two parts: deterministic matching based on a set of key variables to find the unique matches, followed by probabilistic matching on the residuals. Deterministic matching uses a set of matching variables (sex, birth day, month and year, and area unit of residence) and matched records have the same unique values of the matching variables. In contrast, probabilistic matching evaluates all possible matches and uses statistical techniques to achieve matches. The link-rate for 2006-2013 is not reported anywhere.

57 Area units are aggregations of meshblocks. They are non–administrative areas that are in between meshblocks and territorial authorities in size (Statistics New Zealand 2013b). In urban areas, area units are approximately the size of suburbs, and in our dataset they have an average population of 1530. In this paper, we have used 2013 area unit boundaries.

58 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 seven will be rounded to either six (with probability 2/3) or nine (with probability 1/3).

Chavira 1992, Casey and Dustmann 2010, Mowen and Stansfield 2016) We explore the behaviours of all of the Level 2 ethnic groups in our study, as we feel that the opportunity of multi-ethnic affiliation and changing ethnic identity is not concentrated only within small or minority ethnic groups.

Data on self-reported ethnic identification are collected in all censuses and each person can choose a single or multiple response.[59] We take every ethnicity that is reported for our sample of adolescents. Therefore, the adolescent's composite ethnicity is comprised of a binary variable (belongs to the ethnic group=1, otherwise=0) for each ethnic group. We do not include the 'not further defined' categories in our analysis. Moreover, due to the small number of individuals reporting as 'Tokelauan' and 'Other Pacific People', we combined these two groups. For the same reason, we also combined 'Middle Eastern', 'Latin American', 'African' and 'Other ethnicity' ethnic groups into a single group. These ethnicity assignments yield altogether 14 distinct ethnic groups.

[59] The ethnic classifications have changed considerably between 1991 and 2013. Up to three responses were recorded for each individual in 1991 and 1996 compared with up to six in the later Censuses. The format and wording of the Census ethnicity question changed twice between 1991 and 2001. Some significant changes have been identified, including increased multiple responses in 1996 and 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). Moreover, the treatment of responses of 'New Zealander' to the Census ethnicity question has also changed over time. In 2001, 'New Zealander' was counted in the New Zealand European category, whereas in 2006, New Zealander was instead included as a new category. The increase in counts for the New Zealand European category from 2006 to 2013 is partly attributable to fewer people identifying themselves as 'New Zealander' in 2013 (Statistics New Zealand 2017). These changes affect the comparability of ethnicity data in New Zealand over time. However, we deal with this complication by adding inter-censal fixed effects in our regression.

Table 4.1: Ethnic Group Classification and Counts in Auckland New Zealand 1991-2013

Ethnic group code (Level 1)	**Ethnic Group code description (Level 1)**	**1991**	**1996**	**2001**	**2006**	**2013**	**Ethnic group code (Level 2)**	**Ethnic Group code description (Level 2)**	**1991**	**1996**	**2001**	**2006**	**2013**
01	European	625,614	1262,403	667,755	672,055	733,469	10	European not further defined	150	21	228	195	141
							11	NZ European	574,932	536,606	616,859	611,901	696,966
							12	Other European	50,532	725,776	50,668	59,959	36,362
02	Māori	85,926	105,213	127,704	137,304	142,767	21	NZ Māori	85,926	105,213	127,704	137,304	142,767
03	Pacific Peoples	83,370	107,262	163,632	190,581	209,652	30	Pacific Island not further defined	<6	<6	9	6	9
							31	Samoan	41,784	51,639	76,584	87,840	95,916
							32	Cook Island Māori	17,466	21,234	31,077	34,371	36,546
							33	Tongan	12,456	17,958	32,535	40,140	46,971
							34	Niuean	9,354	11,466	16,038	17,667	18,555
							35	Tokelauan	504	627	1,488	1,848	1,959
							36	Fijian	1,506	3,174	4,155	5,847	8,493
							37	Other Pacific Island	300	1,164	1,755	2,868	1,212
04	Asian	18,984	49,211	80,958	134,462	96,766	40	Asian not further defined	<6	45	81	30	21
							41	Southeast Asian	1,806	6,561	9,363	15,909	10,911
							42	Chinese	9,738	23,505	38,025	60,186	39,456
							43	Indian	7,209	16,905	23,484	39,262	34,064
							44	Other Asian	231	2,240	10,086	19,105	12,335
05	MELAA	360	1,578	4,779	10,023	7,344	51	Middle eastern	282	1,194	3,624	6,897	3,759
							52	Latin American/Hispanic	33	204	474	1,194	2,658
							53	African	45	180	681	1,932	927
06	Other	108	198	279	3,687	510	61	Other ethnicity	108	198	279	3,687	510

Source: Statistics New Zealand (2013)

Notes: The data used for the Level 1 calculation have been constructed from Level 2 data sheets (using a bottom-up approach), so that the total responses counts at both levels are the same. In the case where the ethnicity count is less than six, data is suppressed and treated as zero according to the confidentiality rules of Statistics New Zealand.

4.4 Method

In our analysis, following Akerlof and Kranton (2000), each adolescent makes a choice in respect of each ethnic group '*i*' (that is, they choose/do not choose to declare that they belong to ethnic group '*i*'). Akerlof and Kranton (2000) propose a general utility function that includes 'identity' (an individual's sense of self) as a motivation for an individual's behaviour. Individuals may choose their social categories and they have a choice over identity, and this choice might be more or less conscious. In this model, there might be costs associated with choosing a specific identity (for example, disapproval from individuals choosing other options). Akerlof and Kranton treat an individual's identity, which is a basis for their utility function, as a function of their own given characteristics, the social category they belong to, and the behaviour towards them. We adopt this framework in developing our empirical model of the ethnicity choices of adolescents.

Given a set of *m* (=14 in our case) possible ethnicities from which an adolescent can choose, the classic approach is to assume that this choice is unique, that is, in the present context it would then be assumed that only one ethnicity can be selected. There are many choice models that have been developed to describe this situation, of which the multinomial logit model is one of the most popular (for example, Train 2009). However, we have already noted that individuals in New Zealand may identify with more than one ethnicity. In 2013 a maximum of six ethnicities could be stated. One way of modelling this situation is to consider every possible subset a unique choice. However, with adolescents being able to select up to six ethnicities out of 14, the number of subsets is very large (6476, including the case that none are selected) even though in practice many combinations are unlikely to be present in the data.[60] This potentially large number of choices makes the multinomial logit model and related models unwieldy and computationally burdensome.

Another complication is that the choice of a particular ethnicity is likely to be dependent on what other ethnicities have been selected. What matters is whether the utility attached to identifying with several ethnicities simultaneously is the sum of utilities attached to each of

---

[60] In 2013 about 90 percent of the population self-identified as belonging to only one ethnicity and only 0.05 percent self-identified as belonging to four or more ethnicities (Source: https://www.stats.govt.nz/tools/nz-dot-stat, Table: Ethnic group (detailed single and combination) by age group and sex, for the census usually resident population count, 2013; accessed 6/7/2019).

these ethnicities or not. If there is linear additivity of utility associated with specific ethnicities, the multinomial choice model of selecting a subset of ethnicities can then be decomposed into a set of independent binary choice model for selecting each of the ethnicities. However, if adding a certain ethnicity leads to lower utility overall (for example, because it is costly to maintain links with several disparate networks) or higher utility overall (for example, because the ethnicities share the same language), the choice model becomes considerably more complicated. Estimation methods for this case have only been emerging in recent years and require sophisticated data mining algorithms (see, for example, Benson *et al.* 2018). We will therefore adopt the assumption that selection of each ethnicity 'i' is independent of selection of any other ethnicity. This assumption reduces the multinomial subset choice model to a set of binary single choice models.

The theoretical underpinning of the ethnic choice model is the Random Utility Model (RUM) (McFadden 1984). The RUM has been used in many previous studies of choice (for example, Bhat and Guo 2004, 2007, Beine and Parsons, 2015). In the RUM, individuals are rational and attempt to maximise their utility. Individuals make decisions by comparing levels of utility associated with each possible alternative they have. In the ethnic choice model, individuals have m (=14) potential ethnic groups to affiliate themselves with (or not) and the choice depends on the characteristics of the chooser, as well as family and neighbourhood characteristics. In the ethnic choice model, $U_p(x)$ is the utility that person p obtains from choosing an ethnic group x (as opposed to not). Thus, utility level $U_p(x)$, can be represented as:

$$U_p(x) = W_p(x|Z, E) - C_p(x|Z, E) + V_p \qquad (1)$$

where $W_p(x|Z, E)$ is the deterministic component of utility and is a function of observed characteristics (including non-ethnic characteristics (Z) and ethnic affiliations (E) assigned by the parents at the previous census), and $C_p(x|Z, E)$ is the individual's cost of affiliating with ethnic group x. V_p represents unobserved individual-specific differences in utility. The unobserved component is assumed to be an independent and identically distributed (i.i.d.) random term drawn from an extreme value distribution. The individual adopts ethnicity x if and only if $U_p(x) > U_p(\text{not } x)$.

We base our empirical analysis on the assumption that a rational individual will always choose the option that gives them the maximum utility (McFadden 1984). We also assume that adolescents' decisions to affiliate with each ethnic group is independent of their choices to affiliate with other ethnic groups, other than to the extent that E enters into the utility function in Equation (1). To overcome this apparent paradox, we assume that the instantaneous decision to adopt an ethnicity x at time t (the current census) depends on an adolescent's previous ethnic affiliations at time t–1 (the previous census), but does not affect their instantaneous decisions about other ethnic affiliations. This assumption significantly simplifies the analysis, which would otherwise require a multinomial logit specification that would need to include all possible combinations of single and multiple-ethnic affiliations, as discussed above. Moreover, it is unclear whether a multinomial logit model (or similar) would meet the required independence of irrelevant alternatives IIA) assumption.[61] Some might argue that no choice is independent of other choices. However, as we include the individual's ethnic identity at the previous census (most likely stated by the adolescent's parent) as control variables in our analysis, we believe that we capture the most salient aspects of the interdependence of choices within the model.

Based on these assumptions, we use logistic regression to investigate the ethnic identity choices of adolescents. Our dependent variables are binary variables for each ethnic group, and represent whether or not the adolescent identifies with that group (1=identifies with the group, 0=otherwise), regardless of whether they also identify with one or more other groups. It is not possible to know who completes an individual's Census form. Past studies have worked on the assumption that an adult, most probably a parent (Brunsma 2005), fills Census forms for children under the age of 15. In our analysis, the linked individuals are aged 13-17 years in the previous Census and aged 18-22 in the current Census. We assume that once they reach the age of 18, young people report their ethnic identity in the Census form themselves. Therefore, we expect to capture the determinants of a young person's ethnic identity choice made by themselves, when they choose their ethnic identity for the first time.

We use data pooled across four pairs of linked censuses. Throughout this paper we use 'previous' to refer to data and individual ethnic choice from the first census in each inter-censal

---

[61] Other multinomial choice models such as the mixed logit model do not make this IIA assumption but are computationally burdensome when the number of potential choices (ethnicity combinations) is large.

pair and 'current' for the same in the second census in each pair. We take as the dependent variable each adolescents' ethnic response (1=identifies with the group, 0=otherwise) to each ethnic group, regardless of whether they also identify with one or more other groups, in the more recent census.

As noted in the literature review, the adolescents' ethnicity can be affected by independent variables defined at the individual, family, or neighbourhood level. All independent variables have values as observed at the start of each inter-censal period. The individual-level independent variables included in our analysis are the adolescent's ethnicity or ethnicities in the previous census, their age, sex, and whether they were born in New Zealand.[62] Family-level variables are limited to the ethnic identity of their parents (with some caveats, see below). The neighbourhood-level variables included in our analysis are the percentage shares of each ethnic group,[63] and the ethnic diversity of the area unit the adolescent resides in. We use the Entropy diversity measure as our measure of the ethnic diversity of each area unit (Mondal *et al.* 2020).[64] We expect that adolescents are exposed to more ethnicities if they live in more ethnically diverse areas. This may influence their ethnic identity choices (such as socialising and spending time with ethnically diverse population).

The Census records who the child's parents in the households are, but these data are not available for the whole population. The information is not available for children who were coded as an adult, or who were not present at home in the previous census, or when there was a change in parents in the intervening period. Hence missing data may lead to selection bias in our regression models and we therefore create our own proxy variable for the ethnic identity of the parent (who was the census respondent) for all adolescents in our sample. To proxy for the parents' ethnicity, we identified the number of parent-aged males and females (aged 30 to 60 years) of each ethnicity, in each adolescent's household. We dropped households with more

---

62 We derive the binary variable 'New Zealand born' (New Zealand born=1, otherwise=0) from the Census variable 'Birthplace'.

63 These proportions are based on the total number of ethnicities reported in the area unit and not the total number of individuals.

64 The entropy diversity measure is calculated using the formula $D_a = -\sum_{g=1}^{G} \frac{P_{ga}}{P_{.a}} \ln\left(\frac{P_{ga}}{P_{.a}}\right)$, (where P_{ga} refers to the population of group g (=1, 2,…G) in area a (= 1,2,….A), and $P_{.a}$ is the total population of the area).To allow for the calculation of D_a even in the case of there being groups who have zero members at some point in time, we define 0*ln(1/0)= $\lim_{q\to 0}[\, q(\ln(1/q)] = 0$. See also Theil (1972).

than one female or male adult from the analysis and hence limit the sample to the households with only one or two adults (and thereby assume that they are the parent/s).

To reduce the potential for over-fitting the regression models, we limit the explanatory variables for parents' ethnicities to include only ethnicities that match the dependent variable ethnicity. For example, in the model where Tongan is the dependent variable, we include only Tongan mother and Tongan father as explanatory variables, and not other ethnicity variables for the parents.

By definition, for adolescents in households with no adult females or no adult males have one parent's ethnicity undefined. To avoid any resulting bias, we include dummy variables to capture households with no adult females, and households with no adult males.

For the seven-year gap between 2006 and 2013 census, the variables we use in our study cover a seven–year period instead of five years. While this is a limitation of the study, it is not a serious one, as most of the variables are relatively time invariant or, in the case of age, are known with little measurement error. Moreover, we use inter-censal fixed effects in the analysis to control for inter-censal bias.[65]

The standard for ethnicity statistics was developed in 2005. The 'New Zealander' response, which was previously included in the 'European' category, was moved to the 'Other ethnicity' category in 2006 (Statistics New Zealand 2007). Thus, as a result, the number of people reporting a European or New Zealand European ethnicity has reduced in size and proportion, with a subsequent increase in the 'Other ethnicity' category. This is because it was New Zealand Europeans who were most likely to call themselves 'New Zealander' in the Census

---

[65] We also ran the same regressions separately for each inter-censal period. The results are consistent with our results from the data pooled across all censuses except for the combined 'MELAA and Other' ethnic category. For example, for the years 1991-1996 and 1996-2001, mother's ethnicity is not statistically significant for this group. The same for the years 2001-2006 and 2006-2013 are statistically significant. Moreover, the odds of choosing 'MELAA and Other' ethnic category in the current census, having belonged to the same category in the previous census are much greater in 1991-1996 and 1996-2001, than that in 2001-2006 and 2006-2013. For the same reason, we use interaction effects between year dummy variables and the 'MELAA and other' category, and interactions between year dummies and the 'MELAA and other' ethnic group proportions in the area unit the individuals live in.

(Statistics New Zealand 2007, Brown and Gray 2009). Considering the fact that we have combined the MELAA group with the 'Other' group in our analysis, we include interaction variables (interactions between individuals belonging to combined MELAA and Other ethnic group and their presence in each inter-censal period, and interactions between group proportion of combined MELAA and Other ethnic group in the area unit individuals reside in with their presence in in each inter-censal period) to account for any bias in the results that might arise due to the inter-censal issues.

Individuals residing in same area units are likely to be similar in terms of unobserved characteristics. Thus, our logistic regression uses standard errors clustered at the area unit level. In our model, *n* is the total number of individuals across area units, *k* is the number of clusters. Thus, n_k is number of individuals in cluster *k* and $\sum_k n_k = n$. For individual l in cluster *k*, Y_{lk} is the binary response for any given ethnic identity and X is a vector of *m* explanatory variables. Thus, our regression model can be expressed as:

$$logit\left(\frac{p_{lk}}{1-p_{lk}}\right) = \alpha + \beta' X + u_l + u_k \ ; l = 1 \dots . n, k = 1 \dots . \text{N} \qquad (2)$$

$$\text{where} \quad u_l, \sim N(0, \sigma_l^2)$$

$$u_k \sim N(0, \sigma_k^2)$$

where α is the fixed intercept term, β'_m is the effect of variable X_m on the response, u_l is the stochastic error term associated with individual l, u_k is the component of the error term that is common to all individuals in area unit *k*. The error terms u_l and u_k are assumed normally distributed random variables with zero mean and constant variance. p_{lk} is the probability that binary response for individual l in group k (that is, Y_{lk}) is equal to 1, given X_m and the random effects u_l and u_k.

4.5 Results and Discussion

To find the determinants of adolescents' ethnic identity choices, we run logistic regression with clustered standard errors for data pooled across all the censuses.[66] We report the results across Tables 4.2A, 4.2B and 4.2C. [67]

[66] Results for analyses based on individual Census waves are available on request.

[67] Tables 4.2A, 4.2B and 4.2C are reporting results from the same regression. We have broken down the results into different tables according to variables at different levels for easy readability.

Table 4.2A: Clustered Logistic Regression of Current Ethnicity – Effect of Previous Census Ethnicity

Variables	(1) NZ European	(2) Other European	(3) NZ Māori	(4) Samoan	(5) Cook Island Maori	(6) Tongan	(7) Niuean	(8) Fijian	(9) Other PI	(10) SE Asian	(11) Chinese	(12) Indian	(13) Other Asian	(14) MELAA
(1) NZ European	17.592***	1.893***	1.449***	0.924	0.621***	0.834	0.424***	1.213	0.480**	0.278***	0.450***	0.567***	0.786	2.076***
	(0.759)	(0.112)	(0.094)	(0.101)	(0.084)	(0.133)	(0.079)	(0.290)	(0.143)	(0.062)	(0.055)	(0.093)	(0.270)	(0.204)
(2) Other European	4.168***	11.550***	0.810**	0.771*	0.695*	0.923	0.602*	1.332	0.756	0.895	0.822	1.282	1.318	0.953
	(0.249)	(0.833)	(0.086)	(0.113)	(0.153)	(0.239)	(0.182)	(0.470)	(0.270)	(0.248)	(0.163)	(0.329)	(0.570)	(0.134)
(3) NZ Māori	0.550***	1.952***	129.232***	1.181	2.049***	1.411**	0.730*	1.619*	1.042	0.331***	1.191	0.695*	1.262	0.478***
	(0.028)	(0.104)	(6.755)	(0.141)	(0.255)	(0.193)	(0.129)	(0.411)	(0.259)	(0.097)	(0.181)	(0.144)	(0.409)	(0.058)
(4) Samoan	0.258***	1.264***	0.606***	322.146***	0.833	1.596***	1.462*	1.145	2.849***	0.057***	3.145***	0.595***	0.659	0.237***
	(0.014)	(0.103)	(0.052)	(37.246)	(0.145)	(0.279)	(0.321)	(0.294)	(0.684)	(0.027)	(0.430)	(0.116)	(0.332)	(0.046)
(5) Cook Island Māori	0.348***	0.869	0.785*	1.427**	483.660***	1.256	1.913***	0.881	6.022***	0.236***	1.135	0.738	0.237	0.374***
	(0.026)	(0.103)	(0.104)	(0.216)	(70.503)	(0.264)	(0.477)	(0.464)	(1.452)	(0.103)	(0.258)	(0.192)	(0.252)	(0.103)
(6)Tongan	0.234***	0.854	0.742**	1.372*	1.141	444.389***	2.108**	1.526	1.012	0.094***	0.245***	0.767	0.621	0.383***
	(0.020)	(0.121)	(0.099)	(0.231)	(0.250)	(59.565)	(0.643)	(0.599)	(0.457)	(0.047)	(0.101)	(0.233)	(0.415)	(0.093)
(7) Niuean	0.292***	1.103	0.739***	1.400*	1.795***	2.777***	1,060.290***	1.001	0.917	0.596	1.048	0.357**	0.847	0.587*
	(0.026)	(0.169)	(0.085)	(0.274)	(0.403)	(0.539)	(206.153)	(0.498)	(0.391)	(0.270)	(0.253)	(0.177)	(0.641)	(0.187)
(8) Fijian	0.517***	2.128***	1.220	2.338**	1.336	1.345	1.035	491.647***	2.256		1.718	5.890***	6.253***	0.471
	(0.084)	(0.387)	(0.273)	(0.827)	(0.657)	(0.651)	(0.312)	(154.598)	(3.051)		(0.950)	(1.474)	(4.353)	(0.357)
(9) Other PI	0.436***	1.504	0.739	3.134***	3.367***	0.958	0.416***	5.634***	497.769***	0.347	0.613	0.578		0.174*
	(0.101)	(0.378)	(0.275)	(1.078)	(1.304)	(0.360)	(0.121)	(3.675)	(177.473)	(0.356)	(0.333)	(0.258)		(0.175)
(10) SE Asian	0.310***	0.547***	0.343***	0.228*	0.117***	0.481	0.260**			77.576***	9.444***	0.707	2.940**	0.620**
	(0.033)	(0.120)	(0.100)	(0.177)	(0.053)	(0.284)	(0.175)			(17.789)	(1.804)	(0.234)	(1.547)	(0.137)
(11) Chinese	0.176***	0.351***	0.452***	0.623**	0.319***	0.280***	0.480*	0.946	0.771	3.343***	151.374***	0.521**	1.337	0.726**
	(0.013)	(0.060)	(0.070)	(0.120)	(0.126)	(0.136)	(0.195)	(0.443)	(0.516)	(0.751)	(23.602)	(0.136)	(0.498)	(0.107)
(12) Indian	0.166***	0.556***	0.448***	0.461***	0.354***	0.719	0.277*	16.165***	1.152	0.251***	0.570**	271.978***	2.511**	0.795
	(0.015)	(0.087)	(0.080)	(0.126)	(0.128)	(0.243)	(0.184)	(4.282)	(0.641)	(0.083)	(0.154)	(44.847)	(1.112)	(0.120)
(13) Other Asian	0.348***	0.505***	0.996	0.370		1.216	0.546	0.200**		0.840	2.686***	2.075***	210.174***	0.422***
	(0.041)	(0.106)	(0.290)	(0.255)		(0.573)	(0.544)	(0.142)		(0.319)	(0.620)	(0.558)	(74.211)	(0.114)
(14) MELLA	7.261***	1.482***	2.772***	1.280	1.044	1.332	0.620	2.249	1.516	0.776	1.234	1.804**	1.538	13.831***
	(0.512)	(0.191)	(0.342)	(0.329)	(0.417)	(0.630)	(0.351)	(1.251)	(0.823)	(0.254)	(0.304)	(0.487)	(0.634)	(2.097)

Table 4.2A *continued*

Variables	(1) NZ European	(2) Other European	(3) NZ Māori	(4) Samoan	(5) Cook Island Maori	(6) Tongan	(7) Niuean	(8) Fijian	(9) Other PI	(10) SE Asian	(11) Chinese	(12) Indian	(13) Other Asian	(14) MELAA
MELAAD1	0.169***	0.911	0.468	0.167***										49.560***
	(0.071)	(0.366)	(0.254)	(0.104)										(18.011)
MELAAD2	0.050***	3.362***	0.118*	0.301**						4.233*		1.864		71.690***
	(0.018)	(1.473)	(0.141)	(0.144)						(3.587)		(1.956)		(24.346)
MELAAD3	0.078***	0.822	0.207***	0.457		0.708				0.230**	0.526	0.514		4.781***
	(0.026)	(0.278)	(0.071)	(0.336)		(0.509)				(0.150)	(0.572)	(0.654)		(1.197)
Obs.	126,600	126,600	126,600	126,600	124,800	126,400	126,200	124,800	123,400	126,100	126,400	126,600	126,100	126,500
Pseudo R-squared	0.623	0.278	0.656	0.849	0.823	0.842	0.844	0.582	0.603	0.708	0.810	0.859	0.826	0.272

Notes

The table reports odds ratios.

We have dropped the three ‘not further defined’ ethnic groups. We have combined ‘Middle Eastern’, ‘Latin American’, ‘African’ and the ‘Other’ ethnic groups into one group MELAA. We have also combined the ‘Tokelauan’ with the ‘Other Pacific Islander’ ethnic group into one group ‘Other PI’. Thus, our analysis includes 14 Level 2 ethnic groups instead of 21.

Regressions have been run with inter-censal fixed effects.

*p<0.1; **p<0.05; ***p<0.01

Clustered Standard errors in parenthesis.

Tables 4.2A, 4.2B and 4.2C are reporting results from the same regression. We have broken down the results into different tables according to variables at different levels for easy readability.

Blank cells are where variables have been omitted due to perfect collinearity, usually due to small cell sizes.

MELAAD1 - Individuals who belonged to the combined ‘MELAA and Other’ ethnic group in the period 1991-1996

MELAAD2 - Individuals who belonged to the combined ‘MELAA and Other’ ethnic group in the period 1996-2001

MELAAD3 - Individuals who belonged to the combined ‘MELAA and Other’ ethnic group in the period 2001-2006

Table 4.2B: Clustered Logistic Regression of Current Ethnicity – Effect of Individual and Family-Level Variable

Variables	(1) NZ European	(2) Other European	(3) NZ Māori	(4) Samoan	(5) Cook Island Maori	(6) Tongan	(7) Niuean	(8) Fijian	(9) Other PI	(10) SE Asian	(11) Chinese	(12) Indian	(13) Other Asian	(14) MELAA
Sex	0.892***	0.935**	0.989	1.002	0.808***	1.018	0.914	1.084	1.062	0.903	1.001	1.153	0.885	1.129***
	(0.020)	(0.025)	(0.036)	(0.059)	(0.066)	(0.089)	(0.106)	(0.158)	(0.158)	(0.084)	(0.067)	(0.110)	(0.119)	(0.040)
Age	0.975***	1.006	0.952***	0.914***	0.973	0.949	0.955	1.007	0.924*	0.972	0.983	0.993	1.024	1.063***
	(0.007)	(0.009)	(0.011)	(0.019)	(0.028)	(0.030)	(0.033)	(0.044)	(0.041)	(0.027)	(0.023)	(0.027)	(0.042)	(0.013)
NZ Born	3.958***	0.210***	1.802***	1.450***	1.516***	0.989	1.158	1.151	1.101	0.860	0.968	1.006	0.523***	0.693***
	(0.172)	(0.012)	(0.144)	(0.182)	(0.232)	(0.158)	(0.293)	(0.237)	(0.251)	(0.127)	(0.114)	(0.137)	(0.105)	(0.053)
Ethnicity Mother	2.550***	2.236***	8.586***	13.866***	12.729***	22.729***	17.484***	119.469***	48.204***	27.319***	20.000***	11.890***	43.653***	1.906***
	(0.099)	(0.109)	(0.869)	(2.954)	(3.493)	(7.177)	(7.271)	(86.750)	(43.846)	(6.412)	(4.003)	(2.648)	(19.545)	(0.216)
Ethnicity Father	2.033***	2.872***	8.731***	7.157***	4.821***	10.978***	4.988***	41.613***	5.732***	4.658***	7.772***	9.726***	13.155***	1.800***
	(0.073)	(0.161)	(1.142)	(1.328)	(1.155)	(2.529)	(1.381)	(32.645)	(3.842)	(1.443)	(1.667)	(2.047)	(4.677)	(0.187)
No Female Household	1.524***	1.544***	1.656***	2.410***	2.552***	1.814***	2.038***	2.196***	1.902***	4.066***	1.888***	2.279***	3.055***	1.084
	(0.056)	(0.076)	(0.090)	(0.201)	(0.292)	(0.242)	(0.376)	(0.437)	(0.409)	(0.514)	(0.200)	(0.290)	(0.598)	(0.067)
No Male Household	1.541***	1.419***	1.640***	2.060***	2.098***	2.033***	2.392***	1.818***	1.514**	1.530***	1.423***	2.018***	2.385***	0.992
	(0.044)	(0.053)	(0.071)	(0.168)	(0.224)	(0.235)	(0.360)	(0.320)	(0.259)	(0.198)	(0.137)	(0.235)	(0.445)	(0.043)

Notes

The table reports odds ratios.

We have dropped the three 'not further defined' ethnic groups. We have combined 'Middle Eastern', 'Latin American', 'African' and the 'Other' ethnic groups into one group MELAA. We have also combined the 'Tokelauan' with the 'Other Pacific Islander' ethnic group into one group 'Other PI'. Thus, our analysis includes 14 Level 2 ethnic groups instead of 21.

Regressions have been run with inter-censal fixed effects.

*p<0.1; **p<0.05; ***p<0.01

Clustered Standard errors in parenthesis.

Tables 4.2A, 4.2B and 4.2C are reporting results from the same regression. We have broken down the results into different tables according to variables at different levels for easy readability.

Blank cells are shown where variables have been omitted due to perfect collinearity, usually due to small cell sizes.

'Ethnicity Mother' and 'Ethnicity Father' are dummy variables that are equal to 1 when the parent has stated the same ethnicity as the ethnicity that is given by the dependent variable, and 0 otherwise.

Table 4.2C: Clustered Logistic Regression of Current Ethnicity: The Effect of Neighbourhood-Level Characteristics

Variables	(1) NZ European	(2) Other European	(3) NZ Māori	(4) Samoan	(5) Cook Island Maori	(6) Tongan	(7) Niuean	(8) Fijian	(9) Other PI	(10) SE Asian	(11) Chinese	(12) Indian	(13) Other Asian	(14) MELAA
Entropy	0.799***	1.260***	1.022	1.788***	1.666**	2.020**	2.461***	1.583	1.075	2.253***	0.954	1.263	1.836	1.779***
	(0.052)	(0.081)	(0.105)	(0.322)	(0.341)	(0.602)	(0.755)	(0.599)	(0.517)	(0.699)	(0.190)	(0.282)	(0.782)	(0.222)
NZ European Gr	1.026***	1.030***	1.027**	1.055***	1.043*	1.113***	0.996	1.052	1.080**	1.023	0.995	1.045*	0.971	0.983
	(0.008)	(0.010)	(0.011)	(0.016)	(0.024)	(0.028)	(0.028)	(0.037)	(0.041)	(0.034)	(0.020)	(0.025)	(0.046)	(0.015)
Other European Gr	1.052***	1.041***	1.012	1.062***	1.074**	1.053	1.041	1.041	1.190***	0.985	1.005	1.043	1.036	0.982
	(0.009)	(0.011)	(0.013)	(0.020)	(0.031)	(0.034)	(0.039)	(0.054)	(0.061)	(0.036)	(0.025)	(0.029)	(0.061)	(0.017)
NZ Māori Gr	1.009	1.017**	1.035***	1.042***	1.049***	1.067***	1.009	1.024	1.108***	1.003	1.001	1.051***	0.960	0.992
	(0.005)	(0.007)	(0.009)	(0.012)	(0.018)	(0.022)	(0.022)	(0.028)	(0.033)	(0.024)	(0.016)	(0.020)	(0.038)	(0.011)
Samoan Gr	1.016*	1.023**	1.022*	1.059***	1.014	1.091***	0.986	1.056	1.090**	1.009	0.985	1.031	0.958	0.981
	(0.009)	(0.010)	(0.012)	(0.017)	(0.024)	(0.027)	(0.028)	(0.038)	(0.038)	(0.035)	(0.023)	(0.027)	(0.047)	(0.016)
Cook Island Māori Gr	1.009	1.001	1.019	1.083***	1.099**	1.123***	0.981	1.076	1.045	0.970	1.014	1.021	0.969	0.916***
	(0.013)	(0.016)	(0.021)	(0.032)	(0.041)	(0.045)	(0.039)	(0.055)	(0.069)	(0.055)	(0.039)	(0.043)	(0.093)	(0.029)
Tongan Gr	1.028***	1.007	1.038**	1.052**	1.060**	1.178***	0.993	1.082*	1.040	1.035	0.949*	1.052	0.985	0.996
	(0.010)	(0.015)	(0.019)	(0.022)	(0.031)	(0.038)	(0.035)	(0.048)	(0.055)	(0.048)	(0.026)	(0.037)	(0.067)	(0.027)
Niuean Gr	1.000	1.114***	1.010	1.080**	1.084*	1.103**	1.137**	0.941	1.210**	1.125**	1.095**	1.011	1.065	0.952
	(0.017)	(0.022)	(0.030)	(0.040)	(0.052)	(0.055)	(0.063)	(0.068)	(0.100)	(0.065)	(0.043)	(0.055)	(0.172)	(0.050)
Fijian Gr	0.972	1.183***	1.099	1.013	0.944	0.978	1.100	1.713***	0.976	0.638**	1.020	1.231	0.733	0.861*
	(0.040)	(0.070)	(0.078)	(0.115)	(0.122)	(0.139)	(0.167)	(0.312)	(0.246)	(0.116)	(0.111)	(0.158)	(0.227)	(0.075)
Other Pi Gr	1.015	1.021	1.066	1.095	0.987	0.894	1.244**	1.176	1.029	1.146	0.922	1.018	0.933	0.895
	(0.039)	(0.043)	(0.050)	(0.096)	(0.091)	(0.111)	(0.112)	(0.163)	(0.147)	(0.139)	(0.128)	(0.116)	(0.170)	(0.081)
SE Asian Gr	0.988	1.024	1.030	1.022	1.081	1.050	1.002	1.030	1.020	1.139**	1.045	1.052	0.973	1.046
	(0.017)	(0.021)	(0.027)	(0.040)	(0.057)	(0.067)	(0.077)	(0.087)	(0.082)	(0.061)	(0.048)	(0.073)	(0.078)	(0.029)
Chinese Gr	1.022**	1.032***	1.007	1.046**	1.031	1.100***	0.987	0.984	1.159***	0.971	1.031	1.075***	0.991	0.967*
	(0.009)	(0.011)	(0.014)	(0.022)	(0.030)	(0.039)	(0.037)	(0.046)	(0.055)	(0.037)	(0.024)	(0.029)	(0.053)	(0.016)
Indian Gr	1.032***	1.020*	1.037***	1.087***	1.063**	1.109***	0.997	1.039	1.105*	1.028	1.011	1.108***	0.932	0.971*
	(0.009)	(0.012)	(0.014)	(0.023)	(0.028)	(0.037)	(0.037)	(0.048)	(0.061)	(0.036)	(0.026)	(0.032)	(0.050)	(0.015)

Table 4.2C *continued*

Variables	(1) NZ European	(2) Other European	(3) NZ Māori	(4) Samoan	(5) Cook Island Maori	(6) Tongan	(7) Niuean	(8) Fijian	(9) Other PI	(10) SE Asian	(11) Chinese	(12) Indian	(13) Other Asian	(14) MELAA
Other Asian Gr	1.025	0.998	1.009	1.006	0.996	1.186**	0.896	1.113	1.085	1.051	1.047	0.999	0.943	0.983
	(0.017)	(0.023)	(0.031)	(0.052)	(0.071)	(0.083)	(0.094)	(0.106)	(0.110)	(0.065)	(0.050)	(0.065)	(0.077)	(0.026)
MELAA Gr	1.091***	1.051***	1.031*	1.093***	1.059*	1.144***	1.061	1.056	1.087	1.032	1.007	1.041	0.969	0.876***
	(0.011)	(0.017)	(0.018)	(0.030)	(0.032)	(0.044)	(0.044)	(0.052)	(0.081)	(0.044)	(0.026)	(0.039)	(0.059)	(0.018)
MELAA Gr D1	0.979	0.971	0.904	1.427**	1.601	0.607	1.532*	1.425	0.697	2.527***	0.925	1.748***	0.455	0.981
	(0.101)	(0.084)	(0.169)	(0.246)	(0.692)	(0.347)	(0.338)	(0.610)	(0.667)	(0.495)	(0.336)	(0.333)	(0.477)	(0.498)
MELAA Gr D2	0.893	0.913	1.100	1.157	1.194	1.445**	1.259	1.258	0.852	0.869	1.379*	1.154		2.093***
	(0.072)	(0.137)	(0.121)	(0.155)	(0.245)	(0.215)	(0.262)	(0.693)	(0.374)	(0.293)	(0.255)	(0.238)		(0.381)
MELAA Gr D3	1.000	0.997	1.046	1.107	0.973	0.815*	0.969	1.038	0.693*	0.933	0.899	1.024	1.076	0.986
	(0.042)	(0.056)	(0.059)	(0.093)	(0.109)	(0.095)	(0.166)	(0.156)	(0.152)	(0.122)	(0.082)	(0.116)	(0.209)	(0.061)

Notes

The table reports odds ratios.

We have dropped the three ‘not further defined’ ethnic groups. We have combined ‘Middle Eastern’, ‘Latin American’, ‘African’ and the ‘Other’ ethnic groups into one group MELAA. We have also combined the ‘Tokelauan’ with the ‘Other Pacific Islander’ ethnic group into one group ‘Other PI’. Thus, our analysis includes14 Level 2 ethnic groups instead of 21.

*p<0.1; **p<0.05; ***p<0.01

Clustered Standard errors in parenthesis.

Tables 4.2A, 4.2B and 4.2C are reporting results from the same regression. We have broken down the results into different tables according to variables at different levels for easy readability.

‘Gr’ refers to group proportion. For example ‘Tongan Gr’ refers to ethnic group proportion of Tongan group in the area unit an individual resides.

Blank cells are where variables have been omitted due to perfect collinearity, usually due to small cell sizes.

MELAA Gr D1 - Group proportion of combined ‘MELAA and Other’ ethnic group in area unit in the period 1991-1996

MELAA Gr D2 - Group proportion of combined ‘MELAA and Other’ ethnic group in area unit in the period 1996-2001

MELAA Gr D1 - Group proportion of combined ‘MELAA and Other’ ethnic group in area unit in the period 2001-2006

Individual-level Characteristics

We find that ethnicity in the previous Census is statistically significant and positively related to the choice of each ethnic identity (Table 4.2A). Adolescents are highly likely to identify with the same ethnicity as they were identified with by their parents in the previous Census. The odds ratios are highest for this variable for all groups. For example, holding all other independent variables constant, the odds of choosing Niuean when they were recorded as Niuean in the previous census are about 1000 times the odds of choosing Niuean when they were not recorded as Niuean in the previous census.

We observe that odds of choosing an ethnicity is affected by whether or not the adolescent was also previously affiliated with other ethnic groups, both positively and negatively. These are the off-diagonal elements in Table 4.2A. Adolescents belonging to any of the Level 2 ethnic categories under Pacific People (for example, Samoan, Cook Island Māori, Tongan etc.) as well as Asian (for example, Indian and Chinese) ethnic groups in the previous census, had lower than average odds of identifying themselves as NZ Māori or New Zealand European in the current census. As the inter-censal changes might affect the results for the combined 'MELAA and Other' ethnic group, we include interaction terms in our model.[68] There was an increase in the 'Other' ethnic group in 2006, due to the fact that New Zealand European individuals were most likely to call themselves 'New Zealander', which was then included in the 'Other ethnic' group in 2006. The odds of choosing NZ European, Other European and combined MELAA and Other ethnic groups increase if the individual belonged to the combined 'MELAA and Other' category in the previous census.[69]

The odds of being a Samoan, Cook Island Māori, Tongan, Fijian, Tokelauan or Niuean in the current census increases, if the child belonged to the ethnic groups under the Level 2 Pacific People category in the last census. However, the odds of choosing Cook Island Māori in the current census increases if the parent of the child reported that the child was NZ Māori in the previous census. Cook Island Māori people have a long history of inter-marriage with NZ Māori, and they have been to some extent absorbed into NZ Māori communities (Hooper 1961).

---

[68] By interacting the binary variable representing individuals belonging to the 'MELAA and Other' category with binary variables representing presence in each inter-censal period.

[69] We also ran the same regression for each inter-censal period separately. We found that the odds of choosing NZ European increases if a person belonged to the combined 'MELAA and Other' ethnic group in the previous census for the period 2006-2013, but not for 2001-2006.

Alternatively, our results might also be due to the fact that some New Zealand Māori can trace their ancestry to the Cook Islands (Walrond 2015).

We observe positive and significant complementarity between Chinese/ South East Asian. This might be because the Chinese foreign-born proportion in New Zealand come from China (51 percent), Malaysia (6 percent), Taiwan (5 percent) and Hong-Kong (4 percent) (New Zealand Ministry for Culture and Heritage 2015).

We also observe positive and significant complementarity between Chinese/Samoan, and Indian/ Fijians. The complementarity between Chinese and Samoan ethnic identity can be attributed to the increasing cultural assimilation through intermarriages (Wai 2015). Coming to the two way complementarity between Fijian/ Indians, might be due to the fact the ancestors of the Fiji Indians in New Zealand were Indians who arrived in Fiji as labourers and gradually worked up the social and economic ladder, eventually settling in New Zealand after the political turmoil of 1987 and 2000 in Fiji (Friesen *et al.* 2005, Pio 2007, Leckie 2015). Moreover, Indo-Fijians or Fiji Indian ethnic group is one of the largest Fijian groups in New Zealand (Swarbrick 2015. This also might simply be because Fiji Indian, Fijian Indian and Indo-Fijian individuals might choose both Fijian and Indian ethnic group options.

Sex is statistically significant for some ethnic groups, but not others (Table 4.2B). Ceteris paribus, males are statistically significantly less likely to choose to affiliate with a NZ European (OR=0.89), Other European (OR=0.94), or Cook Island Māori (OR=0.81) ethnicity, but more likely to choose the MELAA or Other ethnicity (OR=1.13).

For age, the odds ratios are all close to one, varying from 0.91 (Samoan) to 1.06 (MELAA and Others) (Table 4.2B). This is not surprising because the sample of observations is already in a narrow age range (18-22). Age is statistically significant for NZ European (OR=0.98), NZ Māori (OR=0.95), Samoan (OR=0.91) and MELAA and others group (OR=1.06). NZ European, NZ Māori and Samoan are the most common ethnicities (see Table 4.1). Older adolescents have lower odds of choosing these common ethnicities than younger adolescents, because they may select an ethnicity away from that assigned by their parents as they identify with the ethnicity of a partner or the group they socialise with.

For place of birth (Table 4.2B), we find that adolescents who are born in New Zealand, have higher odds of reporting their ethnic identity as NZ European (OR=3.96), NZ Māori (OR=1.80), Samoan (OR=1.45) or Cook Island Māori (OR=1.52). In 2013, almost two-thirds of Pacific people living in New Zealand, were born in New Zealand, and 77.4 percent of the individuals living in New Zealand, who reported as Cook Islands Māori are New Zealand born (Statistics New Zealand 2014). People living in New Zealand and belonging to these ethnic groups are more likely to be born in New Zealand than any other country in the world. In 2013, around 81 percent of the adolescents of Samoan ethnicity, living in Auckland, were New Zealand born. Thus, it is likely that the odds of affiliating themselves with these ethnic groups increase if the adolescent is born in New Zealand. On the other hand, those with an Other Asian or MELAA ethnicity are mostly recent migrants to New Zealand. Hence the odds of New Zealand born selecting these ethnicities are relatively low (and significant at the 1 percent level).

Family-level Characteristics

For all ethnic groups, we find that parents' ethnicity, has a positive and significant effect (Table 4.2B). The odds ratios are high for these variables. For example, the log odds of affiliating themselves as Tongan are 22.7 times higher for those with a Tongan mother (than those without a Tongan mother).

For most groups, we find that the mother's ethnic identity has a larger coefficient than the ethnic identity of the father.[70] This is in line with the extant literature (Salisbury 1970, Nelsen 1990, Cholil 2009) that children's cultural identity is passed more along the maternal than the paternal side.

Neighbourhood-level Characteristics

The entropy index measures the diversity of the area unit the adolescent is located in. It can be seen from Table 4.2C that a young person is less likely to identify with being NZ European in an ethnically diverse area, (OR = 0.8). On the other hand, for those who identify with being Tongan, Niuean and SE Asian, the effect of diversity of the area unit has an odds ratio that is greater than two. Hence Table 4.2c shows that adolescents residing in more ethnically diverse areas have higher odds of choosing many of the ethnicities, with the exception of the New Zealand European group. New Zealand Europeans are the numerically dominant group in Auckland (59.3 percent in 2013, see Statistics New Zealand 2014). Thus, it might be that, when adolescents see more diversity around them, they feel more comfortable being different and

---

[70] Except for NZ Māori. The coefficient size for NZ Māori mother is (8.59) and father is (8.73). They are statistically significantly different with a p-value <0.01.

adopting the non-dominant ethnicity (perhaps in addition to the dominant ethnicity). And, if they are in a less diverse area, adopting the non-dominant ethnicity would mark them out as different, which imposes a cost on them. So, in that case, they are less likely to adopt the non-dominant ethnicity.

Not surprisingly, the odds of choosing an ethnicity generally increase in area units where the ethnic group makes up a higher proportion of the ethnic mix. The one exception is the combined 'MELAA and Others' group (Table 4.2C). This can be seen from the highlighted main diagonal of this block of odds ratios. We also observe that odds of choosing an ethnicity is often affected by the ethnic group sizes of some other ethnic groups as well. This effect is shown by the off-diagonal elements in Table 4.2C. The statistically significant effects are mostly above average odds, in some cases similar to the off-diagonal effects shown in Table 4.2A. This reinforces the existence of complementarity between ethnic groups such as among Pacific Island communities and, interestingly, between New Zealand Māori and Indian ethnicities. Generally, they represent the co-location of some ethnic groups at the area unit level (Mondal *et al.* 2020).

4.6 Conclusion

The main objective of this paper is to identify the determinants of ethnic identity choices among adolescents in Auckland. We link adolescents between consecutive Censuses, where in the first Census their parents are likely to have recorded the adolescent's ethnicity, and in the second Census the adolescents are likely to have recorded their own ethnicity. To the extent that this assumption holds, we are capturing the ethnic affiliation choice at the time that the adolescent is first making this choice for themselves, that is, when they reach the stage of ethnic identity achievement as described by Phinney (1989, 1990).

We find significant relationships between the adolescents' ethnic identity and the ethnic identity assigned to them by their parents five to seven years previously, their age, sex, having been born in New Zealand, ethnic diversity of their area unit (suburb), and ethnic group-proportions in the area unit they live in, as well as their parents' ethnic identity. The results differ somewhat for different ethnicities, but we also identify patterns of complementarity between ethnicities and ethnic groups that accord with other research.

A limitation of this work might be that we did not link the parents' ethnicity directly to that of the adolescents. Instead we imputed the parental ethnicity variables. With our imputed parental ethnicity variables, we don't know the actual ethnicities of each adolescent's parents and so this is measured with error. However, any resulting measurement error is likely to bias the coefficients on parent's ethnicity towards zero - that is, the results will be over-conservative for this variable. Including only the households with no more than one female and one male adult in the analysis reduces this measurement error. Similarly, siblings may have an effect on an adolescent's choice of ethnic identity, but we did not control for this. As with parents, these data are available for a subset of the population. Another limitation of this work is that we assumed that each transition stage is five years, to match with the census timeframes. The time periods used in the analysis are discrete and not continuous. We are not claiming that the decisions or changes occur at exact five-year intervals. Instead, we assume that the decision process occurs at some point within the inter-censal period, and the final decision is only observed in the census years.

Our analysis could be extended using the available data on parent's ethnicity and/or for siblings, present for a subsample of the census data, to check the consistencies with the results reported in this paper. Moreover, we combined the results for 'MELAA' and 'Other' ethnic groups, due to the small number of adolescents reporting these ethnicities. Future research could investigate these ethnicities in more detail, perhaps making use of qualitative methods given the small sample size. We were also unable to control for effects arising from the adolescent's 'peer group' as there are no variables in the Census that could capture peer group effects (but see, for example, Jugert *et al.* 2019). Despite these limitations, our work represents the most comprehensive investigation in New Zealand to date of the effects of individual, household and community-level variables on adolescents' ethnic identity choices.

Our study contributes to a small but growing literature on adolescent ethnic identity development in New Zealand and elsewhere. There have been several past studies about ethnic-self-identity formation among adolescents (Phinney 1989, 1990, Phinney and Chavira 1992) in relation to theories regarding different stages of identity formation. These studies are mainly based on U.S. data, so our study contributes in a novel context. Moreover, our study provides a baseline for future analysis in exploring the influence of changes in social circumstances on self-identified ethnicity over time (that is, when moving from child to adolescent).

Understanding ethnic mobility is important, given the increasing ethnic diversity of Western countries like New Zealand. Increases in ethnic mixing and intermarriage will lead to increases in multiple ethnicity and potentially to increases in ethnic mobility. This study presents a novel attempt to facilitate understanding of these changes using linked inter-censal data for adolescents.

References

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715–753**.**

Brown, P., Callister, P., Carter, K., & Engler, R. (2010). Ethnic mobility: Is it important for research and policy analysis? *Policy Quarterly*, *6*(3), 45–51.

Brown, P. and Gray, A. (2009). Inter-ethnic mobility between the 2001 and 2006 Censuses: The statistical impact of the New Zealander response. In Statistics New Zealand, *Final report of a review of the official ethnicity standard 2009* (pp 27-36). Wellington, New Zealand: Statistics New Zealand.

Beine, M., & Parsons, C. (2015). Climatic factors as determinants of international migration. *The Scandinavian Journal of Economics*, *117*(2), 723–767.

Benson, A.R., Kumar, R., & Tomkins, A. (2018). A Discrete Choice Model for subset selection. *Association for Computing Machinery*, 37-45. https://doi.org/10.1145/3159652.3159702

Bhat, C. R., & Guo, J. (2004). A mixed spatially correlated logit model: formulation and application to residential choice modeling. *Transportation Research Part B: Methodological*, *38*(2), 147–168.

Bhat, C. R., & Guo, J. Y. (2007). A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B: Methodological*, *41*(5), 506–526.

Brunsma, D. L. (2005). Interracial families and the racial identification of mixed- race children: Evidence from the early childhood longitudinal study. *Social Forces*, *84*, 1131–1157.

Callister, P., R. Didham & D. Potter (2005) 'Ethnic intermarriage in New Zealand. *Official Statistics Research Series*, (1), 1–73.

Callister, P., R. Didham & A. Kivi (2008). Who are we? The conceptualisation and expression of ethnicity. *Official Statistics Research Series*, 4, 1–56.

Carter, K. N., Hayward, M., Blakely, T., & Shaw, C. (2009). How much and for whom does self-identified ethnicity change over time in New Zealand? Results from a longitudinal study. *Social Policy Journal of New Zealand*, 36, 32–45.

Casey, T., & Dustmann, C. (2010). Immigrants' identity, economic outcomes and the transmission of identity across generations. *The Economic Journal*, *120*(542), F31–F51.

Cholil, S. (2009). The politico-religious contestation: hardening of the Islamic law on Muslim-non- Muslim marriage in Indonesia. In G. W. Jones, C.H. Leng & M. Mohamad (Eds.), *Muslim-Non-Muslim marriage: Political and cultural contestations in Southeast Asia* (pp. 139-158). Singapore, Malaysia: Institute of Southeast Asian Studies.

Coope, P. and A. Piesse (2000). *1991–1996 Intercensal consistency study*. Wellington, New Zealand: Statistics New Zealand.

de Snyder, N. S., Lopez, C. M., & Padilla, A. M. (1982). Ethnic identity and cultural awareness among the offspring of Mexican interethnic marriages. *The Journal of Early Adolescence*, *2*(3), 277–282.

Didham, R. (2016). Ethnic mobility in the New Zealand census, 1981-2013: A preliminary look. *New Zealand Population Review*, *42*, 27–42.

Eschbach, K., & Gómez, C. (1998). Choosing Hispanic identity: Ethnic identity switching among respondents to high school and beyond. *Social Science Quarterly*, *79*(1), 74–90.

Fitzpatrick, K. M., & Hwang, S.-S. (1992). The effects of community structure on opportunities for interracial contact. *Sociological Quarterly*, *33*(1), 51–61.

Frideres, J., & Goldenberg, S. (1982). Ethnic identity: Myth and reality in Western Canada. *International Journal of Intercultural Relations*, *6*(2), 137–151. https://doi.org/10.1016/0147-1767(82)90044-X

Friesen, W., Murphy, L., & Kearns, R. (2005). Spiced-up Sandringham: Indian transnationalism and new suburban spaces in Auckland, New Zealand. *Journal of Ethnic and Migration Studies*, *31*(2), 385–401. https://doi.org/10.1080/1369183042000339981

Hooper, A. (1961). The migration of cook islanders in New Zealand. *The Journal of the Polynesian Society*, *70*(1), 11–17.

Howard, S. & R. Didham (2005). Ethnic intermarriage and ethnic transference amongst the Maori population. *Official Statistics Research Series*, (1), 1–23.

Hwang, S.S., & Murdock, S. H. (1991). Ethnic enclosure or ethnic competition: Ethnic identification among Hispanics in Texas. *The Sociological Quarterly*, *32*(3), 469–476.

Jugert, P., Leszczensky, L., Pink, S. (2019) Differential influence of same- and cross-ethnic friends on ethnic-racial identity development in early adolescence. *Child Development*. https://doi.org/10.1111/cdev.13240

Kickett-Tucker, C.S. (2008). How aboriginal peer interactions in upper primary school sport support aboriginal identity. *The Australian Journal of Indigenous Education*, *37*(1), 138–151

Kickett-Tucker, C. S. (2009). Moorn (black)? Djardak (white)? How come I don't fit in mum? Exploring the racial identity of Australian aboriginal children and youth. *Health Sociology Review*, *18*(1), 119–136. https://doi.org/10.5172/hesr.18.1.119

Kukutai, T. (2007). White mothers, brown children: Ethnic identification of Māori-European children in New Zealand. *Journal of Marriage and Family*, *69*(5), 1150–1161.

Kukutai, T. (2008). *Ethnic self-prioritisation of dual and multi-ethnic youth in New Zealand*. Wellington, New Zealand: Statistics New Zealand.

Kukutai, T., & Didham, R. (2009). In search of ethnic New Zealanders: National naming in the 2006 census. *Social Policy Journal of New Zealand*, (36), 46–62.

Leckie, J. (2015). *Fijians – Te Ara Encyclopedia of New Zealand*. Retrieved from https://teara.govt.nz/en/21445

Lee, J., & Bean, F. (2010). *The diversity paradox*. New York, NY: Russell Sage Foundation.

McFadden, D. L. (1984). Chapter 24: Econometric analysis of qualitative response models. In Z. Griliches & M. D. Intriligator (Eds.), *Handbook of econometrics* (Vol. 2, pp. 1395–1457). Amsterdam, Netherlands: North-Holland.

Mondal, M., Cameron, M.P., & Poot, J. (2020). Cultural and economic residential sorting of Auckland's population, 1991-2013: An entropy approach. *Journal of Geographical Systems*. https://doi.org/10.1007/s10109-020-00327-1

Mondal, M., Cameron, M.P., & Poot, J. (forthcoming). Group-size bias in the measurement of residential sorting. In R. Patuelli & S. Suzuki (Eds.), *A broad view of regional science - Essays in honor of Peter Nijkamp*. Berlin, Germany: Springer Nature.

Mowen, T. J., & Stansfield, R. (2016). Probing change in racial self-identification: A focus on children of immigrants. *Sociology of Race and Ethnicity*, *2*(3), 323–337.

Nelsen, H. M. (1990). The religious identification of children of interfaith marriages. *Review of Religious Research*, *32*(2), 122–134.

New Zealand Ministry for Culture and Heritage. (2015). *Te Ara- The Encyclopedia of New Zealand is the complete guide to our peoples, environment, history, culture and society*. Retrieved from https://teara.govt.nz/en/

Phinney, J. S. (1989). Stages of ethnic identity development in minority group adolescents. *The Journal of Early Adolescence*, *9*(1–2), 34–49.

Phinney, J. S. (1990). Ethnic identity in adolescents and adults: Review of research. *Psychological Bulletin*, *108*(3), 499–514.

Phinney, J. S. (1992). The multigroup ethnic identity measure: A new scale for use with diverse groups. *Journal of Adolescent Research*, *7*(2), 156–176.

Phinney, J. S., & Chavira, V. (1992). Ethnic identity and self-esteem: an exploratory longitudinal study. *Journal of Adolescence*, *15*(3), 271–281.

Pio, E. (2007). Inspirational cameos. *Journal of Immigrant & Refugee Studies*, *5*(1), 71–94. https://doi.org/10.1300/J500v05n01_05

Qian, Z. (2004). Options: Racial/ethnic identification of children of intermarried couples. *Social Science Quarterly*, *85*(3), 746–766.

Rosenthal, D. A., & Hrynevich, C. (2007). Ethnicity and ethnic identity: A comparative study of Greek, Italian, and Anglo-Australian adolescents. *International Journal of Psychology*, *20*(3-4), 723–742. https://doi.org/10.1080/00207598508247566

Rumbaut, R. G. (1994). *The crucible within: Ethnic identity, self-esteem, and segmented assimilation among children of immigrants*. Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=1879626

Saenz, R., Hwang, S.S., Aguirre, B. E., & Anderson, R. N. (1995). Persistence and change in Asian identity among children of intermarried couples. *Sociological Perspectives*, *38*(2), 175–194.

Salisbury, W. (1970). Religious identity and religious behaviour of the sons and daughters of religious intermarriage. *Review of Religious Research, 11*(2), 128-135.

Salmond, C. and Crampton, P. (2002). *NZDep2001 Index of Deprivation*. Wellington, New Zealand: University of Otago.

Simpson, L., Jivraj, S., & Warren, J. (2016). The stability of ethnic identity in England and Wales 2001–2011. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *179*(4), 1025–1049. https://doi.org/10.1111/rssa.12175

Statistics New Zealand. (2004*). Ethnic population projections: Issues and trends*. Retrieved from http://archive.stats.govt.nz/browse_for_stats/population/estimates_and_projections/ethnic-pop-projections-issues-and-trends.aspx

Statistics New Zealand. (2007). *Profile of New Zealander responses, ethnicity question: 2006 Census*. Retrieved from http://archive.stats.govt.nz/Census/about-2006-census/profile-of-nzer-responses-ethnicity-question-2006-census.aspx

Statistics New Zealand. (2009). *Final report of a review of the official ethnicity standard: Statistics New Zealand*. Wellington. Retrieved from http://archive.stats.govt.nz/methods/data-integration/data-integration-projects/intercensal-ethnic-mobility-2001-2006.aspx

Statistics New Zealand. (2013a). *2013 Census quickstats about a place*. Retrieved from www.stats.govt.nz

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

Statistics New Zealand. (2014a). *2013 Census quickstats about culture and identity*. Retrieved from www.stats.govt.nz

Statistics New Zealand (2014b). *Linking censuses: New Zealand longitudinal census 1981–2006.* Retrieved from 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 from http://www.stats.govt.nz/Census/2001-census-data/change-in-ethnicity-question.aspx. Accessed January 11, 2020

Stephan, C. W., & Stephan, W. G. (1989). After intermarriage: Ethnic identity among mixed-heritage Japanese-Americans and Hispanics. *Journal of Marriage and Family*, *51*(2), 507–519.

Swarbrick, N. (2015). *Indians: Indian communities - Te Ara Encyclopedia of New Zealand*. Retrieved from https://teara.govt.nz/en/21611

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

Tizard, B., & Phoenix, A. (1995). The Identity of mixed parentage adolescents. *Journal of Child Psychology and Psychiatry*, *36*(8), 1399–1410.

Train, K.E. (2009). *Discrete choice methods with simulation*. Cambridge, United Kingdom: Cambridge University Press.

Tsang, A. K. T., Irving, H., Alaggia, R., Chau, S. B. Y., & Benjamin, M. (2003). Negotiating ethnic identity in Canada: The case of the 'satellite children'. *Youth & Society*, *34*(3), 359–384. https://doi.org/10.1177/0044118X02250124

Umaña -Taylor, A. J., Gonzales-Backen, M. A., & Guimond, A. B. (2009). Latino adolescents' ethnic identity: Is there a developmental progression and does growth in ethnic identity predict growth in self-esteem? *Child Development*, *80*(2), 391–405.

Wai, T. A. M. L. (2015). *Reflections on the experiences of the Chinese community in Samoa*. Retrieved from https://www.wgtn.ac.nz/chinaresearchcentre/programmes-and-projects/china-symposiums/china-and-the-pacific-the-view-from-oceania/9-Ming-Leung-Wai-Chinese-Experience-in-Samoa-paper-sp2-15-REAL-FINAL-2.pdf

Walrond, C. (2015). *Cook Islanders - Te Ara Encyclopedia of New Zealand*. Retrieved from https://teara.govt.nz/en/21199

Waters, M. C. (1989). The everyday use of surname to determine ethnic ancestry. *Qualitative Sociology*, *12*(3), 303–324.

Chapter 5: Projecting the spatial distribution of ethnic groups in Auckland: Development of a spatial dynamic microsimulation model

5.1 Introduction

The residential choice preferences of individuals constitute very important topic of study, as residential household location is one of the key components of urban dynamics. The literature on residential sorting suggests that individuals choose where to locate based on a variety of factors (Uyeki 1964; Schelling 1971; Duncan and Duncan 1955). There are patterns of residential sorting observed on the basis of ethnicity and race (Schelling 1971; Ho and Bedford 2006; Johnston et al. 2011; Mondal et al. 2020), educational qualification (Farley 1977; Denton and Massey 1988; Domina 2006), occupational status (Duncan and Duncan 1955; Simkus 1978), and income (Fischer 2003).

There have been many studies linking ethnic diversity with residential sorting. Schelling (1971) explained that all factors leading to residential sorting are interrelated. An individual looking for houses is usually informed about available housing choices by people who they are in close contact with. Individuals prefer to stay in close contact with people with whom they share similar preferences. Networks are formed mostly in terms of ethnicity, religion and language use, for easier communication and trust. This results in people clustering together with others of the same ethnicity, resulting in residential sorting. Also, individuals with similar level of education have similar types of jobs, resulting in similar incomes. Individuals with similar types of income may cluster together due to having similar housing affordability and house prices and rents are spatially clustered too due to the bid rent model (Alonso, 1964; McCann 2013). Mondal et al. (2020) captured these mutually reinforcing aspects of sorting by income, education and occupation in Auckland, New Zealand, by calculating a combined economic index of residential sorting. However, they found that ethnic residential sorting is much more prevalent in Auckland than economic sorting throughout their study period (1991-2013).

Most of the research studying ethnic identity transitions, ethnic residential sorting and the dynamics of these processes looks either backwards or at the present (Rees et al. 2017). It is crucial to understand and measure existing residential sorting patterns and their dynamics, to meaningfully quantify demand for current housing, local transport, and infrastructural and community facilities, as well as services such as education and healthcare (Mondal et al. 2020; Mondal et al. forthcoming). It is equally important to have knowledge about future residential sorting patterns to enhance the efficiency and efficacy in planning for future public services and housing demands (Cameron and Poot 2019). However, there are only a limited number of extant studies relating to projecting the future ethnic diversity or future ethnic residential sorting patterns for local areas (Rees et al. 2017).

Our understanding of residential sorting, and its causes and impacts, is relatively limited (Bruch and Mare 2006). Broader understanding of changing residential sorting patterns requires an examination at different spatial levels, as the different geographic scales portray different dimensions of residential sorting (Reardon et al. 2019). Urban households are likely to take spatial features at different spatial scales into account when deciding on their residential location, right down to the neighbourhood level. Hence projections of ethnic diversity also require assessing diversity at the neighbourhood level (O'Sullivan 2009), making the task of ethnic population projection more difficult Consequently, the way neighbourhoods are represented and conceptualised might have an impact on the outcomes of residential sorting models. However, projecting future ethnic diversity or residential sorting at a neighbourhood scale is additionally challenging, due to high data requirements. Moreover, methods for small-area population projections are currently under-developed (Cameron and Cochrane 2017).

In this paper, we describe and evaluate a dynamic spatial microsimulation model (MSM) of the Auckland region that captures future ethnic diversity at a fine (i.e. small area) spatial scale and with the maximum feasible disaggregation of ethnic groups, and which is constructed from actual census data from 1991-2006. The 2011 census was delayed until 2013 due to a large earthquake in Christchurch making the inter-censal gap seven years (instead of five) and so we did not use data from the actual census data from 2006-2013. Using the 1996-2001 New Zealand Linked Census data, we use the model to project the ethnic distribution in 2006 in Auckland, New Zealand, i.e. out of sample and incorporating changes in ethnicity-specific population growth along with ethnic and spatial mobility. We validate our model by comparing our simulated results to the actual 2006 census data. This work represents the first attempt to develop a spatial dynamic MSM to project the future ethnic spatial distribution at such a fine spatial (area unit) scale in New Zealand. This model also includes more disaggregated ethnic groups than those used in previous studies (in New Zealand as well as elsewhere), which more adequately captures the heterogeneity among the choices and preferences within the broad ethnic groups (Mondal et al. 2020). We develop and run our model in Stata, which is in itself a novel approach to dynamic microsimulation modelling. As Stata is available inside the secured Statistics New Zealand Datalab, we could run our model without anonymising the data,[71] which prevents any bias arising due to the anonymisation. Moreover, by using Stata inside the datalab, we were also able to use the entire 1996-2001 Auckland population as our base population for our model, rather than a sample of the population.

The remainder of the paper is organised as follows. Section 5.2 presents detailed information about different types of MSMs and how they have been used in previous research. Sections 5.3 and 5.4 describes the data and the methods we employ respectively. Section 5.5 describes the results and the validation of the MSM model, and Section 5.6 concludes.

5.2 Literature Review

Microsimulation is a method that can be used to estimate and project populations and their attributes. *Micro* refers to individual units, e.g. people (Mot 1992), households (Rogers et al. 2014) or firms (Finta 1987). Simulation refers to the process by which attributes are assigned to those units (Lomax and Smith 2017). The unit of analysis in the MSM is referred to as the unit record. The base sample of a MSM can come either from a survey or can be synthesised from various data sources (Zaidi and Rake 2001). MSMs have previously been used for tax-benefit analysis (Lambert et al. 1994; Spielauer 2011), projecting future socio-economic development trends under current (or forecast) policies (Favreault and Smith 2004; Harding 2007), modelling lifetime earning distributions (Smith et al. 2007; Holmer et al. 2014), and in studies of wealth accumulation (Caldwell et al. 1998). MSMs have also been used to assess the future performance and sustainability of long-term public programmes such as pensions, healthcare and educational financing (Goldman et al. 2009; Rowe and Wolfson 2000; Wolfson and Rowe 2013).

In New Zealand there are a range of government and non-government organisations that produce population projections at national (Statistics New Zealand 2016), sub-national/district level (Cameron and Cochrane 2016a), or at small area level (Cameron and Cochrane 2016b). Cameron and Poot (2019) calculate and discuss ethnic population projections for New Zealand regions with the cohort change ratio method. Statistics New Zealand also generates ethnic population projections down to the Territorial Authority level, but only for the highly aggregated (one digit) ethnic classification. The MSM that is the developed in this chapter is

[71] A confidentiality rule required by Statistics New Zealand to take any data out of the secured Statistics New Zealand Datalab.

the first model that generates ethnic projection at a disaggregated level of ethnicities and for small spatial areas (area units).

5.2.1 Types of MSMs

MSMs differ in terms of the overall setup of the model (static or dynamic), the estimation of transitional probabilities, exclusion or inclusion of behavioural responses of the micro-units (arithmetical or behavioural), treatment of time (discrete/continuous), and whether they are explicitly spatial or not.

Static MSMs usually take a cross-section of the population at a specific point in time, and measure the immediate effects of policy changes without modelling any of the specific processes that result in changes over time (Lambert et al. 1994; Spielauer 2011). This type of MSM has been mainly used to evaluate tax-benefit systems (Pechmen and Okner 1974) and also to analyse the redistribution impacts of reforming existing tax systems (Paulus et al. 2009). For example, Immervoll et al. (2007) used a static MSM to estimate marginal and participation tax rates[72] in response to increasing traditional welfare and the introduction of in-work benefits in 15 countries of the European Union in 1998. Eggink et al. (2016) used a static MSM to forecast the use of publicly funded long-term elderly care in Netherlands from 2008 to 2030.

In contrast, *dynamic MSMs* are able to simulate changes over time for a population, by ‘ageing’ unit records based on the probabilities of numerous real-life events occurring. This type of model can therefore estimate the effects of policies in both the long-term and the short-term (Lomax and Smith 2017). For example, Favreault and Smith (2004) designed DYNASIM (Dynamic Simulation of Income Model) III in order to analyse the long-term distributional consequences of retirement and ageing from 1992 to 2040 in the US. In the UK, PENSIM is a national dynamic microsimulation model designed to study the impact of policy changes on the income distribution of pensioners, for 1935-1985 birth cohorts for the period until 2030 (Hancock et al. 1992; Holmer et al. 2014).

Dynamic MSMs can be probabilistically dynamic or implicitly dynamic. *Probabilistic dynamic models* use event probabilities to project the characteristics of each unit record in the simulated database into the future. These event probabilities (or transition probabilities) are probabilities that govern the change in the variables studied from one time period to the next. For example,

---

[72] Difference between the current household taxes and benefits and the household taxes and benefits when the individual earning is set to zero, divided by the earnings (Immervoll et al. 2007).

Ballas et al. (2005a) used a probabilistic model to project population change from 1991-1996 and between 1996-2002 at the District Electoral Division (DED) level in Ireland. Probabilistic dynamic MSMs require modellers to undertake the difficult task of determining the interdependencies between individual attributes and events, and so they require high quality suitable data, which are seldom available (Ballas et al. 2005b). In contrast, *implicitly dynamic models* use independent small area projections and apply static simulation techniques to create small area microdata. For example, Ballas et al. (2005b) used data from the 1971, 1981 and 1991 British population censuses to estimate small area data for 2001, 2011 and 2021 in Wales. They then used these estimates, in combination with national survey data, to simulate future trends in car ownership, demography, and employment at the small area level.

Arithmetical MSMs are generally used to simulate distributional effects in response to changes in taxes, benefits and wage changes. This type of model takes as constant the individual's behavioural responses to the policy change being examined, i.e. the individual's behavioural responses to the policies are not included in the model (Bourguignon and Spadro 2006). Hence the behavioural responses are considered exogenous, i.e. determined outside the model. Arithmetical models have been used to examine indirect taxes and tax reforms (Creedy 1999; Sahn 2003), to estimate incidence of public spending in health and education (Demery 2003), and also to compare fiscal policy effects (Callan and Sutherland 1997; Atkinson et al. 1988, 2002). For example, Atkinson et al. (1988) analysed the effect of replacing the French tax-benefit system with that of the British, for a given sample of French households.

In contrast, *behavioural MSMs* explicitly consider the changes in the behaviour of individuals in response to policy changes. These models are based on economic theory and may be policy-specific (Creedy and Duncan 2002). Behavioural MSMs have been used to evaluate the effects of direct tax reforms (Blundell et al. 2000; Das and van Soest 2001; Bonin et al. 2002) as well as indirect tax reforms (Creedy 1999; Liberati 2001; Kaplanoglou and Newbery 2003). The main advantages of behavioural MSMs are the ability to account for the heterogeneity within the population of interest, and the identification of both the mean and the distributional impact of a reform. However, these models require the estimation of a policy-specific behavioural model and they are often not generalizable for the evaluation of other policies (Zucchelli et al. 2010).

Dynamic MSMs can be represented in discrete or continuous time. In case of *discrete-time dynamic MSMs*, each individual's characteristics are simulated at fixed time intervals. These models usually include a probability model or a transition matrix for the simulations[73] (Willekens 2006). In New Zealand, Milne et al. (2015) developed a dynamic discrete-time MSM that modelled child development from birth to age 13, focusing on factors that influence health service use, early literacy and conduct problems of children. They used 2006 New Zealand Census data and three New Zealand child cohort studies[74] to build their model and transition probability estimates.

Continuous-time dynamic MSMs treat time as continuous and thus are able to estimate an exact time at which each event occurs. In these models, individuals are assigned characteristics that can change at any time. These models use survival functions to model the length of time that an individual will remain in his/her current state, and to simulate the timing of events (Willekens 2006). Although these models have theoretical advantages, they have higher data requirements than discrete time MSMs (Zaidi and Rake 2001). In Canada, Rowe and Wolfson (2000) used a dynamic continuous-time MSM called 'LifePaths' to model health care treatment, student loans and public pensions. Their analysis started with the cohort born in 1892 and extended for two centuries. In Australia, DYNAMOD is a dynamic continuous-time MSM developed by the National Centre for Social and Economic Modelling (NATSEM), and was designed to project population characteristics and the implications of policy changes over a 50-year period (King et al. 1999).

A dynamic MSM can be classified as either open or closed, based on whether new individuals are introduced to the base population as the simulation progresses, or not. In an *open MSM model* such as LifePaths in Canada, new individuals are generated if an individual in the initial population is selected to form a marital union. This differs from a *closed MSM model*, such as DYNACAN in Canada, which generates a new unit only when a baby is born (Zaidi and Rake 2001).

Dynamic MSMs can also be non-spatial or spatial in nature. *Spatial dynamic MSMs* are used to project the *geographical* trends in socio-economic activities, by combining spatial information into a dynamic MSM. For example, the SVERIGE model (Vencatasawmy et al. 1999; Rephann 2004) was the first national-level dynamic spatial MSM, and was developed

---

[73] Demographic modules in discrete-time dynamic MSMs are usually constructed using annual transition probability matrices.

[74] The Christchurch Health and Development Study, the Dunedin Multidisciplinary Health and Development Study, and the Pacific Islands Families Study.

from longitudinal socio-economic information on every resident in Sweden from 1985-1995. The model was used to study the spatial consequences of public policies at different geographical levels (national, regional and local). The model included specific events in a person's life, generated through deterministic models of behaviours that are functions of individual, household and regional socio-economic characteristics. Holm et al. (2002) studied population composition change in Sweden, by simulating development of all individuals in Sweden with respect to variations in demographic processes such as mortality, fertility and immigration using a spatial dynamic MSM. Their model was executed for 110 years (1990-2100).

Finally, MSMs differ in terms of the base population that they use. Some MSMs use Census or other survey data to form a base population. However, Census data sometimes do not provide all the variables necessary for analysis, so data may also be used from multiple alternative sources, generated for diverse purposes that are not compatible by design. In these cases, a *synthetic population* closely representing the actual population is formed as the initial base population in the MSM (Zaidi and Rake 2001). The synthetic unit records may be generated using existing datasets and a variety of techniques like iterative proportional fitting, linear programming, or complex combinatorial optimisation methods (Williamson et al. 1998; Ballas and Clarke 2000; Ballas 2001). For example, DYNACAN in Canada, DYNAMOD 2 in Australia, and PENSIM in the UK all use census or survey unit records as the base population, whereas NEDYMAS in Netherlands and LifePaths in Canada uses a synthetic database of unit records created using the census and other data sources (Li, J. and O' Donoghue 2012).

5.2.2 Previous MSMs Projecting Ethnic Population Change

Dynamic MSMs have been used to project the future ethnic composition of the population for several countries. For example, Demosim is a spatial dynamic MSM developed and maintained by Statistics Canada, which has been used to project the Canadian ethnocultural population composition. Demosim produces dynamic population projections at various spatial levels including provinces, territories, census metropolitan areas, and smaller geographical areas, based on individual demographic characteristics including age, sex, and place of birth (Statistics Canada 2018). Malenfant et.al. (2015) used the Demosim model to provide an insight into the ethnocultural makeup of the Canadian population in 2031 at different spatial scales. Taking 20 percent of the 2006 Canadian census as the base population, they calculated transition probabilities for mortality, immigration, internal migration, emigration, and highest level of schooling. They found that there would be a significant increase in ethnocultural diversity over time, within both the Canadian-born and the foreign-born population, especially in certain metropolitan areas such as Toronto and Vancouver.

Davis and Lay-Yee (2019) built a dynamic MSM (SociaLab) to simulate societal change in New Zealand from 1981 to 2038, to address social and policy questions related mainly to education, employment, personal/household income, household deprivation, and housing tenure. They worked with linked microdata from the New Zealand Longitudinal Census from 1981-2006 to build, calibrate, and validate their model. They considered individual demographic characteristics like age, sex, place of birth, religion, and ethnicity as predictor variables to estimate the pattern of changes in states and attributes throughout the life course for the New Zealand population. They used four broad ethnic groups (Māori, Pacific, Asian and NZ European/Other), considering them as time-invariant (i.e. the ethnic group/s of each individual remain constant throughout the simulation). The results from their model show that from 2006 to 2038, New Zealand will be ageing and becoming more ethnically diverse, which continues the observed trend over the past several decades (see also Mondal et al. 2020, who show similar past trends for Auckland, New Zealand's largest city). Also, the currently observed changing patterns in living arrangements, shifting away from the nuclear family, was projected to continue.

In the study most closely related to ours, Ardestani (2013) built a hybrid geosimulation model (a combination of an agent-based model and a microsimulation model) to investigate residential segregation in Auckland, New Zealand over the period 1991 to 2006. The author used New Zealand Census data to inform, calibrate and validate the model, and examined and measured the changes in ethnic residential segregation for four major ethnic groups.[75] Ardestani (2013) took into account the link between micro level (individual preference) and macro-level (number of groups, group size and proportion) elements to model and predict (until 2021) the changing ethnic residential patterns of the Greater Auckland Urban Area at both meso (territorial authorities[76]) and macro level (the entire Auckland urban area). He simulated several scenarios based on different assumptions about population growth, mobility rates of each ethnic group, housing vacancy rates, and freedom of movement (as a proxy for income). The study found that the ethnic population was projected to be consistently clustered over time

[75] European, Asian, Pacific, and Māori.
[76] Auckland City, Manukau, North Shore, Waitakere, and Papakura.

in all the area units in the Auckland urban area. Results also showed that the number of area units with a majority of Asians and Māori population will increase in the future in all of the territorial authorities they studied. In the Waitakere area, the author projected that there would be several area units where the Pacific people would be the largest group. It was also projected that in the Manukau area there would be an absolute decline in the European population.

In a follow-up study, Ardestani et al. (2018) used a multi-scaled agent based model to simulate the relocation of residents in five central territorial authorities (TA) of the Auckland urban area, to study the dynamics of residential segregation. They again focused on four major ethnic groups (Europeans, Asians, Pacific people and Māori). They found that a high fertility and migration scenario leads to lesser levels of residential segregation than a low fertility and migration scenario. They also found that, in the low fertility and migration scenario, residential segregation observed in the whole Auckland urban area was lesser than the residential segregation observed separately in some of the TAs (e.g. Manukau). They also looked into the impact of housing vacancy rates on the dynamics of residential segregation, and found that a reduction in housing vacancy rates leads to higher degrees of residential sorting at both the territorial authority and metropolitan area scales.

As noted earlier, studies relating to the spatial ethnic distribution of future population at the local level have been rare worldwide generally, and in New Zealand specifically. Among the relevant studies in New Zealand, Ardestani (2013) and Ardestani et al. (2018) did not investigate the residential segregation patterns at the micro (area unit) level, which offers a platform for more insightful findings (e.g. Mondal et al. 2020; Mondal et al. forthcoming). Moreover, they focused only on broad ethnic groups (Europeans, Asians, Pacific people and Māori), which ignores the diversity within these ethnic groups (especially for the Asian ethnic group) (Mondal et al. 2020). They also do not consider inter-ethnic mobility (changes in ethnic affiliation over time), which plays an important role in social change and is an increasingly popular and important area of research both internationally and in New Zealand (Carter et al. 2009).

In contrast, our paper develops and validates a spatial dynamic microsimulation model that can be used to investigate the future ethnic residential sorting in Auckland at a fine geographical (area unit) scale, and using more disaggregated ethnic groups than these previous studies. This is necessary to capture the underlying ethnic and spatial heterogeneity in choices and characteristics (Mondal et al. 2020). Ethnic mobility is experienced by a surprisingly large proportion of people in New Zealand.[77] Changes in ethnic identification are linked to historical socio-political experiences throughout an individual's life (Didham 2016). Thus, we also explicitly incorporate ethnic mobility into our model.

5. 3 Data

Auckland is the most ethnically diverse region in New Zealand and accounts for more than one-third (33.4 percent) of the New Zealand population. The major ethnic groups present in Auckland in 2018 were European (53.5 percent), Asian (28.2 percent), Pacific Peoples (15.5 percent), Māori (11.5 percent), MELAA[78] (2.3 percent), and Other (1.1 percent) (Statistics New Zealand 2020).[79] Because of its high ethnic diversity, we chose Auckland as our area of focus for this research.

We used the 1996-2001 and 2001-2006 New Zealand Linked Census (NZLC) data, for the Auckland region, which links successive censuses into longitudinal census datasets. Throughout this paper we use 'previous' to refer to data from the first census in each inter-censal period and 'current' for the same in the following census. The link rate for individuals from the 1991 to 1996 Census was 72 percent, 1996 Census to 2001 Census was 69.5 percent and from the 2001 to 2006 Census was 70.3 percent (Statistics New Zealand 2014).[80] The NZLC is the most comprehensive source of socio-demographic information on individuals (e.g. sex, age, ethnicity, education, place of residence etc.). The Census is usually conducted every five years, and collects information from all individuals present in New Zealand on census night (Statistics New Zealand 2018). These individual-level data can be aggregated to

---

[77] 22 percent of people changed their ethnicity in New Zealand in the 2006 Census (see Table 5.4).

[78] Middle Eastern/Latin American/African.

[79] The most recent population census was held on March 6, 2018. Linked longitudinal 2018 census data required for this analysis were not available at the time of writing of this report. Percentages do not sum to 100 percent, as people can report more than one ethnicity.

[80] The link rate for 2013 are unavailable. A census pair '*t*, *t*+5' refers to a pair of censuses where individual records in census (*t*+5) are linked to those of the previous census (*t*). For example, if we are looking at linking records from the 1996 Census to those from the 1991 Census, we refer to this as the 1991–1996 census pair. Though the terms "matching" and "linking" are used interchangeably, a 'link' refers to a record pair where the connection has been assessed as probable. A 'match' refers to a record pair where the connection is true. The matching process is comprised of two parts: deterministic matching based on a set of key variables to find unique matches, followed by probabilistic matching on the residuals. Deterministic matching uses a set of matching variables (sex, birth day, month and year, and area unit of residence) and matched records have the same unique values of the matching variables. In contrast, probabilistic matching evaluates all possible matches and uses statistical techniques to achieve matches (Statistics New Zealand 2014).

population statistics at different spatial scales such as the area unit[81] level. Our analysis is based on data aggregated to the area unit level, using 2013 Auckland area unit boundaries.[82] The Auckland region was comprised of 413 land-based area units in 2013, of which 409 had a non-zero usually resident population. We dropped area units with no usually resident population. The 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.[83]

In New Zealand, ethnicity is defined as the ethnic group that people feel a sense of belonging to. Ethnicity in New Zealand is not a measure of race, ancestry, nationality or citizenship. It is a measure of cultural affiliation. Ethnicity is self-recognised, and individuals can choose up to six ethnic groups in the census.[84] Individuals are also able to choose a different ethnicity in each census from any they have chosen previously (Statistics New Zealand 2015).

The New Zealand Standard Classification of Ethnicity orders ethnicity into rankings of four levels (Statistics New Zealand, 2013). The Level 1 classification of ethnicity has six categories and Level 2 has 21, which are shown in Table 5.1. The Level 1 ethnic groups are too broad and potentially mask heterogeneity in the characteristics of the ethnic groups, particularly for the Asian and the Pacific broad ethnic groups (Mondal et. al 2020). Thus, we use Level 2 ethnic groups in the analysis to better capture the diversities within the broad ethnic groups. Due to the small number of individuals reporting as 'others' and 'not further defined' groups (among those who are European, Asian or Pacific Islanders), we combined these four groups. Thus, we have eighteen ethnic groups, rather than twenty-one, in the analysis. We do not use finer Level 3 ethnic groups as the cell size is too small for some groups to develop a suitable model.

[81] Meshblocks are the smallest geographic units for which Statistics New Zealand collects and processes data. Area units are non-administrative aggregations of adjacent meshblocks with common boundaries (Statistics New Zealand 2013). An area unit is approximately the size of a suburb in urban areas.

[82] We use 2013 Area Unit boundaries as our proto-type microsimulation model will be used to project the ethnic diversity in Auckland from 2013 into the future.

[83] *Disclaimer*: The results in this paper are not official statistics. They have been created for research purposes from Census unit record data in the Statistics New Zealand Datalab. The opinions, findings, recommendations, and conclusions expressed in this paper are those of the authors, not Statistics NZ. Access to the anonymised data used in this study was provided by Statistics NZ under the security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation, and the results in this paper have been confidentialized to protect these groups from identification and to keep their data safe. Careful consideration has been given to the privacy, security, and confidentiality issues associated with using unit record census data.

[84] Individuals could choose up to three ethnic groups until 1996, which was increased to six in later censuses.

Table 5.1: Ethnic group classification in New Zealand

Ethnic group code (Level 1)	Ethnic Group code description (Level 1)	Ethnic group code (Level 2)	Ethnic Group code description (Level 2)
01	European	10	European not further defined
		11	NZ European
		12	Other European
02	Māori	21	NZ Māori
03	Pacific Peoples	30	Pacific Island not further defined
		31	Samoan
		32	Cook Island Māori
		33	Tongan
		34	Niuean
		35	Tokelauan
		36	Fijian
		37	Other Pacific Island
04	Asian	40	Asian not further defined
		41	Southeast Asian
		42	Chinese
		43	Indian
		44	Other Asian
05	MELAA	51	Middle eastern
		52	Latin American/Hispanic
		53	African
06	Other	61	Other ethnicity

Source: Statistics New Zealand (2013)

Two issues affect the use of ethnicity data. First, the format and wordings of the Census ethnicity question has been inconsistent between censuses. For instance, the ethnicity question in 2001 differed substantially from that in 1996.[85] These inconsistencies particularly affect the 'European' ethnic groups (including 'New Zealand European') and the 'Māori' ethnic group. In the 1996 data, the count for 'Other Europeans' was much higher than in the 2001 data. This was because the difference in format of the ethnicity question resulted in increased multiple

[85] In the 1996 Census, the ethnicity question had a different format compared to that used in 1991 and 2001. In 1996, there was an option to choose 'Other European' with additional drop down answer boxes for 'English', 'Dutch', 'Australian', 'Scottish', 'Irish' and 'Other'. These options were absent in 1991 and 2001 Censuses. Moreover, the first two answer boxes were in a different order in 1996 from that in 1991 and 2001. In 1996, 'NZ Māori' was listed first and 'NZ European or Pākehā' was listed second. 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' (Statistics New Zealand, 2017).

responses, and a consequent reduction in single responses. This also resulted in some respondents answering the 1996 question on the basis of ancestry rather than ethnicity. The count for the 'New Zealand European' category was much lower in 1996 than in 2001, which 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).

Second, there has also been inconsistency in the treatment of responses of 'New Zealander' to the Census ethnicity question.[86] 'New Zealander' was included explicitly as a new category in 2006, but not in 2001 or earlier. In 2001, individuals considering themselves to be a 'New Zealander' were likely to have been counted in the 'New Zealand European' ethnic category (Statistics New Zealand 2017).

The individual level independent variables included in our analysis for the ethnic transition module are the individual's ethnicity in the previous census, their age, sex and whether they were born in New Zealand. The neighbourhood level variables are the ethnic diversity and the percentage share of the different ethnic groups in the area unit they reside in. All independent variables were observed at the start of each inter-censal period.

As mentioned earlier, the Census collects self-reported ethnic identification and each individual can affiliate themselves with single/multiple ethnic groups. In our models, we consider every Level 2 ethnicity that the person reports as their ethnic group. Thus, in the analysis the individual's ethnicity is an 18x1 vector of binary variables, with one variable for each of the eighteen ethnic groups (belongs to ethnic group 'i'=1, otherwise=0).

The Normalised Entropy Score measure[87] of residential sorting based on individual's reported ethnicity, is a measure of ethnic diversity in each area unit used in the analysis (Mondal et. al.

---

[86] In 2005, a standard for ethnicity statistics was developed. Previously, the 'New Zealander' response was included in the 'European' category, which was later moved to the 'Other ethnicity' category (Statistics New Zealand 2007a). New Zealand Europeans were most prone to calling themselves 'New Zealander' in the census (Statistics New Zealand 2007b; Brown and Gray 2009), resulting in an increase in the 'Other ethnicity' category, and a consequent reduction in the size and proportion of people reporting as European or New Zealand European.

[87]Normalised Entropy diversity measure is calculated using the formula $D = -\sum_{g=1}^{G} \frac{P_{ga}}{P_{..}} \ln\left(\frac{P_{ga}}{P_{..}}\right) / \ln(G)$, (where P_{ga} refer to the population of group g (=1, 2,…G) in area a (= 1,2,….A), $P_{..}$ is the total Auckland population).To allow calculation of D even in the case of there being groups who have zero members at some point in time, we define 0*ln(1/0)= $\lim_{q\to 0}[q(\ln(1/q)] = 0$.

2020). The normalised diversity index ranges from 0 (when only one ethnic group is present in the area unit) to 1 (when all ethnic groups are equally represented in area unit) (Nijkamp and Poot 2015). We also use the proportion of the population that identifies with each ethnic group, calculated at the area unit level.

5.4 Methodology

In this section, we describe the construction and calibration of a microsimulation model which can be used to project the future spatial distribution of ethnic diversity in Auckland, taking ethnic and spatial mobility into consideration. Models for future residential mobility and sorting need to capture realistic trends and their applicability to real urban areas (Ardestani et al. 2018). Thus, it is recommended to use real-world data, based on the same administrative spatial boundaries that the data are collected and in modelling how individuals interact with each other in the real world (Batty 2010). Our model is a *discrete-time (*runs in five year time steps) *probabilistic (*uses transitional probabilities to project forward) and *dynamic* (includes time effect) *MSM*. Our model is an *open dynamic MSM* as, in addition to people moving between area units within Auckland, it allows individuals to move out of Auckland (out-migration) as well as move into Auckland from other parts of the world (in-migration). As we study the geographical trends in the ethnic diversity of Auckland, our model is *spatial* in nature.

The MSM model we describe here is a *validation model*, which uses data from the 1996-2001 linked Census to simulate and project variables in 2006, which is then validated against actual 2006 census data. This model can then be used to develop a *projection model*, which will use data from the 2006-2013 linked Census to simulate and project predictor and predicted variables from 2006 and 2013 respectively, to predict variables in 2018 and 2023 and so on. The projection model is beyond the scope of the present chapter and will be developed in follow-up work. The validation model is comprised of two modules: (1) an ethnic transition module; and (2) a location transition module. For each of these modules, we break the population into two age groups: (1) children/adolescents (0-17 years); and (2) adults (above 17 years).

Table 5.2 shows the details of the variables used in the analysis. As the decision to move is effected by duration of stay (Poot 1987), we include number of years the resident has lived in the origin area unit as an explanatory variable in the location transitional module along with all other variables included in the ethnic transition module.

Table 5.2: Variables used in the analysis

Module	Predicted Variable	Level of variables	Predictor variables (all evaluated at the time of the previous census)
Ethnic Transition	Ethnic affiliation in current census (1=belongs to ethnic group *'i'*, 0=otherwise)	Individual	Ethnicity, Age, Sex, NZ-born
		Neighbourhood	Ethnic diversity in area units, Ethnic group size proportions in area unit.
Location Transition	Moved[88] (1=moved, 0=otherwise)	Individual	Ethnicity, Age, Sex, NZ-born, years at address
		Neighbourhood	Ethnic diversity in area units, Ethnic group size proportions in area unit.

[88] We created the binary variable 'moved' (1=if individual changed area unit during the intercensal period, 0=otherwise) from the census data on usual-resident location in the current census and the variable 'address five years ago' for the same individual.

For out-migration, we consider individuals who were present in Auckland in the previous census, but absent in Auckland in the current census. Thus, our out-migration includes people who moved from Auckland to elsewhere in New Zealand. For emigration (people who moved from Auckland to overseas) and for people who died between the previous census and the current census we did not have individual level data. This is because the linked census data provides information in the current census (2001 in this case) for only the individuals who have been linked in both 1996 and 2001 census. Thus, to account for emigration and death, we apportioned the number of emigrants from Auckland and number of deaths in Auckland to each area unit according to area unit population[89]. For in-migration, we identify individuals who were not present in the previous census in Auckland but present in Auckland in the current census. We use the census characteristics of these individuals in the model to proxy for inward-migration and births. Thus, as they are in practical terms conflated with out-migration and in-migration respectively, we account for both mortality and fertility in this model.

In our model, we capture individual ethnic transitions as well as their spatial mobility i.e. individuals making choices regarding their ethnicity and location. Figures 5.1 and 5.2 outline the theoretical framework for the ethnic transition and location transition modules respectively. In our model, the ethnic transition module runs first, before the location transition module.

[89] Emigration was calculated as a residual of 1996-2001 population change.

Figure 5.1: Theoretical framework – Ethnic Transition

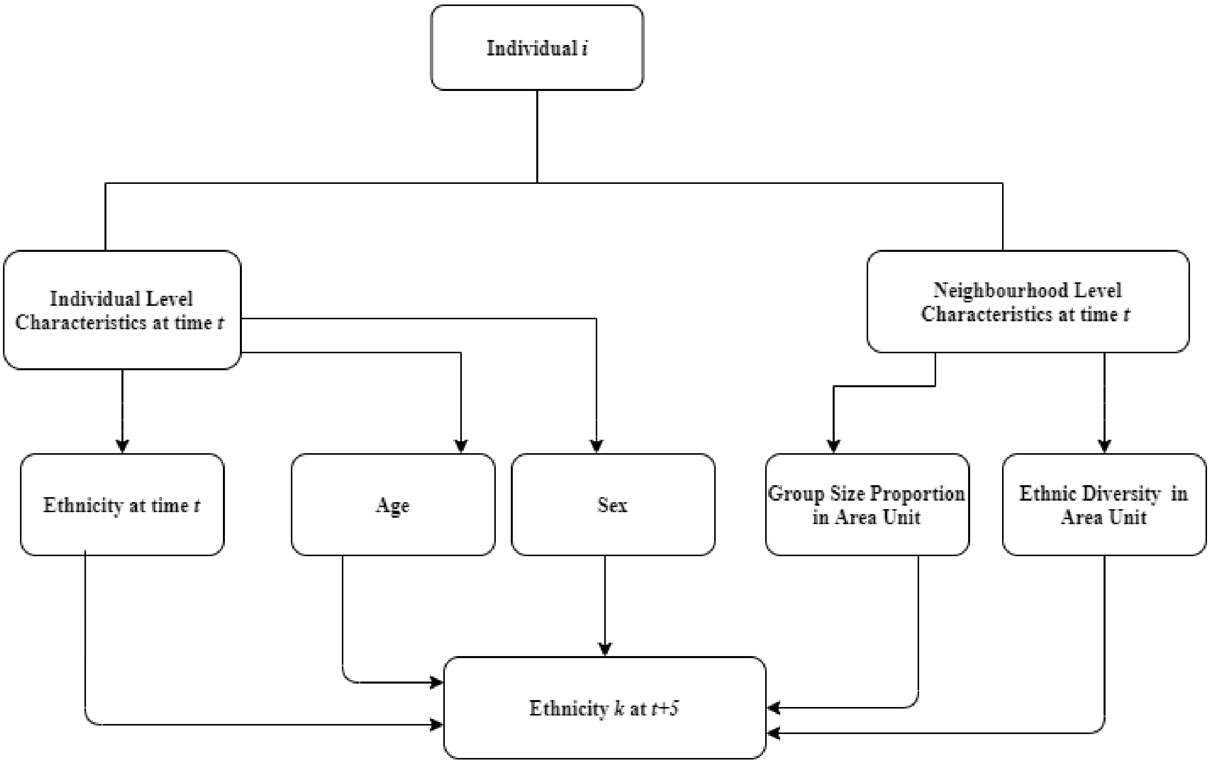
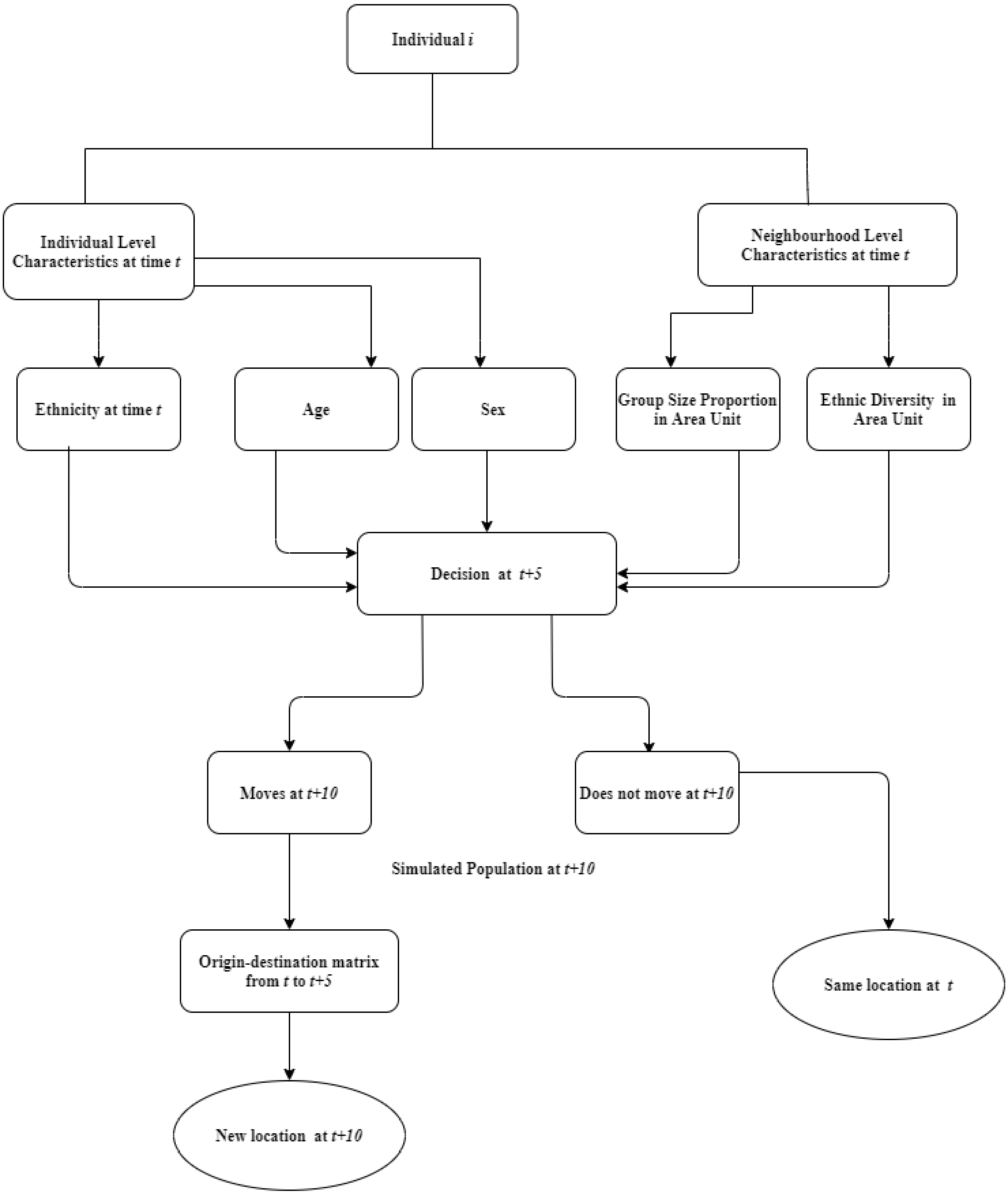


Figure 5.2: Theoretical framework – Location Transition



The ethnic transition module runs a separate logistic regression equation for each ethnicity. We take the individual's ethnic response, which is binary (1= belongs to ethnic group 'i', 0= otherwise), in the current census as the dependent variable. This variable represents whether or not the individual identifies with that group (1=identifies with the group, 0=otherwise), regardless of whether they also identify with one or more other groups. This substantially simplifies the analysis relative to a multinomial logit specification, which would require that every possible combination of ethnic affiliations be an option. This also allows us to include possible multiple ethnic affiliation for individuals, without requiring an order of priority for the determination of the ethnic choices, i.e. each individual's each choice in regards to each ethnicity is given equal importance. From the logistic regression models, we obtain the predicted probabilities of an individual belonging to ethnic group 'i' in the current census. We then assign uniformly distributed random variables (over the interval 0 and 1) to each individual. Comparing the predicted probabilities with the random variables the model determines whether the individual identifies with that ethnicity in the projected year.

The location transition module proceeds in two stages, following Willekens' (2016) migrant pool model for projecting migration. In the first stage, the number of out-migrants (i.e. people who change their usual residence) is projected. In the second stage, the people who changed their location are then distributed over possible destinations using a distribution function that is solely dependent on the destination but not on the origin. Specifically, we first use logistic regression to obtain predicted probabilities of moving for each individual in the current census. Similar to our ethnic transition model, we assign a uniformly distributed random variable to each individual. Then, comparing the values of the random variable and the predicted probabilities, the model determines whether the person is a mover or not in the projected year. In the second step, movers are allocated their destination area units based on a column-standardised origin-destination matrix (with a zero diagonal), with a different origin-destination matrix for each ethnicity. For individuals with multiple ethnicities, one of their ethnicities is chosen at random. The destination for each migrant is determined again using a uniformly distributed random variable, with the origin-destination matrix used as a lookup table to determine their destination. For those individuals where 'outside Auckland' (out-migration or death) is selected as the destination, they are removed from the dataset.

5.4.1 Projection Evaluation

As mentioned earlier, in our validation model we use predictor variables from 1996 and predicted variables from 2001 and use the validation model to project and simulate 2006 data.

We evaluate the performance of our model in two ways. First, we compare the proportion of people who changed their ethnicity, the proportion of people who changed their location, and the proportion of people who moved out of Auckland in 2006 in our simulated data to those in the actual 2001-2006 linked census data.

Second, we calculate measures of residential sorting based on the simulated data for 2006 as well as the actual 2006 census data and use different forecast error measures to estimate forecast error and bias in the model.

Measures of residential sorting

There are many different measures that can be used as indicators of residential sorting (see e.g. Massey and Denton 1988; Nijkamp and Poot 2015; Reardon and Firebaugh 2002). We choose entropy-based measures of residential sorting and diversity, following the influential contribution by Theil and Finezza (1971). Entropy measures are conceptually and mathematically attractive and are least biased by group size (Reardon and Firebaugh 2002; Modal et.al. forthcoming). The formulas of the measures of residential sorting and diversity used in the analysis are detailed in Table 5.3. In order to observe the extent to which one ethnic group is over or under represented in an area unit, we calculate diversity (entropy) index (E_a) of the population in area unit a in terms of the given ethnic group classifications. We normalise the entropy diversity index to an evenness index I_a that varies between zero and one, following Nijkamp and Poot (2015). The value of the diversity evenness index is zero (i.e. $E_a = 0$) when only one of the groups is present in area unit a and is one (i.e. $E_a = 1$) when all groups are equally represented in area unit a. We calculate the Entropy Index of spatial sorting of group g (EIS_g), which measure the area-population weighted average of one minus the relative entropy of the areas $\left(\frac{E_{ga}}{\bar{E}_g}\right)$ with respect to group g. This index varies between zero (when the group is distributed proportionally to the total population in all area units) to one (when all areas in which group g is represented contain no other group). We also calculate an overall measure of residential sorting *(H*)*, for Auckland, by taking the group-population weighted average of the EIS_g values. This is an alternate way of calculating the *Theil's Multi-group Segregation Index H* (Theil 1972; Theil and Finezza 1971; White 1986). This calculation gives approximately the same value as H, but is easier to interpret. We also calculate the normalised diversity (entropy) index I* of the whole Auckland population in terms of the given ethnic group classifications.

Table 5.3: Summary Measures of Residential Sorting

Entropy diversity (area unit)	$E_a = -\sum_{g=1}^{G} \frac{P_{ga}}{P_a} \ln \frac{P_{ga}}{P_a}$
Normalised Entropy diversity (area unit)	$I_a = -\frac{\sum_{g=1}^{G} \frac{P_{ga}}{P_a} \ln \frac{P_{ga}}{P_a}}{\ln(G)}$
Entropy Index of Segregation	$EIS_g = \sum_{a=1}^{A} \frac{P_a}{P}\left(1 - \frac{E_{ga}}{\bar{E}_g}\right)$
	Where: $E_{ga} = -\frac{P_{ga}}{P_a} \ln\left(\frac{P_{ga}}{P_a}\right) - (1 - \frac{P_{ga}}{P_a})\, ln\left(1 - \frac{P_{ga}}{P_a}\right)$
	$\bar{E}_g = -\frac{P_g}{P}\, ln\left(\frac{P_g}{P}\right) - (1 - \frac{P_g}{P})\, ln\left(1 - \frac{P_g}{P}\right)$
Normalised Entropy diversity (city)	$I^* = -\frac{\sum_{g=1}^{G} \frac{P_{g.}}{P}\, ln \frac{P_{g.}}{P}}{ln\,(G)}$
Theil's multi-group spatial sorting index	$H^* = \sum_{g=1}^{G} \frac{P_g}{P} EIS_g$

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 $\sum_{a=1}^{A} \pi_{ga} = 1$. P_a is the total number of people in area unit a. P_g as the number of members of group g in Auckland and P to be the total number of people in Auckland. Comparing group g with all other groups combined, we denote the entropy of area a as (E_{ga}) and whole Auckland city as $\bar{E}_g$. Comparing 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).

Forecast error measures

Following Cameron and Cochrane (2017) and Wilson (2015), we estimate multiple measures of forecast error and bias. Projection error is defined as the difference between the index values based on the simulated population (M_t) and the actual population (A_t), standardised by the actual population size. Thus, the projection Percentage Error at time *t*+5 (PE_{t+5}) is given as:

$$PE_{tt+5} = \frac{M_{t+5} - A_{t+5}}{A_{t+5}} x100\%$$

To report projection accuracy, we use the weighted mean absolute percentage error (WMAPE) as our primary measure. This is a weighted mean of the absolute Percentage Errors (PE_t), with weights equal to the actual group size proportions of the population in the year projected (Siegel 2002; Wilson 2012). WMAPE is preferable in cases where the population sizes vary widely. In our study, population size of an area unit in Auckland varies from less than 9 to over 3000. WMAPE at projected year t+5 is defined as:

$$WMAPE_t = \sum_g \left(\left| PE^g_{tt+5} \right| \frac{P_{g.t+5}}{P_{t+5}} \right)$$

where g is the number of groups, $P_{g.t+5}$ is the population size of each group and P_{t+5} is size of the whole population in the year of projection $t+5$.

The population projection error distribution is likely to be right-skewed due to the small numbers of unusually high errors resulting in the mean being a poor representation of the average error (Tayman and Swanson 1999). Thus, we also report the median absolute percentage error (MedAPE_t) and the median algebraic percentage error ($\mathrm{MedALPE}_t$). MedAPE_t is the middle of a set of ranked absolute PE_t values. MedAPE_t is a measure of precision of projection because it is not influenced by the direction of the error. On the other hand, $\mathrm{MedALPE}_t$ measures the middle of a set of ranked non-absolute (i.e. algebraic) PE_t, values. This measure preserves the negative and the positive percentage error values. Hence, it is a measurement of projection bias. Both$\mathrm{MedAPE}_t$and $\mathrm{MedALPE}_t$values are not affected by extreme outliers.

5.4.2 Calibration Process

After performing the initial stages of model coding, we calibrated the initial model so that the simulated results would be as close as possible to the actual data. We assume that if the proportion of people changing their location, proportion of people in each ethnic group and proportion of each ethnic group changing their location in the simulated data is close to the actual data, then the model should project close to the correct levels of ethnic diversity and residential sorting. The calibration processes undertaken are described below.

Step 1: Calibrating the proportion of 'movers'

We observed that the percentage of people changing locations in our initial model was less than that observed in the actual data. We took the difference between the actual and the simulated proportion of people changing their location as our first calibration constant. We then subtracted this calibration constant from the previously generated uniformly distributed random variable, thereby ensuring that the model would increase the number of 'movers'. The model then uses this adjusted random variable to the predicted probabilities to determine whether the person is a mover or not.

Step 2: Calibrating the proportion of people in each ethnic group

We calculated the difference between the proportion of people in each ethnic group between the simulated data and the actual data. We considered the difference for each ethnic group as a calibration constant for that ethnic group. For the cases where the model simulated too much of an ethnic group, we added the respective calibration constant onto the uniformly distributed random variable. We subtracted the calibration constants to the random variable if the model simulated too few of an ethnic group. This process was repeated several times, aiming to minimise the sum of the absolute differences between actual and simulated proportions.

Step 3: Calibrating the proportion of people in each ethnic group who are 'movers'

We calculated the differences between the proportion of people changing location in the simulated data and the actual data for each ethnic group. We treated these differences for each ethnic group as ethnic-specific calibration constants. We then subtracted the calibration constant for ethnicity i from the predicted probability of 'moving' for people who belong to ethnicity i. For people belonging to multiple ethnic groups we subtracted all of the ethnic-specific calibration constants that apply to them from the predicted probability of 'moving'. Again, this process was repeated several times, aiming to minimise the sum of the absolute differences between actual and simulated proportions.

5.5 Results

Our ultimate aim is to use our microsimulation model to build a *projection model* that will project the population forward with minimum error. To this end, we validated the ability of the current model to replicate known 2006 census outcomes. Table 5.4 shows that the counts of people in Auckland in the actual (655,767) and simulated (678,807) 2006 Census data are similar. Table 5.4 also shows that 22 percent of people, who were in Auckland in 2001, changed their ethnicity in the actual 2006 Census, whereas the same for the simulated 2006 Census is 25 percent. Considering the inconsistencies in the ethnic categorisations in the 1996-2001 Census data already mentioned,[90] which was used to parameterise the initial model, these percentages are very close. The percentage of people moving from one area unit in 2001 to a different area unit in 2006 in the actual 2006 Census (40 percent) and the simulated data (40 percent) are very similar. The difference in the percentage of people moving out of Auckland between the actual (9 percent) and the simulated (8 percent) data is only one percentage point.

Table 5.4: Comparison between simulated data and the actual 2006 Census data

Variable	Actual	Model	Difference (Model- Actual)
Ethnic change	22%	25%	3%
Location change	40%	40%	0%
Movement out of Auckland	9%	8%	-1%
People in Auckland	655,767	678,807	23,040

Table 5.5 shows that in terms of overall ethnic residential sorting in Auckland, our simulated value for Theil's multi-group spatial sorting index *(H*)* is very close to the actual value, the difference being just 0.004. Table 5.5 also shows that the simulated ethnic diversity in Auckland *(I*)* very closely matches the actual ethnic diversity observed in Auckland in 2006.

[90] See section 5.3 for details.

Table 5.5: Actual and Simulated Spatial sorting in Auckland, 2006

Measures of Residential Sorting	Actual	Model	Difference (Model-Actual)
Theil's multi-group index (*H**)	0.087	0.083	-0.004
Evenness Index (*I**)	0.621	0.632	0.012

Table 5.6 summarises all three forecast error measures (WMAPE, MedAPE and MedALPE) for both the Entropy Index of Segregation measure for the ethnic groups and the Normalised Entropy Diversity measure for area units. The WMAPE estimates are larger than the MedAPE for both the simulated sorting of ethnic groups and the simulated ethnic diversity of the area units. This might indicate that the absolute errors are largest for area units and ethnic groups with larger populations (Cameron and Cochrane 2017). Overall, the model shows a moderate degree of accuracy in terms of projecting ethnic group sorting. The negative MedALPE (-23.5 percent) value reflects that there is downward bias in the simulated values of the Entropy Index of Segregation measure (Table 5.6, column (A)), potentially resulting from the partial observability of all characteristics that might affect the ethnic transitions. Table 5.6 (column (B)) demonstrates that the model performs well in terms of the simulated Normalised Entropy diversity measure for area units, with the WMAPE and the MedALPE value being just 4.35 percent and 2.13 percent respectively.

Table 5.6: Model Performance

Error Measure	***EIS*** **(A)**	***I*** **(B)**
WMAPE (%)	19.72	4.35
MedAPE (%)	23.48	3.69
MedALPE (%)	-23.48	2.13

5.6 Conclusion

The main aim of this paper was to describe the development and calibration of a microsimulation model that can be used for projecting the future spatial ethnic distribution in Auckland. The model described in this paper takes both ethnic and spatial mobility into consideration. Data from the 1996-2001 NZLC was used to simulate census data for 2006. The simulated results were then compared to the available actual 2006 Census data.

We have demonstrated that census data can be used to inform, calibrate and validate our model, which is capable of reproducing the dynamics of residential sorting in Auckland, without detailed information on all the elements of an individual's residential decision-making process. The results show that our model is capable of projecting the ethnic future spatial distribution in Auckland with minimum error.

Results from the location transition module are fairly close to the actual data. However, our ethnic transition module shows a moderately lower degree of accuracy. We interpret this as an error caused by the inconsistencies in the ethnic categorisation in the census data that was used as the base data in our model. We infer this from the fact that the way both the ethnic and location transition modules work is similar.

This model is not without its limitations. First, with a given set of predictor variables, logistic regression models are used to predict the probability of a certain level of event occurring. Hence, only the people who have been linked in the 1996-2001 NZLC could be used in the logistic regression model. However, the base population for the model is comprised of the whole Auckland population in the 2001 Census, whether linked/non-linked in the 1996-2001 NZLC data. Thus, to the extent that unlinked and linked people differ in ways that are relevant to the transitions we estimate, that will generate some bias in the results. However, some of this bias will be attenuated through the process of calibration.

Second, due to the too few people reporting as belonging to the 'Not further defined (NFD)' and 'Other' ethnic groups, we combined these into one broad ethnic group called 'Others'. As the 'NFD' groups are a disaggregated Level 2 category in the ethnic classification under each broad Level 1 ethnic category, they are likely to behave more like the other sub-groups within their Level 1 broad ethnic group than they would to the 'Other' Level 1 ethnic group with which they have been merged. This problem could be eliminated by removing these ethnic groups from the model. However, we preferred to retain these ethnic groups at this stage of model development. A future extension to this work could be to remove these ethnic groups and observe how the results change, or to merge them into other Level 1 groups. Another future prospect for this work could be to consider these as separate ethnic groups. This would be easier if the same model were extended to consider the future ethnic diversity of the whole of New Zealand, wherein the problem of small cell counts for these groups would be reduced.

Third, an individual's location decision and ethnic choices are dependent on a variety of factors other than the ones that are used in the model, one of these being their completed education level. Although data on the completed education for adults is available in the Census, the same data for children transitioning to adulthood is not available. Including education within the model would require the addition of a module on educational attainment. We initially attempted to parameterise such a model, but it performed poorly.[91] Thus, we have not included education as one of our predictor variables in the model. As a future prospect for research, it would be interesting to see how adding an additional 'educational transition' module to the model alters the results.

Fourth, ethnic identity of the parents is important for the adolescents (Mondal et al. 2020). However, the NZLC does not have this data. Thus, we could not include this variable in the model.

In spite of these limitations, this paper has contributed significantly to the limited evidence on projecting ethnic diversity at a local and sub-ethnic group level in Auckland, New Zealand, and internationally. Our model was developed using Stata, which extends the number of resources previously used to build the microsimulation models. Our future focus will be to concentrate on further calibration of the model, looking at the co-location of individuals with others in their own ethnic groups. Future research by the same authors will use the final calibrated model and the 2006-2013 NZLC data to project the future ethnic spatial distribution in Auckland forward to 2038.

[91] Further details are available from the authors on request.

References

Alonso, W. (1964). *Location and land use.* Cambridge, MA: Harvard University Press

Ardestani, B.M. (2013). *Using a hybrid model for investigating residential segregation: an empirical and simulation-based study (*(Doctoral thesis). Auckland, New Zealand: University of Auckland, New Zealand). Retrieved from https://researchspace.auckland.ac.nz/handle/2292/21618

Ardestani, B. M., O'Sullivan, D., & Davis, P. (2018). A multi-scaled agent-based model of residential segregation applied to a real metropolitan area. *Computers, Environment and Urban Systems, 69*, 1–16. https://doi.org/10.1016/j.compenvurbsys.2017.11.002

Atkinson A., Bourguignon, F., & Chiappori, P. A. (1988). What do we learn about tax reforms from international comparisons? France and Britain. *European Economic Review,* 32 (2-3), 343-52.

Atkinson, T., Bourguignon, F., O'Donoghue, C., Sutherland, H. & Utili, F. (2002). Microsimulation of social policy in the European Union: Case study of a European minimum pension. *Economica*, *69*, 229-243.

Ballas, D. and Clarke, G.P. (2000). GIS and microsimulation for local labour market policy analysis. *Computers, Environment and Urban Systems*, *24*,305-30.

Ballas, D. (2001). A spatial microsimulation approach to local labour market policy analysis (unpublished PhD thesis). School of Geography, University of Leeds.

Ballas, D., Clarke, G., & Wiemers, E. (2005a). Building a dynamic spatial microsimulation model for Ireland. *Population, Space and Place*, *11*(3), 157-172. doi:10.1002/psp.359.

Ballas, D., Rossiter, D., Thomas, B., Clarke, G.P., and Dorling, D. (2005b). *Geography matters: Simulating the local impacts of national social policies.* York, United Kingdom: York Publishing Services.

Batty, M. (2010). Towards a new science of cities. *Building Research and Information*, *38* (1), 123-126

Bedford, R., & Ho, E. (2008). *Asians in New Zealand: Implications of a changing demography*. (Outlook Series Paper 07). Wellington, New Zealand: Asia New Zealand Foundation. Retrieved from https://www.asianz.org.nz/assets/Uploads/fd30383787/Asians-in-New-Zealand-Implications-of-a-changing-demography.pdf

Birkin, M., Turner, A., Wu, B., Townend, P., Arshad, J., & Xu, J. (2009). MoSeS: A grid-enabled spatial decision support system. *Social Science Computer Review*, *27*(4), 493–508. https://doi.org/10.1177/0894439309332295

Blundell, R., Duncan, A., McCrae, J. & Meghir, C. (2000). The labour market impact of the working families' tax credit. *Fiscal Studies, 21*(1), 75–104.

Bonin, H., Kempe, W., & Schneider, H. (2002). *Household labour supply effects of low-wages subsidies in Germany* (IZA Discussion Paper 637). Bonn, Germany: Institute for the Study of Labor.

Bourguignon, F., & Spadaro, A. (2006). Microsimulation as a tool for evaluating redistribution policies. *Journal of Economic Inequality*, *4*, 77-106.

Brown, P. & Gray, A. (2009). Inter-ethnic mobility between the 2001 and 2006 Censuses: The statistical impact of the New Zealander response. In Statistics New Zealand, *Final report of a review of the official ethnicity standard 2009* (pp 27-36). Wellington, New Zealand: Statistics New Zealand.

Bruch, E. E., & Mare, R. D. (2006). Neighborhood choice and neighborhood change. *American Journal of Sociology*, *112*(3), 667–709. https://doi.org/10.1086/507856

Caldwell, S., Clarke, G., & Keister, A. (1998). Modelling regional changes in US household income and wealth: A research agenda. *Environment and Planning C: Government and Policy*, *16*(6), 707- 722. doi:10.1068/c160707.

Callan, T., & Sutherland, H. (1997). The impact of comparable policies in European countries: Microsimulation approaches. *European Economic Review*, *41* (3-5), 327–33.

Cameron, M.P., & Cochrane, W. (2016a). *2016 update of population, family and household, and labour force projections for the Waikato Region*, 2013-2063, research report commissioned by Future Proof. Hamilton, New Zealand: University of Waikato.

Cameron, M.P., & Cochrane, W. (2016b). *2016 update of area unit population, household, and labour force projections for the Waikato Region, 2013-2061*, research report commissioned by Future Proof. Hamilton, New Zealand: University of Waikato.

Cameron, M. P. & Cochrane, W. (2017). Using land-use modelling to statistically downscale population projections to small areas. *Australasian Journal of Regional Studies*, *23*(2), pp 195-216.

Cameron, M.P. & Poot, J. (2019). Towards superdiverse Aotearoa: Dimensions of past and future ethnic diversity in New Zealand and its regions. *New Zealand Population Review*, *45*, 18–45.

Carter, K. N., Hayward, M., Blakely, T., & Shaw, C. (2009). How much and for whom does self-identified ethnicity change over time in New Zealand? Results from a longitudinal study. *Social Policy Journal of New Zealand, 36*, 32–45.

Creedy, J. (1999). *Modelling Indirect taxes and tax reform*. Northampton, United Kingdom: Edward Elgar.

Creedy, J. & Duncan, A. (2002). Behavioural microsimulation with labour supply responses. *Journal of Economic Surveys*, *16*, 1-39.

Das, M. & van Soest, A. (2001). Family labour supply and proposed tax reforms in the Netherlands. *De Economist*, *149*, 191–218.

Davis, P., & Lay-Yee, R. (2019). *Simulating societal change: Counterfactual modelling for social and policy Inquiry*. Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-04786-3

Denton, N.A., Massey, D.S. (1988) Residential segregation of blacks, Hispanics and Asians by socioeconomic status and generation. *Social Science Quarterly, 69*, 797–817.

Demery, L. (2003). Analysing the incidence of public spending. In F. Bourguignon and P. da Silva L. (Eds.), *The impact of economic policies on poverty and income distribution: Evaluation techniques and tools* (pp 41-68). Washington, DC: World Bank Publications.

Didham, R. (2016). Ethnic mobility in the New Zealand census, 1981-2013: A preliminary look. *New Zealand Population Review*, *42*, 27–42.

Domina, T. (2006) Brain drain and brain gain: Rising educational segregation in the United States, 1940–2000. *City Community*, 5, 387–407. https://doi.org/10.1111/j.1540-6040.2006.00190.x

Duncan, O.D., & Duncan, B. (1955) Residential distribution and occupational stratification. *American Journal of Sociology*, *60*, 493–503.

Eggink, E., Woittiez, I., & Ras, M. (2016). Forecasting the use of elderly care: A static micro-simulation model. *The European Journal of Health Economics: HEPAC: Health Economics in Prevention and Care*, *17*(6), 681–691.

Farley, R. (1977) Residential segregation in urbanized areas of the United States in 1970: An analysis of social class and racial differences. *Demography, 14*, 497–518.

Favreault, M., & Smith, K.E. (2004). *A primer on the dynamic simulation of income model (DYNASIM3).* The Urban Institute. Retrieved from https://www.urban.org/sites/default/files/publication/71226/410961-A-Primer-on-the-Dynamic-Simulation-of-Income-Model-DYNASIM-.PDF

Finta, N. (1987). *Preliminary computation with real data in Hungaria*. Budapest, Hungary: Hungarian Ministry of Agriculture.

Fischer, M. J. (2003). The relative importance of income and race in determining residential outcomes in U.S. urban areas, 1970-2000. *Urban Affairs Review*, *38*, 669–696. https://doi.org/10.1177/1078087403038005003

Goldman, D.P., Zheng, Y., Girosi, F., Michaud, P.C., Olshansky, S.J., Cutler, D., & Rowe, J.W. (2009). The benefits of risk factor prevention in Americans aged 51 years and older. *American Journal of Public Health*, *99*(11), 2096-2101. doi:10.2105/AJPH.2009.172627

Hancock, R., Mallender, J., & Pudney, S. (1992). Constructing a computer model for simulating the future distribution of pensioners' incomes for Great Britain. In R. Hancock & H. Sutherland (Eds.) *Microsimulation models for public policy analysis:*

New frontiers (pp. 33-66). London, England: Suntory-Toyota International Centre for Economics and Related Disciplines.

Harding, A. (2007*). APPSIM: The Australian dynamic population and policy microsimulation model*. Canberra, Australia: National Centre for Social and Economic Modeling.

Ho, E., & Bedford, R.D. (2006) The Chinese in Auckland: Changing profiles in a more diverse society. In Li W (Ed.) *From urban enclave to ethnic suburb* (pp 203-233). Honolulu, Hawaii: University of Hawaii Press.

Holm, E., Holme, K., Lindgren, U., & Makila, K. (2002). *The SVERIGE spatial microsimulation model* (Discussion paper 595). Umea, Sweden: Department of Social and Economic Geography, Umea University.

Holmer, M., Janney, A., Cohen, B. (2014). *PENSIM overview*. Washington, DC: U.S. Department of Labor, Office of Policy Research. Retrieved from http://www.polsim.com/doc/overview.pdf

Iceland, J., Weinberg, D.H., & Steinmetz, E. (2002). *Racial and ethnic residential segregation in the United States 1980-2000. Appendix B.* Washington, DC: US Census Bureau.

Immervoll, H., Kleven, H. J., Kreiner, C. T., & Saez, E. (2007). Welfare reform in European countries: A microsimulation analysis. *The Economic Journal*, *117*(516), 1–44. https://doi.org/10.1111/j.1468-0297.2007.02000.x

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*, 1–23. https://doi.org/10.1111/j.1467-9663.2009.00577.x

Kaplanoglou, G. & Newbery, D.M. (2003). Indirect taxation in Greece: Evaluation and possible reform. *International Tax and Public Finance*, *10*, 511–533.

King, A., Baekgaard, H., & Robinson, M. (1999). *Dynamod2: An overview*. (Technical Paper 19*)*. Canberra, Australia: University of Canberra. Retrieved from https://www.natsem.canberra.edu.au/storage/tp19.pdf

Lambert, S., Percival, R., Schofield, D., & Paul, S. (1994). *An introduction to STINMOD: A static microsimulation model* (STINMOD technical paper No. 1). Canberra, Australia: National Centre for Social and Economic Modelling (NATSEM).

Liberati, P. (2001). The distributional effects of indirect tax changes in Italy. *International Tax and Public Finance*, *8*(1), 27–51.

Li, J., & O'Donoghue, C. (2012). *A methodological survey of dynamic microsimulation models*. (UNUMERIT Working Papers; No. 002). Maastricht, Netherlands: UNU-MERIT, Maastricht Economic and Social Research and Training Centre on Innovation and Technology.

Lomax, N., & Smith, A. (2017). Microsimulation for demography. *Australian Population Studies*, *1*(1), 73–85.

Massey, D.S., Denton, N.A. (1988). The dimensions of racial segregation. *Social Forces, 67*, pp. 281-315.

Marois, G., & Bélanger, A. (2014). Microsimulation model projecting small area populations using contextual variables: An application to the Montreal metropolitan area, 2006-2031. *International Journal of Microsimulation*, *7*(1), 158–193.

Malenfant, É. C., Lebel, A., & Martel, L. (2015). *Projections of the diversity of the Canadian population, 2006 to 2031*. Retrieved from https://www150.statcan.gc.ca/n1/pub/91-551-x/91-551-x2010001-eng.htm

McCann, P. (2014). *Modern urban and regional Economics*. Oxford, England: Oxford University Press.

Milne, B., Lay Yee, R., McLay, J. M., Pearson, J., Von Randow, M., & Davis, P. (2015). Modelling the early life-course (MELC): A microsimulation model of child development in New Zealand. *International Journal of Microsimulation*, *International Microsimulation Association*, *8*(2), 28–60.

Mondal, M., Cameron, M.P., & Poot, J. (2020). Cultural and economic residential sorting of Auckland's population, 1991-2013: An entropy approach. *Journal of Geographical Systems.* doi:10.1007/s10109-020-00327-1

Mondal, M., Cameron, M.P., & Poot, J. (2020). *Determinants of ethnic identity among adolescents: Evidence from New Zealand* (Working paper in Economics No.5/20). Hamilton, New Zealand: University of Waikato.

Mondal, M., Cameron, M.P., & Poot, J. (forthcoming). Group-size bias in the measurement of residential sorting. In R. Patuelli & S. Suzuki (Eds.) *A broad view of regional science - Essays in honor of Peter Nijkamp.* Berlin, Germany: Springer Nature.

Mot, E. S. (1992). *Survey of microsimulation models*. The Hague, Netherlands: VUGA.

Nijkamp, P., & Poot, J. (2015). Cultural diversity: A matter of measurement. In P. Nijkamp, J. Poot J and J. Bakens (Eds.) *The economics of cultural diversity* (pp. 17-51). Cheltenham, England: Edward Elgar.

O'Sullivan (2009). Changing neighborhoods–neighborhoods changing: A framework for spatially explicit agent-based models of social systems. *Sociological Methods & Research*, *37* (4) , 498-530.

Paulus, A., Čok,M., Figari, F., Hegedüs, P., Kump, N., Lelkes, O., Levy, H., Lietz, C., Lüpsik, S., Mantovani, D., Morawski, L., Sutherland, H., Szivos, P., & Võrk, A. (2009). *The effects of taxes and benefits on income distribution in the enlarged EU.* EUROMOD Working Paper (No. EM8/09). Colchester, England: University of Essex.

Pechman, J., & Okner, B. (1974). *Who bears the tax burden?* Washington, DC: The Brookings Institution.

Poot, J. (1987). Estimating duration-of-residence distributions: Age, sex and occupational differentials in New Zealand. *New Zealand Geographer*, *43*(1), 23–32. https://doi.org/10.1111/j.1745-7939.1987.tb01205.x

Rake, K., & Zaidi, A. (2001). Dynamic microsimulation models: a review and some lessons for SAGE. Retrieved from http://www.lse.ac.uk/collections/SAGE/pdf/SAGE_DP2.pdf

Reardon, S.F., Firebaugh, G. (2002) Measures of multigroup segregation. *Sociological Methodology*, *32*, 33–67.

Reardon, S.F., Farrell, C.R., Matthews, S.A., O'Sullivan, D., Bischoff, K., & Firebaugh G. (2009). Race and space in the 1990s: Changes in the geographic scale of racial residential segregation, 1990-2000. *Social Science Research*, *38* (1), 55-70.

Rees P.H., Wohland P., Norman P., Lomax N., & Clark S.D. (2017). Population projections by ethnicity: Challenges and solutions for the United Kingdom. In D. A. Swanson (Ed.), *The frontiers of applied demography* (Volume 9, pp. 383-408). Applied Demography Series. Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-43329-5_18

Rephann, T. (2004). Economic-demographic effects of immigration: Results from a dynamic spatial microsimulation model. *International Regional Science Review*, *27*(4), 379-410. doi: 10.1177/0160017604267628.

Rowe, G., & Wilfson, M., (2000). Public pensions-Canadian analysis based on the Lifepaths generational accounting framework. 6th Nordic Seminar on Microsimulation Models, Copenhagen, Denmark. Retrieved from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.515.9275&rep=rep1&type=pdf

Rogers, S., Rineer, J., Scruggs, M., Wheaton, W., Cooley, P., Roberts, D., & Wagener, D. (2014). A geospatial dynamic microsimulation model for household population projections. *International Journal of Microsimulation*, *7*(2), 119-146.

Sahn, D., & Younger, S. (2003). Estimating the incidence of indirect taxes in developing countries. In F. Bourguignon and P. da Silva L. (Eds.) *The impact of economic policies on poverty and income distribution: Evaluation techniques and tools* (pp. 27-40). Washington, DC: World Bank Publications.

Schelling, T.C. (1971) Dynamic models of segregation. *Journal of Mathematical Sociology*, *1*, 143–186. https://doi.org/10.1080/0022250X.1971.9989794

Siegel, J. S. (2002). Applied Demography: Applications to business, government, law and public policy. San Diego, CA: Academic Press.

Simkus, A.A. (1978). Residential segregation by occupation and race in ten urbanized areas, 1950-1970. *American Sociological Review*, *43*, 81–93. https://doi.org/10.2307/2094763

Spielauer, M. (2011). What is social science microsimulation? *Social Science Computer Review*, *29*(1), 9-20.

Statistics Canada. (2018). *Demosim*. Retrieved from https://www.statcan.gc.ca/eng/microsimulation/demosim/demosim

Statistics New Zealand. (2007a). *Profile of New Zealander Responses, Ethnicity Question: 2006 Census*. Retrieved from http://archive.stats.govt.nz/

Statistics New Zealand. (2007b). *Survey of dynamics and motivations for migration in New Zealand – information releases.* Retrieved from http://www.stats.govt.nz/

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

Statistics New Zealand (2014). *Linking censuses: New Zealand longitudinal census 1981–2006.* Retrieved from www.stats.govt.nz.

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

Statistics New Zealand. (2016). *National Population Projections: 2016(base)–2068.* Retrieved from http://infoshare.stats.govt.nz

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

Statistics New Zealand. (2018). *Census*. Retrieved from https://www.stats.govt.nz/topics/census

Statistics New Zealand. (2020). *Ethnic group (detailed total response - level 3) by age and sex, for the census usually resident population count, 2006, 2013, and 2018 Censuses (RC, TA, SA2, DHB)*. Retrieved from http://nzdotstat.stats.govt.nz

Tayman, J. & Swanson, D.A. (1999). On the validity of MAPE as a measure of population forecasting accuracy. *Population Research and Policy Review*, *18*, 523-528.

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

Theil, H., Finezza, A.J. (1971). A note on the measurement of racial integration of schools by means of informational concepts. *Journal of Mathematical Sociology*, *1*, 187-194.

Uyeki, E.S. (1964) Residential distribution and stratification, 1950-1960. *American Journal of Sociology*, *69*, 491–498.

Vencatasawmy, C.P., Holme, K., Rephann, T., Esko, J., Swan, N., Öhman, M., Åström, M., Alfredsson,E., & Siikavaara, J. (1999). *Building a spatial microsimulation model*. Retrieved from https://pdfs.semanticscholar.org/3f01/36d2633e1ca80c0a228aca492e25c1a9c9dc

Willekens, F. (2006) *Description of the micro-simulation model (continuous-time micro-simulation)*. Deliverable D8 (first part), MicMac Bridging the micro-macro gap in population forecasting. The Hague, Netherlands: NIDI.

Willekens, F. (2016). Migration flows: Measurement, analysis and modeling. In M.J. White (Ed.), *International handbook of migration and population distribution* (pp. 225-24). Dordrecht, Netherlands: Springer. https ://DOI 10.1007/978-94-017-7282-2_11

Williamson, P., Birkin, M. & Rees, P. (1998). The estimation of population microdata by using data from small area statistics and samples of anonymised records. *Environment and Planning A, 30*, 785–816

Wilson, T. (2012). Forecast accuracy and uncertainty of Australian Bureau of Statistics state and territory population projections. *International Journal of Population Research*, 1-16. https://doi.org/10.1155/2012/419824

Wilson, T. (2015). Short-term forecast error of Australian local government area population projections. *Australasian Journal of Regional Studies, 21*, 253-275.

Wolfson, M, & Rowe, G. (2013). HealthPaths: Using health trajectories to estimate and simulate the relative importance of determinants of health-adjusted life expectancy (HALE): a statistical analysis. *The Lancet*, *381*(2). doi.org/10.1016/S0140-6736(13)61402-6.pdf

White, M.J. (1986). Segregation and diversity measures in population distribution. *Population Index, 52*,198-221.

Zaidi, A., & Rake, K. (2001). Dynamic microsimulation models: a review and some lessons for SAGE. (Discussion Paper. 2). London, England: The London School of Economics, Simulating Social Policy in an Ageing Society (SAGE).

Zucchelli, E., Jones, A.M., & Rice, N. (2010). *The evaluation of health policies through microsimulation methods.* (HEDG working paper number 10/03). York, England: Department of Economics, University of York.

Chapter 6: Conclusion

There has been increasing global migration, coinciding with rising racial, ethnic and religious diversity and related tensions in the past few decades around the world (Pew Research Center 2019). Consequently, the cultural make-up of New Zealand has become more diverse with high population growth, immigration and inter-ethnic marriages (Statistics New Zealand 2019). The patterns of growing diversity of New Zealand vary by region and sub-region. A strategic approach is required to utilise and manage this diversity. Auckland is the most populous and the most diverse region in New Zealand. Through four inter-connected studies, this thesis examined the sub-regional (area unit) ethnic make-up of Auckland, using more disaggregated ethnic groups than those used in previous research in New Zealand. An important motive to examine the residential sorting and diversity at the chosen level of disaggregation of ethnic groups is that the past studies on residential sorting has tended to obscure the expected heterogeneity within the broad ethnic groups. Another innovation in the thesis is that, given the very limited extant studies on diversity and residential sorting in Auckland in terms of population characteristics other than ethnicity, the thesis also addresses economic diversity in terms of income, education and occupation. Moreover, the high ethnic mobility rate among adolescents, along with the rising number of younger individuals with multiple ethnicities in the census, motivated the thesis to examine the factors affecting the self-identified ethnic identity decisions among adolescents in Auckland. This thesis examined the past ethnic diversity observed in Auckland and also proposed, developed and validated a spatial microsimulation model that can be used in projecting the future ethnic diversity in Auckland at a small spatial scale.

6.1. Main Findings

Chapter 2 of this thesis contributed to the relatively few systematic analyses in the literature that are concerned with finding the "best" measure of residential sorting. This chapter provided an empirical analysis using New Zealand census micro-data from 1991 to 2013 in Auckland and demonstrated the sensitivity of the traditional measures of residential sorting to group size. It is important to identify this sensitivity of the common measures of residential sorting to group size in order to correctly compare the sorting indices across groups of varied sizes both cross-sectionally and over time. The analysis showed that the relationship between group size and the Entropy Index of Systematic Segregation was the weakest among all the measures of residential sorting considered, and hence that measure should be preferred in analyses of residential sorting in future applications.

Chapter 3 of this thesis provided descriptive evidence on the long-term patterns of residential sorting and diversity in Auckland not only by ethnicity, but also in terms of economic factors (specifically education, occupation and income) among people aged 22 and above. Using entropy-based measures of residential sorting, this chapter provided evidence of the primacy of residential sorting by ethnicity in comparison to the selected economic factors, although residential sorting had been generally declining over time. The results demonstrated that the larger ethnic groups in Auckland, like the New Zealand European and New Zealand Māori groups, were consistently more evenly dispersed spatially, whereas the smaller ethnic groups such as the African, Latin American/Hispanic, and Tokelauan groups were the least evenly dispersed. The Theil Multi-group Index decomposition results showed the presence of spatial heterogeneity of the finer ethnic groups, i.e. individuals were increasingly tending to co-locate more according to their finer ethnic groups than their broad ethnic groups. As expected, the results showed that the conclusions about residential sorting changed depending on the level of disaggregation of ethnic groups. This demonstrated the importance of considering the appropriate level of aggregation in the studies related to residential sorting.

Chapter 4 of this thesis focussed on the dynamics of self-declared ethnic identities of adolescents in Auckland. Using New Zealand Linked Census data for four inter-censal periods between 1991 and 2013, the same individuals were linked across two consecutive Censuses, where in the first Census their parents were likely to have recorded the adolescent's ethnicity, and in the second Census the adolescents were likely to have recorded their own ethnicity. Each and every ethnicity that an adolescent reported in the later census of the inter-censal period was included in the analysis, and hence logistic regression analysis (rather than multinomial logistic regression) was used with linked data pooled across all the inter-censal periods. Considering adolescents (aged between 13 to 17 in the previous census) who transitioned to adulthood in the later census, the chapter identified the major determinants of the first conscious ethnic affiliation of adolescents as their sex, age, whether New Zealand-born, ethnicity stated at the previous census, parents' ethnicity, and the ethnic makeup of their neighbourhood.

Chapters 2 and 3 provided insight into the past observed patterns of ethnic diversity and residential sorting among people aged 22 and above in Auckland, whereas Chapter 4 identified the factors impacting the transitions in the ethnic identity affiliations in adolescents in the past in Auckland. These results have implications for future spatial ethnic diversity and sorting patterns. Chapter 5 expanded on these ideas and developed a spatial microsimulation model to project the future ethnic diversity and residential sorting in Auckland for all ages and all ethnic groups at a fine level of classification, using New Zealand Linked Census data from 1996-2001. The model was validated using Census data from 2006. The chapter described the calibration processes undertaken in order to validate the model, and summarised the extent of error in the resulting simulated data projected forward to 2006. The transitions in the decision of individuals to change from one ethnicity to other, and change their location, which were used to project the model forward, were predicted using logistic regression analysis. The results show that census data can be used to inform, calibrate and validate our model. The microsimulation model can reproduce the dynamics of residential sorting in Auckland, without detailed information on all the elements of an individual's residential decision-making process.

In this thesis, entropy-based measures of residential sorting, which are the least biased by group size, as shown in Chapter 2, were used to measure residential sorting in Chapter 3. Results showed that area unit diversity in Auckland is greater in terms of ethnicity than other socio-economics variables. Residential sorting was also found to be greater by ethnicity than by other socio-economic variables. Along with other factors, multiple-ethnic affiliation and ethnic mobility results in rising ethnic diversity (Cameron and Poot 2019). As young individuals change their ethnic identity the most (Statistics New Zealand 2009), Chapter 4 identifies the factors that shape an adolescent's ethnic identity. Results showed a significant relationship between adolescents' ethnic identity and their age, sex, country of birth, and their previous ethnic affiliation. These variables were then used in Chapter 5 to calculate the transitional probabilities for the ethnic and locational transition modules for children and adolescents (aged 0-17 years), to project the population in the dynamic microsimulation model. A similar model was constructed for the adult ethnic transitions, and an additional module was separately constructed for locational transitions. The four inter-related chapters in the thesis provide insights about the past sub-regional (area unit) ethnic make-up of Auckland, New Zealand and also constructed and calibrated a dynamic microsimulation to project the future ethnic diversity.

6.2. Policy Implications

The results in this thesis have several important implications for policy.

First, residential sorting has been decreasing in Auckland over time. However, particular ethnic groups tend to co-locate with each other. This recognizes that diverse ethnic groups respond differently to residential choices. The diversity in responses and characteristics are in terms of preferences, health needs, beliefs, behaviours etc. In the areas where specific ethnic groups are particularly clustered, health, education and other social services should be tailored to those ethnic groups. This would improve the quality of services that each ethnic group receives and also make provision of services less costly. While planning for future public services like health, education and community services, policymakers should consider the rising diversity as well as decreasing residential sorting, as many of these services are or can be targeted at particular ethnic groups. The combination of rising diversity and declining residential sorting will make culturally-based provision both more important and more difficult, since spatial targeting becomes less feasible. Where areas are forecasted to be affected mostly by diversity, adequate level of planning for provision of public services and infrastructure is needed.

Second, patterns of cultural diversity in Auckland vary by sub-region. Policies should be spatially targeted to eliminate social-economic disadvantages that arise due to residential sorting. If poor neighbourhoods are concentrated with particular ethnic groups, pre-existing inequalities in terms of earnings, wealth and poverty (Grodsky and Pager 2001) may be intensified. This makes some groups more vulnerable to social problems (Massey and Denton 1993; Halpern-Felsher et al. 1997). If different areas favour one group over another (in terms of employment opportunities, healthcare, education, housing facilities, etc.), policies should be area-specific and also aim at reducing the spatial differences.

Third, individuals co-locate more according to their finer ethnic groups than broader ethnic groups. Policy makers should consider the heterogeneity present within the broad ethnic group levels when planning ethnic-group specific provision of services. This is because, due to the heterogeneity within the broad ethnic groups, the needs and the choices of the subgroups differ, which creates a challenge. There might be subgroups within a broad ethnic group that are improving their standard of living, while other subgroups within the same broad ethnic group face severe hardship. This could lead to misallocation of resources by government attempting to improve standards of living or to address poverty issues. Ethnic-group-specific provisions become especially difficult when the boundaries of ethnic groups are fluid (Callister 2007). Moreover, results show that within-group ethnic residential sorting is more than between group ethnic sorting in Auckland. Thus, policy makers should emphasise on clearly defining the ethnic groups, considering the within-group differences.

Fourth, people's ethnic affiliation shows a great degree of persistence, even inter-generationally. Even though the census treats ethnicity as a cultural affiliation not related to ancestry, in relation to defining an ethnic group both the ancestry and the self-identified ethnic group should be considered. This is because, if policies favour a certain ethnic group, individuals might identify with that ethnic group to benefit from the policy.

Fifth, ethnic diversity has increased in Auckland over time. Cultural diversity brings with it a diverse range of societal norms, customs, and ethics. These may have positive impacts on technological innovation, diffusion of new ideas and increase in production of different types of goods and services (Ozgen 2021). Individuals have different ideas and abilities, which can be important for technological progress. Diversity has productivity-enhancing effects as workers belonging to different backgrounds have different skills, experiences and abilities (Bove and Elia, 2017). Moreover, results show that cultural diversity is higher than economic diversity in Auckland. Thus, policy makers should aim at promoting the skills, talents and innovations that diversity brings and create an inclusive environment in the society, and also find ways to support entrepreneurship within and between the diverse communities.

6.3. Thesis Contribution

This thesis made several important contributions. First, it contributed to the relatively scarce literature on the systematic analysis of presence of group-size bias in the traditional measures of residential sorting. Second, it used entropy measures of spatial diversity and residential sorting, which is new in New Zealand. Third, this thesis not only used more disaggregated groups than previous studies, but the thesis was also the first study in New Zealand to consider residential sorting within and between ethnic groups. This is important as past studies (e.g. Johnston et al. 2002; 2008; 2011) have considered only broad ethnic groups (specifically New Zealand European/Pākehā, Māori, Pacific, and Asian) and have thereby ignored the expected heterogeneity in choices and preferences among the finer ethnic groups that make up each broad ethnic group. For example, the 'Asian' broad ethnic group is insufficient to capture the diversity of the finer Asian ethnic groups, such as the South East Asian, Chinese, and Indian groups within the broader Asian group. As previously explored by Maré and Coleman (2011), more in-depth regression modelling of residential location should be undertaken that uses more disaggregated groups. The results from the thesis provided support that it is becoming increasingly important to look at residential sorting at a finer scale, as the spatial heterogeneity of the finer ethnic groups can create both problems as well as opportunities for public services

(Caldwell et al. 2017). Moreover, this thesis also considered every ethnicity that the person reported as part of their ethnic identity. This was a relatively new approach and supersedes the approach of prioritised ethnicity popularly used in past studies (e.g., Johnston et al. 2005; Maré et al. 2012; Maré et al. 2016). The prioritised ethnicity approach ensure that every person is allocated only one ethnicity despite the fact that they might have reported belonging to multiple ethnic groups. The approach adopted in the thesis avoided the possibility of ignoring a lot of expected diversity arising as a consequence of multiple-ethnic affiliation, as demonstrated in Cameron and Poot (2019).

Additionally, this thesis extended the existing knowledge of residential sorting in New Zealand by looking at long-term trends (close to a quarter century) in residential sorting. This thesis also contributed to the very limited existing literature on residential sorting in terms of characteristics other than ethnicity in New Zealand by also focussing on economic characteristics. Moreover, following Florida and Mellander (2018) this thesis used an overall measure of economic sorting in Auckland by means of a combination of income, occupation and qualification. This is also a novel approach in New Zealand.

Moreover, this thesis contributed in a novel way to the small but growing literature on adolescent ethnic identity development in New Zealand. Unlike most of the past research in New Zealand and elsewhere (e.g. Coope and Piesse 2000; Qian 2004; Kickett-Tucker 2009), which considered a generic minority or a limited number of ethnic groups, this research considered all of the fine ethnic groups as multiple ethnic affiliation and ethnic mobility is not restricted to minority ethnic groups.

This thesis also contributed to the relatively limited number of studies conducting microsimulation modelling of ethnic diversity in New Zealand and elsewhere. This thesis developed and validated a spatial dynamic microsimulation model using actual census data from 1996-2001 which can be used to project the future ethnic diversity at a finer spatial scale and including finer ethnic groups than that used in previous research related to ethnicity in New Zealand or elsewhere (e.g. Ardestani 2013, Malenfant et.al. 2015, Ardestani et al. 2018, Davis and Lay-Yee 2019).

One of the most novel contributions of this thesis was the use of statistical software Stata to run the spatial dynamic microsimulation model. Previous similar research, like Demosim and DYNACAN in Canada, DYNAMOD 2 in Australia and Pensim in the UK, has used other programming languages. By running the model in Stata, which is available inside the secured

Statistics New Zealand Datalab, we could prevent any bias arising due to the anonymisation[92] of the data. Also, by using Stata inside the datalab, we were able to use the entire 1996-2001 Auckland population as the base population for our model, rather than a sample of the population. This thesis has broadened the possibilities for the development of microsimulation models using software other than those that are traditionally used. This thesis has also showed how census data can be used to develop and validate a model for predicting the future ethnic makeup without comprehensive knowledge of all of the factors related to an individual's locational decisions, which increases the prospects for further research on population projections.

6.4. Limitations and Future Research

This thesis is not without its limitations. The first broad issue concerns the potential problems with the measurement of ethnicity. The inconsistency of the ethnicity question across successive censuses poses problems for calculation with, and interpretation of, past data and also for future projections. The inconsistencies become more relevant in more disaggregated groups. Thus, past diversity and sorting trends and projections based on the available data for the finer ethnic groups that have been affected in the specific years should not be over-interpreted. However, most of the variables used in the regressions were time invariant or structurally deterministic. Moreover, we combined the results for 'MELAA' and 'Other' ethnic groups in Chapter 4, due to the small number of adolescents reporting these ethnicities. These ethnicities could be investigated in more detail in future research, perhaps using qualitative methods, given the small group sizes. By using inter-censal fixed effects, and interactive dummy variables between affected ethnic groups and the inter-censal periods, the inter-censal bias can in principle be controlled for. This was the approach that we have taken in this thesis. The efficacy of the approach was confirmed by the results of our supplementary regressions in Chapter 4. When we ran the same regressions separately for different years the results were consistent with our regression results with data pooled across all years. Additionally, the most recent Census period used in this thesis was 2013. While there was a new Census in 2018, the required longitudinal data for the study was not available at the time of writing. When the

[92] A confidentiality rule required by Statistics New Zealand to take any data out of the secured Statistics New Zealand Datalab.

required data becomes available, it will be interesting to see any differences in the long-term diversity and sorting already observed in the analyses presented here.

Second, in terms of looking at residential sorting by occupation, only individuals who were employed were considered in this thesis. It would be interesting to see this analysis extended further by including the people who are unemployed or not in the labour force (e.g. students, or retirees).

Third, this thesis looked at the past ethnic make-up only for the city of Auckland. As sorting and diversity might not be uniform or similar throughout all the regions (Cameron and Poot 2019), this research could be extended to look at diversity and sorting patterns in New Zealand overall as well as in each region or metropolitan area.

Fourth, a limitation of this work is that we did not link the parents' ethnicity directly to that of the adolescents. The required data is not available for the whole adolescent population in the Census. Specifically, the data on parent's ethnicity is not available for children who were coded as an adult, who were absent in the previous census, or when there was a change in parents in the intervening periods. Thus, to avoid selection bias arising from missing data in the regression model, we imputed the parental ethnicity variables for all adolescents in the sample. Though we did not know the actual ethnicities of the parents, any resulting measurement error arising from this approach will lead to regression coefficients for parental ethnicity that are attenuated, i.e. over-conservative. By including only the households with no more than one male or female adult in the analysis, this measurement error was reduced. However, that potentially introduces a new source of measurement error, if adolescents living in households with more than two adults present differ systematically from those living in households with two or fewer adults. We also did not include data on siblings of the adolescent, also due to the fact that the data is available for only a subsample of the population. Siblings may have an effect on an adolescent's ethnic identity choice. Thus, this analysis could be extended using the available data on parent's ethnicity and/or for siblings, present for a subsample of the census data, to check the consistencies with the results reported in this thesis.

Fifth, an individual's location decision depends on a variety of factors, other than the ones that are used in the microsimulation model, one of them being their completed education level. We had data on the completed education for adults. But due to the lack of the same data for children transitioning to adulthood, we did not include education as one of our predictor variables in the models. Adding an additional 'educational transition' module in the model and observing the changes in the results could be an interesting direction for future research.

Sixth, this thesis is largely descriptive and the empirical work reported here does not necessarily reveal the causal mechanisms underlying diversity and sorting. The causes and consequences of current and future residential sorting for individual well-being and social capital are also interesting topics for future research. Our analysis on residential sorting and diversity could also be extended by including even more finer-grained ethnic groups (e.g. at Level 3 of the ethnic classification), other cultural variables (e.g. language and religion), or by combining cultural and socio-economic variables through cross-tabulated groups (e.g. ethnicity-income, ethnicity-education etc.). This might lead to the identification of more complex patterns of residential sorting than those observed in this thesis.

Despite the inevitable limitations, this thesis has provided an important update on the study of spatial sorting and diversity in Auckland and presented a novel attempt to facilitate understanding of the dynamics of ethnic diversity and sorting. Given the known limitations of commonly used population projection methodologies at small spatial scales, the method adopted in this thesis for small area population projections has great potential for future research at the national, regional, and sub-regional scales.

References

Ardestani, B.M. (2013). *Using a hybrid model for investigating residential segregation: An empirical and simulation-based study*. (Doctoral thesis). University of Auckland, New Zealand. Retrieved from https://researchspace.auckland.ac.nz/handle/2292/21618

Ardestani, B. M., O'Sullivan, D., & Davis, P. (2018). A multi-scaled agent-based model of residential segregation applied to a real metropolitan area. *Computers, Environment and Urban Systems*, *69*, 1–16. https://doi.org/10.1016/j.compenvurbsys.2017.11.002

Bove, V., & Elia, L. (2017). Migration, diversity, and economic growth. *World Development, 89,* 227-239.

Caldwell, J.T., Ford, C.L., Wallace, S.P., Wang, M.C., & Takahashi, L.M. (2017). Racial and ethnic residential segregation and access to health care in rural areas. *Health & Place, 43*, 104–112.

Callister, P. (2007). *Special measures to reduce ethnic disadvantage in New Zealand – An examination of their role*. Wellington, New Zealand: Institute of Policy Studies.

Cameron, M.P., & Poot, J. (2019). Towards superdiverse Aotearoa: Dimensions of past and future ethnic diversity in New Zealand and its regions. *New Zealand Population Review, 45*, 18-45.

Coope, P., & Piesse, A. (2000). *1991–1996 Intercensal consistency study*. Wellington, New Zealand: Statistics New Zealand.

Davis, P., & Lay-Yee, R. (2019). *Simulating societal change: Counterfactual modelling for social and policy Inquiry*. Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-04786-3

Florida, R., &Mellander, C. (2018). The geography of economic segregation. *Social Sciences, 7*, 1-17.

Grodsky, E., Pager, D. (2001). The structure of disadvantage: individual and occupational determinants of the Black-White wage gap. *American Sociological Review*, *66*, 542–567.

Halpern-Felsher, B., Connell, J., Spencer, M., Aber, J.L., Duncan ,G., Clifford, E., Crichlow, W., Usinger,P., Cole, S., Allen, L., & Seidman, E. (1997). Neighborhood and family factors predicting educational risk and attainment in African American and White children and adolescents. In: J. B. Gunn, G. J. Duncan, J. L. Aber (Eds.) *Neighborhood poverty: Context and consequences for children* (pp. 146-173). New York, NY: Russell Sage Foundation.

Johnston, R., Poulsen, M., & Forrest, J. (2002). Rethinking the analysis of ethnic residential patterns: segregation, isolation, or concentration thresholds in Auckland, New Zealand? *Geogrphical Analysis, 34*, 245–261.

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*, 115–129.

Johnston, R., Poulsen, M., & Forrest, J. (2008). Asians, Pacific Islanders, and ethnoburbs in Auckland, New Zealand. *Geographical Review, 98*, 214–241.

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, 1–23.

Kickett-Tucker, C. S. (2009). Moorn (black)? Djardak (white)? How come I don't fit in mum? Exploring the racial identity of Australian aboriginal children and youth. *Health Sociology Review*, *18*(1), 119–136.

Malenfant, É. C., Lebel, A., & Martel, L. (2015). *Projections of the diversity of the Canadian population, 2006 to 2031*. Retrieved from https://www150.statcan.gc.ca/n1/pub/91-551-x/91-551-x2010001-eng.htm

Maré, D.C., & Coleman, A. (2011). *Estimating the determinants of population location in Auckland.* Wellington, New Zealand: Motu Economic and Public Policy Research. Retrieved from https://papers.ssrn.com/abstract=1861531. Accessed 17 January 2019

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

Maré, D.C., Pinkerton, R.M., & Poot, J. (2016). Residential assimilation of immigrants: A cohort approach. *Migration Studies, 4*, 373–401.

Massey, D.S., Denton, N.A. (1993). *American apartheid: Segregation and the making of the underclass*. Cambridge, MA: Harvard University Press.

Ozgen, C. (2021) The economics of diversity: Innovation, productivity and the labour market. Journal of Economic Surveys, 35, 1168-1216. doi.org/10.1111/joes.12433Pew Research Center. (2019). *A changing world: Global views on diversity, gender equality, family life and the importance of religion*. Retrieved from https://www.pewresearch.org.

Qian, Z. (2004). Options: Racial/ethnic identification of children of intermarried couples. *Social Science Quarterly*, *85*(3), 746–766.

Statistics New Zealand. (2009). *Final report of a review of the official ethnicity standard: Statistics New Zealand*. Wellington. Retrieved from http://archive.stats.govt.nz/methods/data-integration/data-integration-projects/intercensal-ethnic-mobility-2001-2006.aspx

Statistics New Zealand. (2019). *New Zealand's population reflects growing diversity*. Retrieved from https://www.stats.govt.nz

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 2: Mondal, M., Cameron, M.P., & Poot, J. (forthcoming). Group-size bias in the measurement of residential sorting. In R. Patuelli & S. Suzuki (Eds.), A Broad View of Regional Science – Essays in Honor of Peter Nijkamp. Springer Nature

Nature of contribution by PhD candidate	Data analysis, result interpretation, preparing full draft, conference presentation, submission as a book chapter.
Extent of contribution by PhD candidate (%)	70%

CO-AUTHORS

Name	Nature of Contribution
MICHAEL P. CAMERON	Guidance, critical feedback, assisting with book chapter submission.
JACQUES POOT	Critical feedback, guidance and corresponding author.

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
MICHAEL CAMERON	[signature]	12-11-2020.
JACQUES POOT	[signature]	12-11-2020

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.	Chapter 3: Mondal, M., Cameron, M.P., & Poot, J. (2020). Cultural and economic residential sorting of Auckland's Population, 1991-2013: An entropy approach. Journal of Geographical Systems.
Nature of contribution by PhD candidate	Data analysis, result interpretation, preparing full draft, conference presentation, journal submission, journal revision
Extent of contribution by PhD candidate (%)	70%

CO-AUTHORS

Name	Nature of Contribution
MICHAEL P. CAMERON	Guidance, critical feedback, assisting with journal submission.
JACQUES POOT	critical feedback, assisting with journal submission.

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
MICHAEL CAMERON	[signature]	12-11-2020
JACQUES POOT	[signature]	12-11-2020

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 4: Mondal, M., Cameron, M.P., & Poot, J. (2020). Determinants of ethnic identity among adolescents: Evidence from New Zealand. (Working paper in Economics No. 5/20). University of Waikato, Hamilton.

Nature of contribution by PhD candidate	Data analysis, result interpretation, preparing full draft, conference presentation.
Extent of contribution by PhD candidate (%)	70 %

CO-AUTHORS

Name	Nature of Contribution
MICHAEL P. CAMERON	Guidance, critical feedback, proof reading
JACQUES POOT	Critical feedback, helping in improvement of the chapter

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
MICHAEL CAMERON	[signature]	12-11-2020.
JACQUES POOT	[signature]	12-11-2020

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

chapter 5: Mondal, M., Cameron, M. P, & Poot, J. (forthcoming). Projecting the spatial distribution of ethnic groups in Auckland: Development of a spatial dynamic microsimulation model (Working Paper in Economics). University of Waikato.

Nature of contribution by PhD candidate	Development of the model, data analysis, result interpretation, preparing initial draft.
Extent of contribution by PhD candidate (%)	70%

CO-AUTHORS

Name	Nature of Contribution
MICHAEL P. CAMERON	Conceptualising the study, guidance, critical feedback
JACQUES POOT	Guidance, critical feedback

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
MICHAEL CAMERON	[signature]	25-11-20
JACQUES POOT	[signature]	25-11-20